
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Mehrabidavoodabadi, Abbas; Siekkinen, Matti; Ylä-Jääski, Antti
Green Edge Adaptive Mobile Video Streaming with Device-to-Device Communication

Published in:
IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2019

Published: 06/09/2019

Document Version
Peer reviewed version

Please cite the original version:
Mehrabidavoodabadi, A., Siekkinen, M., & Ylä-Jääski, A. (2019). Green Edge Adaptive Mobile Video Streaming with Device-to-Device Communication. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2019* (pp. 1-7). IEEE Communications Society.

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Green Edge Adaptive Mobile Video Streaming with Device-to-Device Communication

Abbas Mehrabi, Matti Siekkinen, and Antti Ylä-Jääski

Department of Computer Science, Aalto University, Espoo, Finland.

Emails: {abbas.mehrabidavoodabadi,matti.siekkinen,antti.yla-jaaski}@aalto.fi

Abstract—Green mobile edge computing (GMEC) refers to the integration of intermittent renewable energy sources into mobile edge computing architecture. It is a promising solution to overcome the challenges posed by grid-powered edge servers. On the other hand, direct device-to-device (D2D) communication and caching helps to alleviate the processing tasks at the edges and reduce traffic in the mobile backhaul. Using either D2D or energy harvesting functionality alone causes some performance loss which can be compensated by enabling both functionalities into the MEC. Motivated by this fact, we investigate how enabling D2D into the green MEC helps to save the grid/renewable energy consumption without sacrificing user experience or the traffic localization benefits, specifically in the case of on demand video streaming. To this end, we formulate a joint optimization of quality of experience (QoE), backhaul traffic, and grid energy consumption in D2D-enabled GMEC and design a heuristic-based algorithm with low complexity to solve the problem. Simulation results using both radio link information of the clients and measured solar harvested energy reveal that integrating D2D into GMEC for DASH indeed helps to significantly save the grid/renewable energy and reduce the backhaul traffic with marginal improvement in quality of video delivered to the clients.

Index Terms—Green computing, multi-access edge computing (MEC), device-to-device (D2D) communication, dynamic adaptive streaming over HTTP (DASH), quality of experience (QoE), backhaul traffic.

I. INTRODUCTION

Due to time-varying characteristics of wireless channels, dynamic adaptive video streaming over HTTP (DASH) is the dominant streaming protocol used at the moment for on-demand mobile video streaming [1]. DASH enables video streaming clients to adapt dynamically to time-varying network conditions by continuously selecting most suitable video bitrate [1]. Nowadays, DASH-driven video streaming solutions are mainly client-based which determine video bitrate based on either the average throughput [4] or on the playback buffer occupancy [3]. Due to lack of coordination, bitrate adaptation solutions following server and network assisted DASH are emerging [2].

Recent advancements in multi-access edge computing (MEC) have facilitated the implementation of network-assisted bitrate adaptation solutions [12]. Collaborative DASH video caching and processing at the edges of the network has been shown to significantly reduce the outbound traffic on the backhaul network [5]. Utilizing the processing capability of the edges, a higher quality video chunk can be transrated to a lower quality when such a video chunk is not readily available at the cache, hence reducing load of the mobile backhaul. In collaborative edge caching for DASH video streaming services, authors

in [6] address the trade-off between the QoE of the mobile clients and created traffic on backhaul network. On the client side, one-hop device-to-device (D2D) communication and caching has been proposed for reducing the backhaul traffic since the requested chunks of clients can be directly fetched from another device in the vicinity [16].

Current MEC solutions are mainly powered by non-renewable energy sources which unavoidably causes the environmental concerns in long term system operation [7]. With the dense deployments of edge devices in the next generation of 5G mobile networks, the integration of green renewable energy such as solar or wind into MEC devices known as green edge computing (GMEC) is a promising solution to overcome the energy efficiency challenges of the current MEC systems [7], [8]. The main challenge of utilizing green energy harvesting is the time-varying characteristic of environmental renewable energy during different time intervals. To adjust the supply and demand of the energy, the energy storages have been utilized in different communication systems [10], [13].

Although GMEC helps to reduce the grid energy consumption, its potential to reduce mobile backhaul load is smaller than that of adopting D2D. However, although enabling D2D may significantly reduce backhaul traffic, it causes higher grid energy consumption at edge servers and hence increases the operational costs for mobile network operators (MNO). Motivated by these shortcomings, we aim to study the impact of enabling D2D functionality into green MEC, which has not yet been explored. Our main contributions are summarized as follows:

- We propose the integration of D2D communication with green mobile edge computing (GMEC) for mobile on-demand adaptive video streaming.
- We formulate a problem to jointly optimize the QoE of mobile clients, backhaul traffic, and grid energy consumption in such a system as an integer nonlinear programming (INLP) model. Due to intractability, we design a low-complexity algorithm using auto-tuned parameterization technique to solve the problem.
- We examine the performance of resulting optimized system using simulations. The results are compared to competing solutions in terms of average grid/green energy consumption, QoE, and backhaul traffic.

II. GREEN EDGE ADAPTIVE MOBILE VIDEO STREAMING WITH D2D

In this section, we first introduce the proposed D2D-enabled GMEC system model for DASH and then define the related mathematical notations.

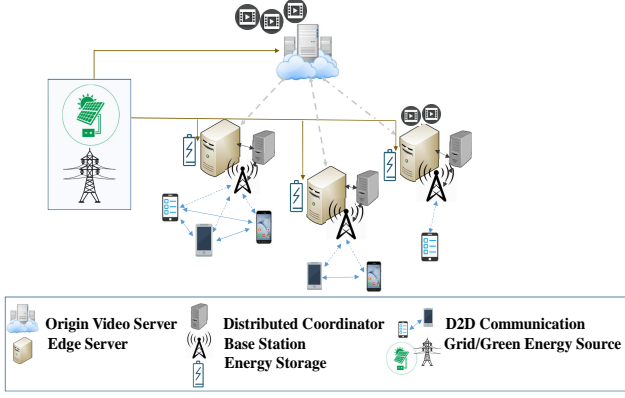


Fig. 1: Green edge adaptive video streaming with D2D.

A. System Model

Fig. 1 illustrates the schematic view of the proposed green edge adaptive mobile video streaming integrated with direct device-to-device (D2D) communication. Edge servers are associated with the cellular base stations (BSs) within the radio access network (RAN) from where the radio resource blocks (RBs) are allocated to the set of connected mobile clients in proportionally fair manner. Multiple set of videos are divided into consecutive chunks with fixed size and are initially stored in different qualities at the origin server in the cloud. After the clients request, the video chunks are cached at the edge servers for the future access of the clients. The distributed coordinators receive clients radio access link information from the BSs and the clients information (arrival, video request) from their application software to solve a joint optimization problem for optimal and fair video bitrate allocation.

In addition to the caching, the requested chunks of the clients can be also transrated at the edges from the chunks with higher bitrates which are available in the cache. Corresponding to the processing, the servers consume an amount of energy. The grid energy generated from the fuel resources and the renewable (green) energy are both injected into the edge servers through one single power line. In our system, we assume that these two sources of energy are separately controlled at the transmission point. The generated energy from the renewable sources is continuously stored in the storage from where it is then transmitted through the edge server power line. A smart controller is used in the system which enables the alternative switch between the grid and renewable energy transmission to the servers. Each time the processing takes place at the edges, the servers first use the renewable energy from the storage for the processing due to low operational costs of renewable energy and also saving the grid energy for the time intervals with high demands. Alternatively, they utilize the grid energy if the renewable energy does not suffice. The chunk transrating operation cannot be performed if none of the storage or grid energy are sufficient for the current processing.

In a hierarchical caching structure, the allocated chunk/bitrate to the client by the system coordinator is first looked-up in the cache of the local or neighborhood devices. If it is not available in the devices, the possibility

TABLE I: Description of system notations.

Notation	Description
K, S, R	Number of edge servers, number of DASH clients and the discrete set of available bitrates on each server
$ T , \Delta t, C$	Total number of scheduling time slots, the duration of each slot and the constant size of each video chunk in seconds
$W_k^{(t)}$	Available resource blocks at base station k in time slot t
Q_k, D_s	Cache size at respectively server k and mobile device s
$M_k^{(t)}, D_s^{(t)}$	Set of available chunks/bitrates in edge server k and mobile device s , respectively, at time slot t
A_s, D_s	Arrival and departure times of client s
$L_s^{max}, L_s^{(t)}$	Video buffer capacity of client s and its level at time slot t
$SNR_{sk}^{(t)}, Thr_{sk}^{(t)}$	Achievable signal-to-noise ratio and the theoretical data throughput by client s from base station k at time slot t
R_s, E_s, F_s	Average video bitrate, the accumulated bitrate switching and the associated fairness value for client s
BT_s	Overall backhaul traffic from video streaming of client s
ρ, ω, γ	Adjustable weighting parameters for average bitrate, switching and fairness, respectively
α, β, θ	Adjustable weighting parameters for QoE, data traffic and grid energy consumption, respectively
δ_S, δ_F	Switching and fairness thresholds
$a_{sk}^{(t)}, y_{sk}^{(t)}$	Binary indicator of allocating client s to server k at time slot t and, the binary decision variable indicating the processing of bitrate client s at edge sever k in time slot t
$x_{sk}^{(t)}, x_{sk}^{(t)}$	Binary decision variables indicating the chunk download from respectively the edge and origin server by client s at server k in time slot t
$r_{sk}^{(t)}, tr_{sk}^{(t)} \in R$	Discrete video bitrate (integer decision variable) of client s at time slot t and the transrated bitrate for client s at edge server k in time slot t
B_k, ϕ	Fixed capacity of energy storage at edge server k and constant processing weight of edge servers
$b_k^{(t)}, e$	The available green energy in the storage of edge server k at time slot t and constant consumed energy per bitrate
$h_k^{(t)}, c_k^{(t)}$	Harvested and consumed green energy by edge server k in time slot t
C_s^G, C_s^S	The overall consumed grid and green energy by client s during its video streaming session

of retrieval or transrating at the edge servers is then investigated. Finally, if the chunk can not be fetched or processed at the edges, it is directly downloaded from the origin server which in turn increases the backhaul traffic.

B. System Notation

We consider the scheduling of S number of DASH mobile video streaming clients in one round which consists of $|T|$ consecutive time slots each with fixed size of Δt . The MCE system includes K edge servers which are associated to the cellular base stations. The available resource blocks at the base station associated to edge server $1 \leq k \leq K$ at time slot t i.e., the number of subcarriers in the frequency domain, is represented by $W_k^{(t)}$. The arrival and departure time slots of client $1 \leq s \leq S$ which are respectively, the time that the client starts the video streaming session and the time that it either finishes or abandons the streaming session, are represented by A_s and D_s . The downlink signal-to-noise ratio of client s from the base station k at time slot t is represented by $SNR_{sk}^{(t)}$. Multiple videos with different popularities are divided into the consecutive chunks each with fixed size of C seconds and are initially stored at the origin server.

The variable $a_{sk}^{(t)}$ indicates the binary allocation of client s to edge server k at time slot t . Binary variables $x_{sk}^{(t)}$ and

$xc_{sk}^{(t)}$ indicate the retrieval of the chunk of client s allocated to server k at time slot t from respectively the edge and the origin server. Furthermore, the binary variable $y_{sk}^{(t)}$ indicates the transrating of the chunk of client s at edge server k in time slot t . Also, the decision variables $r_{sk}^{(t)}$ and $tr_{sk}^{(t)}$ denote respectively the allocated video bitrate to client s by the coordinator and the bitrate from which the bitrate of client s at server k in time slot t is transrated. The list of system notations including the edge/device cache size and energy harvesting/consumption parameters along with their definitions have been summarized in Table I.

C. Quality of Experience and Fairness

According to the research studies, the two major factors that directly impact the QoE of the mobile video streaming users are *the average video quality* and *video bitrate switching* [9]. Average quality perceived by client s during its video streaming session is obtained as follows.

$$R_s = \frac{C}{|D_s - A_s|} \sum_{p=1}^{\lceil (|D_s - A_s|)/C \rceil} a_{sk}^{(A_s + p \cdot C)} \cdot r_{sk}^{(A_s + p \cdot C)} \quad (1)$$

where r is the allocated bitrate to the client by the system coordinator which in some cases may be achieved by transrating from the chunks with higher bitrates available at the edges. The average magnitude of bitrate switching during the streaming session of client s is given by:

$$E_s = \frac{C}{|D_s - A_s|} \sum_{p=2}^{\lceil (|D_s - A_s|)/C \rceil} \sum_{k=1}^K a_{sk}^{(A_s + p \cdot C)} \cdot (r_{sk}^{(A_s + p \cdot C)} - r_{sk}^{(A_s + (p-1) \cdot C)}) \quad (2)$$

Due to the lack of coordination among the connected clients to base station, we explicitly incorporate the fairness into our optimization. More precisely, the bitrates of each client are allocated in a way that their different with the average bitrate of other simultaneous clients is minimized. Denoting F_s as the fairness index associated to client s , the fairness objective is stated as follows.

$$F_s = \sum_{t=A_s}^{D_s} \sum_{k=1}^K a_{sk}^{(t)} \cdot (r_{sk}^{(t)} - \bar{r}_k^{(t)}) \quad (3)$$

where $\bar{r}_k^{(t)}$ is the average bitrate of all clients allocated to base station (edge server) k at time slot t .

D. Backhaul Traffic

The average backhaul traffic due to downloading video chunks from the backhaul network during the video streaming session of client s is obtained as follows.

$$BT_s = \frac{1}{|D_s - A_s|} \sum_{t=A_s}^{D_s} \sum_{k=1}^K a_{sk}^{(t)} \cdot xc_{sk}^{(t)} \cdot r_{sk}^{(t)} \quad (4)$$

E. Energy Consumption Model

At each scheduling round, we assume that the fixed amount of grid energy is allocated to the edge servers which its quantity is randomly chosen from a uniform interval. The quantum of grid energy which is not consumed

in each round is accumulated and utilized in the future rounds when there are high volume of chunk processing.

In addition, the edge servers are also equipped with the energy storage which is used to keep the harvested renewable energy during different time intervals [13]. We assume the energy storages are kind of low-cost ideal lithium batteries with almost no-energy loss which their installation/maintenance normally occurs once per couple of years [14]. The available energy in the storage of server k at the beginning of time slot t follows the linear evolution model [11]:

$$b_k^{(t)} = \min\{b_k^{(t-1)} - c_k^{(t-1)} + h_k^{(t)}, B_e\}, \quad \forall 1 \leq t \leq |T| \quad (5)$$

where $h_k^{(t)}$ and $c_k^{(t)}$ are respectively the amount of harvested energy at the beginning and the amount of consumed renewable energy during time slot t by edge server k . Obviously, the consumed renewable energy by server cannot exceed its available storage energy.

$$c_k^{(t)} \leq b_k^{(t)}, \quad \forall 1 \leq t \leq |T| \quad (6)$$

Since energy is consumed for transrating operations at edges, the average energy consumption at edges during the video streaming session of client s is obtained as follows.

$$C_s = \frac{1}{|D_s - A_s|} \sum_{t=A_s}^{D_s} \sum_{k=1}^K a_{sk}^{(t)} \sum_{k'=1}^K y_{skk'}^{(t)} (tr_{skk'}^{(t)} - r_{sk}^{(t)}) \phi \cdot e \quad (7)$$

In equation (7), variable $tr_{skk'}^{(t)}$ indicates the bitrate available at edge server k' at time slot t which is transrated to the desired bitrate allocated by the coordinator to client s at edge server k ($r_{sk}^{(t)}$). The average consumed energy at the edges during video streaming of the client includes both grid and renewable energy consumption: $C_s = C_s^G + C_s^S$.

III. JOINT OPTIMIZATION PROBLEM

The objective of system model is to optimize for each individual DASH client jointly the perceived QoE/fairness, the backhaul traffic and the consumed grid energy of the servers during its whole video streaming session. The joint optimization problem for individual client s which also takes into account the time-varying energy harvesting constraints is formulated as INLP model.

$$\text{Maximize}_{xc, xe, y, r} \alpha(\rho R_s - \omega E_s - \gamma F_s) - \beta BT_s - \theta(1/e) C_s^G \quad (8)$$

Subject to:

$$\sum_{s'=1}^S a_{s'k}^{(t)} \cdot \left[\frac{r_{s'k}^{(t)}}{Thr_{s'k}^{(t)}} \right] \leq W_k^{(t)}, \quad \forall 1 \leq k \leq K, 1 \leq t \leq |T| \quad (9)$$

$$0 < L_s^{(t)} \leq L_s^{max}, \quad \forall A_s \leq t \leq D_s \quad (10)$$

$$a_{sk}^{(t)} = \begin{cases} a_{sk}^{(t-1)} & \text{if } \lfloor t/C \rfloor \neq 0 \\ 1 & \text{if } \lfloor t/C \rfloor = 0 \text{ AND } k = \arg \max_{k'} \{SNR_{sk'}^{(t)}\} \\ 0 & \text{if } \lfloor t/C \rfloor = 0 \text{ AND } k \neq \arg \max_{k'} \{SNR_{sk'}^{(t)}\} \end{cases} \quad \forall 1 \leq k \leq K, A_s \leq t \leq D_s \quad (11)$$

$$x_{sk}^{(t)} + x_{sk}^{(t)} = 1, \quad \forall 1 \leq k \leq K, A_s \leq t \leq D_s \quad (12)$$

$$\sum_{k=1}^K y_{sk}^{(t)} \leq 1, \quad \forall A_s \leq t \leq D_s \quad (13)$$

$$x_{sk}^{(t)}, x_{sk}^{(t)}, y_{sk}^{(t)} \in \{0, 1\}, \quad \forall 1 \leq k \leq K, A_s \leq t \leq D_s \quad (14)$$

$$r_{sk}^{(t)}, tr_{sk}^{(t)} \in R, \quad \forall 1 \leq k \leq K, A_s \leq t \leq D_s \quad (15)$$

In addition, energy consumption equations (5),(6) and (7) are also included to the set of constraints of above INLP problem. Note that in objective function (8), the last term energy consumption is multiplied by the factor of $(1/e)$ in order to have same magnitude as first two terms.

Constraint (9) states that at any time slot, the overall allocated RBs to the connected clients at the base station does not exceed the total available RBs of the base station at that time slot. Constraint (10) ensures that no stalling (underflow) or overflow happens on the client's video buffer during its whole video streaming session. Relation (11) determines the client to server mapping such that at each time slot, the client is assigned to the nearest edge server from where it receives the highest downlink SNR from the associated BS. Constraint (12) states that at each time slot, the client downloads its allocated chunk/bitrate from either the edge of the network or from the origin server. Inequality (13) constraints that the transrating operation at each time slot can be performed in at most one edge server and finally, the set of constraints (14) and (15) specify the range of decision variables.

It is noteworthy to mention that the fact that optimizing only single term in the objective function causes the drop in some performance metrics was indeed our main motivation for designing the joint optimization problem.

IV. ENERGY-AWARE BITRATE SELECTION ALGORITHM

Due to NP-hardness of the formulated problem in (8)-(15), we design an online greedy-based bitrate selection algorithm with low complexity which uses an auto-tuned parameterization technique for easy practical deployment by MNOs. The pseudo-code of the proposed algorithm named Energy-aware Greedy Bitrate Allocation Algorithm (EGBAA) has been demonstrated in Algorithm 1.

At each time slot, the energy level of storage at the edge servers is updated using relation (5). For the set of active clients in current time slot, the buffer level and backhaul traffic are initialized and the client is then assigned to most suitable edge server according to relation (11). If the client is streaming in the middle of its current chunk, it is assigned to the same edge server and with the same video bitrate as the previous time slot. Otherwise, the subroutine Auto-tuned Bitrate Selection is invoked to decide on the optimal/fair bitrate to the new chunk.

A. Auto-tuned Bitrate Allocation

If the client is at the beginning of downloading the new chunk of video, the selection of most suitable bitrate should be decided for the new chunk. Our proposed auto-tuned bitrate selection heuristic uses two thresholds switching and fairness when deciding on the

selection of most suitable bitrate with respect to other simultaneous streaming clients. The switching threshold δ_S is determined knowing the fact that the highest bitrate switching happens when the bitrates are merely allocated based on the buffer status of the client [12] and the fairness threshold δ_F which determines the fairness level of the allocated bitrate is given as input to the algorithm.

Algorithm 1: Energy-aware Greedy Bitrate Allocation Algorithm (EGBAA)

- 1: **Input:** $|T|, K, R$: Number of time slots, number of edge servers, set of available discrete bitrates at the origin server
- 2: **Output:** Binary allocation $x_{sk}^{(t)}, x_{sk}^{(t)}, y_{sk}^{(t)}$ and integer bitrate allocation $r_{sk}^{(t)}, tr_{sk}^{(t)}$ for each client s , edge server $1 \leq k \leq K$ and time slot $1 \leq t \leq |T|$, VideoQuality, BackhaulTraffic
- 3: **for** each time slot $1 \leq t \leq |T|$ **do**
- 4: **for** each edge server $1 \leq k \leq K$ **do**
- 5: $b_k^{(t)} = b_k^{(t)} + h_k^{(t)}$;
- 6: **for** each client s such that $A_s \leq t \leq D_s$ **do**
- 7: **Initialize** BufferStatus, $R_s = 0, BT_s = 0$;
- 8: **Allocate** client s to server k according to (11)
- 9: **if** $(t - A_s) \bmod C \neq 1$ **then**
- 10: **Allocate** client s to the same server and with the same bitrate as with time slot $t - 1$;
- 11: **Update** BufferStatus, $L_s^{(t)}, R_s, BT_s$
- 12: **if** $(t - A_s) \bmod C = 1$ **then**
- 13: **Call** Subroutine Auto-tuned Bitrate Selection;
- 14: **if** $t == D_s$ **then**
- 15: BackhaulTraffic=BackhaulTraffic+ BT_s ;
- 16: VideoQuality=VideoQuality+ R_s ;
- 17: **Return** VideoQuality, BackhaulTraffic;

Subroutine 1: Auto-tuned Bitrate Selection

- 18: **if** $t - A_s \leq C$ **then**
 - 19: **Allocate** the highest available bitrate;
 - 20: **Update** BufferStatus, $L_s^{(t)}, R_s, BT_s$;
 - 21: **Compute** estThr and threshold δ_S ;
 - 22: **for** each bitrate $r \in R$ in decreasing order **do**
 - 23: **if** $r \leq \max(\text{estThr}, \hat{T}hr_{sk}^{(t)}, L_s^{(t)})$ **AND** allocation of r satisfy (9) and satisfy both switching δ_S and fairness threshold δ_F **then**
 - 24: **if** r achieves the maximum objective value (8) **then** $r_{sk}^{(t)} = r$;
 - 25: **if** $r_{sk}^{(t)} = 0$ **then**
 - 26: **for** each bitrate $r \in R$ in decreasing order **do**
 - 27: **if** $r \leq \max(\text{estThr}, \hat{T}hr_{sk}^{(t)}, L_s^{(t)})$ **AND** allocation of r satisfy (9) and satisfy switching threshold δ_S **then Perform** the same operations as in line 20;
 - 28: **if** $r_{sk}^{(t)} = 0$ **then**
 - 29: **for** each bitrate $r \in R$ in decreasing order **do**
 - 30: **if** $r \leq \max(\text{estThr}, \hat{T}hr_{sk}^{(t)}, L_s^{(t)})$ **AND** allocation of r satisfy (9) **then**
 - 31: **Perform** the same operations as in line 20;
 - 32: **Update** weighting parameters ρ, ω, γ at time slot t ;
 - 33: **Update** the binary decision variables $x_{sk}^{(t)}, x_{sk}^{(t)}, y_{sk}^{(t)}$;
 - 34: **Compute** R_s, E_s, F_s, BT_s and objective value of client s up to time slot t ;
 - 35: **Update** BufferStatus, $L_s^{(t)}$;
 - 36: **Update** the grid and storage energy of the servers;
 - 37: **Return** $R_s, BT_s, x_{sk}^{(t)}, x_{sk}^{(t)}, y_{sk}^{(t)}$;
-

In decreasing order of magnitude, the bitrate in set R

which satisfies the wireless resource allocation constraint at the base station (constraint (9)), satisfies both switching and fairness thresholds and also achieves the highest objective value (8) is selected as the bitrate for the current chunk of the client. If there is no such bitrate available in set R , the selection of most suitable bitrate which satisfies only the switching threshold compromising the fairness threshold is then investigated. And, finally, if there is no such bitrate available, the most suitable bitrate which maximizes the objective value (8) is chosen for the current chunk of the client. Note that the availability of both video chunk (and its corresponding bitrate) in hierarchical caching (device, edge and origin server) and enough grid/green energy at the edges for transrating are considered in the computation of objective value (8).

After the bitrate allocation to the current video chunk of the client, the values of binary decision variables are then determined based on the location from where the chunk/bitrate of the client was retrieved or processed. The weighting parameters of the objective function and the objective value are then computed at the current time slot. The weighting parameter ρ is determined based on how far the selected bitrate is from the highest available bitrate in set R and parameter ω is derived based on how far the selected bitrate is from the one which results in no switching with the previous chunk. The fairness weighting γ is also computed based on how far is the selected bitrate from the average bitrate of other simultaneous clients. The algorithm then updates client's video buffer status/level and green or renewable energy consumption of the servers.

B. Cache Replacement Heuristic

We design an efficient cache replacement heuristic which is run at the edges if some of the allocated chunks are downloaded from the origin server at the current time slot. For each chunk, the heuristic computes a caching likelihood value using two statistical probabilities. First, how likely the chunk will be requested by the clients assigned to the same edge server in future time slots. Second, how percentage of the clients have downloaded their video chunks in the previous time slots with the same bitrate as the chunk in question.

The first probability can be approximated using the information about the retention of the clients with respect to different videos. In practice, the origin video server normally keeps the retention information and can communicate them with the edge servers [15]. The second probability is computed using the clients requests in the previous time slots which can be readily retrieved from the clients access history. After computing the caching likelihood values for all the requested chunks in the current time slot, the heuristic then sorts the chunks in decreasing order of their caching likelihood values and inserts them into the cache until it is filled.

C. Complexity Analysis

At each time slot, the client to edge server mapping takes $O(K)$ time where K is number of edge servers. Our analysis shows the worst case time complexity of $O(K + C \cdot S + |R|^2 + |R| \cdot K + |T|)$ for running the Auto-tuned

Bitrate Selection procedure. Considering all the clients and edge servers, the overall time complexity of the algorithm is therefore of order $O(|T| \cdot S \cdot K \cdot (K + C \cdot S + |R|^2 + |R| \cdot K + |T|))$ during $|T|$ number of time slots.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed D2D-enabled GMEC for DASH video streaming through simulations using downlink radio link information of the mobile clients and the measured solar harvested energy. Particularly, we compare the following strategies.

- **Collaborative Caching and Processing using GMEC with D2D (GMEC_CCP-D2D):** Mobile edge servers collaborate in both caching and processing by utilizing both the grid and renewable energy. D2D communication is also enabled by this solution.
- **Collaborative Caching and Processing using GMEC without D2D (GMEC_CCP):** This strategy provides collaborative edge caching and processing but without the potential of D2D communication. Edge servers perform the chunk transrating by relying on both the grid and renewable energy harvesting.
- **Non-collaborative GMEC with D2D (GMEC_CP-D2D):** Edge servers handle the caching and processing tasks locally without collaboration with the neighborhood servers. Both D2D functionality and renewable energy harvesting are also enabled.
- **Collaborative MEC with D2D (MEC_CCP-D2D):** Although this strategy enables both collaborative edge caching/processing and D2D functionality, the edge servers are powered from only the grid electricity without the possibility of green energy harvesting.

A. Simulation Setup

We consider the scheduling of DASH mobile clients in a typical MEC system consists of 10 edge server (and associated BSs) and 100 number of mobile users (UEs) during 300 time slot. Each time slot has the fixed size of $\Delta t = 1$ second. The downlink radio access link level information of the clients during $5min$ time duration is obtained from the SimuLTE while their D2D signal-to-noise ratio during the same time duration is a random variable selected from the uniform interval $\text{Unif}[-10dB, 10dB]$. We further use the Shannon theorem [17] to obtain the instantaneous achievable throughput of the clients according to their downlink SNRs. There are four videos with different popularities which are divided into the consecutive chunks each with fixed size of $C = 5$ seconds. The chunks are available in ten different qualities $R = \{15, 17, 22, 26, 30, 35, 38, 43, 45, 50Mbps\}$ and are initially stored on the origin server in the cloud. A linear curve is also utilized to simulate the retention of the clients with respect to different videos.

The video buffer of each client has the fixed size of $250Mb$ and the size of the cache at edge servers is fixed at $3Gb$. In one round of scheduling (with $5min$ time duration), the allocated grid energy to the edge servers is selected from the uniform interval $\text{Unif}[1KJ, 1.5KJ]$. Furthermore, the capacity of the energy storage is fixed at $1KJ$. We also use the measured solar energy harvesting

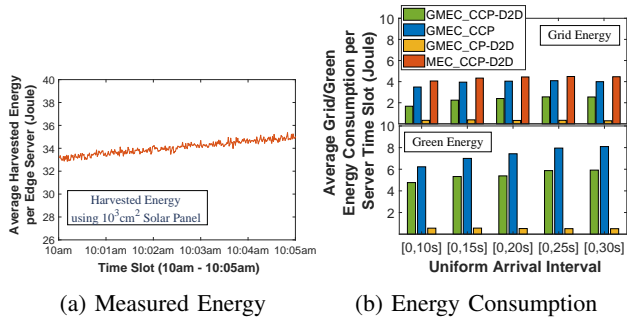


Fig. 2: (a) Measured solar energy harvesting adopted from [18] (b) comparison in term of energy consumption.

during a day in January at Hamburg city in Germany which is extracted from the experimental work in [18]. As depicted in Fig. 2a, the average harvested energy of 35 Joule is obtained during the time interval $10 \text{am} - 10:05 \text{am}$ using 10^3cm^2 -sized solar panel. Unless explicitly mentioned, we set the values of weighting parameters in the optimization problem to $\alpha = \beta = \theta = 1/3$. At each part of simulation, the average of the results taken over 20 runs of simulation with confidence interval of 95% are presented.

It is noteworthy to mention that although we set the weighting parameters in joint optimization problem to equal values, they can be tuned in practice based on joint operational interests of MNO and mobile clients. We should further highlight that despite of achieving our results based on some particular parameters setting, the results indeed verify that D2D-enabled GMEC system and the proposed algorithm are promising for use in general practical system setup.

B. Grid/Renewable Energy Consumption

We have first compared four above-mentioned strategies in terms of average grid/renewable energy consumption per edge server during the whole video streaming session of the clients. Comparison results for different arrival intervals of the clients have been shown in Fig. 2b. As we can see from the top subplot of Fig. 2b for the average grid energy consumption per edge server during 5min video streaming session, our proposed GMEC_CCP-D2D solution results in reducing on average about 40% grid energy consumption compared to GMEC_CCP without D2D functionality. The reason is that enabling D2D functionality into GMEC helps the clients to fetch some of their requested chunks/bitrates from the local or neighborhood devices therefore reducing the amount of processing tasks at the edge servers. Furthermore, the integration of renewable energy into MEC indeed saves on average about 47% grid energy compared to MEC_CCP-D2D where there is no possibility of energy harvesting. However, our solution causes about 85% more grid energy consumption compared to non-collaborative solution (GMEC_CP-D2D). This is because of larger volume of processing from not only the local but also the neighborhood clients in collaborative solution which in turn causes higher volume of energy consumption.

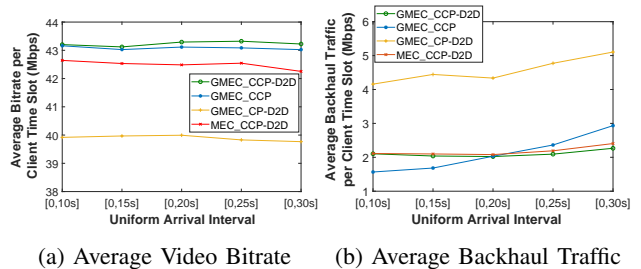


Fig. 3: Comparison in terms of (a) average video bitrate (b) average backhaul traffic per client at each time slot.

As we can see from the bottom subplot in Fig. 2b, enabling D2D into GMEC indeed helps to reduce the average green energy consumption per edge server by about 25% compared to GMEC_CCP. In contrast, our solution however results on average about 90% more green energy consumption compared to non-collaborative.

C. QoE and Backhaul Traffic Comparison

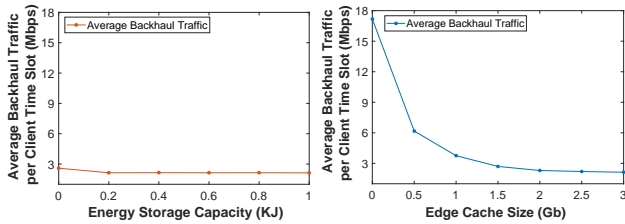
For different arrival intervals of the clients, we have next compared four strategies in terms of the average video bitrate and backhaul traffic per client at each time slot.

As we can see from the results in Fig. 3a, enabling D2D into GMEC yields only marginal improvement in the average video bitrate of the clients compared to GMEC_CCP without D2D. On the other side, our solution helps to slightly improve the average video bitrate compared to no energy harvesting strategy MEC_CCP-D2D and with more improvement compared to non-collaborative solution GMEC_CP-D2D. The reason is that although enabling D2D gives the opportunity for downloading some chunks with higher quality from the neighborhood devices, the potential of either collaboration with neighborhood servers or harvesting the renewable energy provides more opportunities for transcoding at the edges. This in turn provides the clients the opportunity to download more chunks with high quality from the edges. As observed from the results in Fig. 3a, the average improvements of 1% and 8% in bitrate are achieved using our solution compared to MEC_CCP-D2D and GMEC_CP-D2D while almost with the same average bitrate compared to GMEC_CCP without D2D.

We have also shown in Fig. 3b the average created backhaul traffic per client at each time slot using four strategies. As it is seen from the results, the proposed GMEC_CCP-D2D solution yields the average backhaul traffic reduction of about 3% and 53% compared to respectively MEC_CCP-D2D and GMEC_CP-D2D strategies. It is observed that collaborative edge caching/processing has higher potential for reducing the backhaul traffic compared to integrating the renewable energy. Although for few cases the GMEC_CCP solution yields lesser backhaul traffic, our solution achieves higher average bitrate and less energy consumption compared to GMEC_CCP.

D. Impact of Energy Storage Capacity vs. Cache Size

We have also studied the impact of increasing the capacity of energy storage of edge servers on the average backhaul traffic per client at each time slot. For this



(a) Impact of Storage Capacity (b) Impact of Cache Size

Fig. 4: (a) Impact of increasing energy storage capacity and (b) increasing the edge cache size on backhaul traffic.

simulation, we set $\alpha = 1, \beta = 0, \theta = 0$ in order to study the impact of increasing energy storage capacity or cache size on the average backhaul traffic in the best possible way. With fixed average energy harvesting of 35 Joule and fixed edge cache size at 3 Gb , the results when the storage capacity increases from 0 to 1 KJ has been shown in Fig. 4a. As we see, there is slightly reduction in the backhaul traffic when the capacity of energy storage increases.

Next, we increased the edge cache size from 0 to 3 Gb with fixed average energy harvesting magnitude of 35 Joule and fixed energy storage capacity of 1 KJ . The corresponding average backhaul traffics per client time slot have been shown in Fig. 4b. As we can see, the backhaul traffic dramatically decreases by increasing the edge cache size from 0 to 1.5 Gb .

Comparing the results of Fig. 4a and 4b under the same simulation setup, it is concluded that increasing the cache size has larger impact on backhaul traffic reduction compared to boosting the energy storage capacity. The reason is that the number of video chunks that are retrieved from the edge caches are much higher than the number of chunks that are transrated at the edges. This interesting observation can indeed guide the MNOs in the adjustments of edge cache size and energy storage capacity in economically optimal manner.

VI. CONCLUDING REMARKS

In this paper, we investigated the impact of enabling direct device-to-device (D2D) communication into the green mobile edge computing (GMEC) for particularly dynamic adaptive video streaming over HTTP (DASH). To this end, we formulated the joint optimization of QoE of the mobile clients, the backhaul traffic and grid energy consumption in D2D-enabled GMEC subject to time-varying characteristics of renewable energy harvesting during different time periods. We then designed a low-complexity heuristic-based bitrate selection algorithm to solve the optimization problem.

Results of our conducted simulations using SNR of mobile clients and the measured solar energy harvesting pattern reveal that enabling D2D into GMEC is indeed promising to save significantly the renewable/grid energy consumption of edge servers. Our results also confirm that D2D-enabled GMEC for DASH yields noticeable reduction in outbound traffic with marginal improvement in average video bitrate of the clients. Our results can be indeed used as guidelines for mobile network operators

(MNOs) to judge the effectiveness of enabling D2D into green MEC for DASH services in next generation of mobile networks.

ACKNOWLEDGMENT

This research has been financially supported by Lacrimosa project grant number 297892 and the Nokia Center for Advanced Research.

REFERENCES

- [1] ISO/IEC 23009-1. 2014. Dynamic adaptive streaming over HTTP (DASH) Part 1: Media presentation description and segment formats. (2014)
- [2] ISO/IEC 23009-5. 2017. Dynamic adaptive streaming over HTTP (DASH) Part 5: Server and network assisted DASH (SAND). (2017)
- [3] T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson "A buffer-based approach to rate adaptation: Evidence from a large video streaming service", in *Proc. ACM Conf. on SIGCOMM (SIGCOMM'14)*, pp. 187-198, Aug. 2014.
- [4] T. Mangla, N. Theera-Ampornpant, M. Ammar, E. Zegura, and S. Bagchi, "Video through a crystal ball: Effect of bandwidth prediction quality on adaptive streaming in mobile environments", in *Proc. 8th ACM Int. Workshop on Mobile Video*, pp. 1-6, May 2016.
- [5] T. X. Tran, P. Pandey, A. Hajisami, and D. Pompili, "Collaborative multi-bitrate video caching and processing in mobile-edge computing networks", in *Proc. 13th Annual IEEE Conf. on Wireless On-demand network systems and services (WONS)*, pp. 165-172, Feb. 2017.
- [6] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "QoE-traffic optimization through collaborative edge caching in adaptive mobile video streaming", *IEEE Access*, vol. 6, pp. 52261-52276, Sep. 2018.
- [7] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys & Tuts*, vol. 19, no. 4, pp. 2322-2358, Fourth Quarter 2017.
- [8] J. Xu, and S. Ren, "Online learning for offloading and autoscaling in renewable-powered mobile edge computing", in *Proc. IEEE Global Commun. Conf. (Globecom)*, pp. 1-6, Dec. 2016.
- [9] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hossfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming", *IEEE Commun. Surveys & Tuts.*, vol. 17, no. 1, pp. 469-492, First Quarter 2015.
- [10] P.-H. Chiang, R. B. Guruprasad, and S. Dey, "Optimal use of harvested solar, hybrid storage and base station resources for green cellular networks", *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 707-720, Sep. 2018.
- [11] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices", *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590-3605, Dec. 2016.
- [12] A. Mehrabi, M. Siekkinen, and A. Ylä-Jääski, "Edge computing assisted adaptive mobile video streaming", *IEEE Trans. Mobile Comput.*, vol. 18, no. 4, pp. 787-800, Apr. 2019.
- [13] M. Mendil, A. D. Domenico, V. Heiries, R. caire, and N. Hadjsaid, "Battery-aware optimization of green small cells: Sizing and energy management", *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 635-651, Sep. 2018.
- [14] X. Liu, and N. Ansari, "Dual-Battery enabled profit driven user association in green heterogeneous cellular networks", *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 1002-1011, Dec. 2018.
- [15] M. Siekkinen, M. A. Hoque, and J. K. Nurminen, "Using viewing statics to control energy and traffic overhead in mobile video streaming", *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1489-1503, Jun. 2016.
- [16] J. Liu, B. Bai, J. Zhang, K. B. Letaief, and Y. Li, "Joint device caching and channel allocation for D2D-assisted wireless content delivery", in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1-6, Jul. 2018.
- [17] LTE Guideline: http://www.etsi.org/deliver/etsi_tr/136900_136999/136942/08.02.00_60/tr_136942v080200.pdf
- [18] D. Krüger, S. Fischer, and C. Buschmann, "Solar power harvesting - Modeling and experiences", in *GI/ITG KuVS Fachgespräch Drahtlose Sensornetze*, vol. 8. Stuttgart, Germany: OPUS, 2009.