

# An optimized approach to video traffic splitting in heterogeneous wireless networks with energy and QoE considerations



Nadine Abbas<sup>a,\*</sup>, Hazem Hajj<sup>a</sup>, Zaher Dawy<sup>a</sup>, Karim Jahed<sup>b</sup>, Sanaa Sharafeddine<sup>b</sup>

<sup>a</sup> American University of Beirut, Lebanon

<sup>b</sup> Lebanese American University, Lebanon

## ARTICLE INFO

### Keywords:

QoE  
Heterogeneous networks  
Energy consumption  
Traffic splitting  
Multi-RAT

## ABSTRACT

Due to the exploding traffic demands with the ubiquitous anticipated spread of 5G and Internet of Things, research has been active to devise mechanisms for meeting these demands while maintaining high quality user experience. In support of this direction, 3GPP is working towards cellular/WiFi interworking in heterogeneous networks to boost throughput, capacity, coverage and quality of experience. However, the continuous use of multiple wireless interfaces will increase the system performance but at the expense of more energy. As a result, there is a need for a dynamic use of multiple interfaces to provide a balance between energy consumption, throughput and user experience. Previous work in this field has considered improving throughput and reducing energy consumption, but did not consider simultaneously quality of experience as perceived by the end user. In this work, we aim at devising real-time traffic splitting strategies between WiFi and cellular networks to maximize user experience, reduce delay, and balance the needed energy consumption. We develop solutions for cellular/WiFi network resource management using Lyapunov drift-plus-penalty optimization approach. We evaluate the proposed approach using parameters determined via experimental measurements from mobile devices, and using our own test bed implementation to provide an evaluation under realistic operation conditions. Results show the performance effectiveness of the proposed traffic splitting approach in terms of throughput, delay, queue stability, energy consumption and quality of user experience by monitoring the frequency and lengths of video stalls.

## 1. Introduction

According to Cisco Visual Networking Index, the mobile traffic will reach 30.6 Exabytes per month by 2020 (Cisco). Video streaming and downloads are expected to consist more than 75% of all consumer Internet traffic in 2020 (Cisco). Due to the large video demands, operators are working towards satisfying application needs in terms of bandwidth and users' expectations. The user's expectations, noted as quality of experience (QoE), is defined by the International Telecommunication Union (ITU) as the overall acceptability of an application or service, as perceived subjectively by the end-user (Recommendation ITU-T P10/G100, 2016).

To meet these tremendous traffic demands, the research community is currently actively involved in the design of the key components that will lead to the development of the 5G cellular technology, with the ability to accommodate massive connections and high loads with ultra-fast speeds taking advantage of coexistence of multiple wireless interfaces (Andrews et al.; 3GPP TR 22.934 version 11.0.0 Release

11; 3GPP TR 36.932 version 12.1.0 Release 12). Heterogeneous networks (HetNets) are currently under intensive academic and industrial research due to their improved coverage and unprecedented capacity gains as compared to conventional single-tier macro networks. The idea behind HetNets is to overlay existing macro cellular networks with additional infrastructure in the form of smaller low-power low-complexity access nodes, such as WiFi hotspots. Substantial gains in throughput and user experience can be provided by utilizing smarter resource coordination among base stations, better transmission strategies and more advanced techniques for efficient resource management (3GPP TS 22.278 version 8.5.0 Release 8; 3GPP TR 36.839 version 11.0.0 Release 11). Resource allocation in HetNets can be divided between approaches focused on utilizing one wireless interface at a time, which is denoted as network selection, and approaches focused on utilizing multiple interfaces simultaneously which is denoted as traffic splitting. Network selection allows moderate enhancements in terms of throughput while keeping energy consumption low. However, traffic splitting achieves higher throughput and quality

\* Corresponding author.

E-mail addresses: [nfa23@aub.edu.lb](mailto:nfa23@aub.edu.lb) (N. Abbas), [hh63@aub.edu.lb](mailto:hh63@aub.edu.lb) (H. Hajj), [zd03@aub.edu.lb](mailto:zd03@aub.edu.lb) (Z. Dawy), [karim.jahed@lau.edu](mailto:karim.jahed@lau.edu) (K. Jahed), [sanaa.sharafeddine@lau.edu.lb](mailto:sanaa.sharafeddine@lau.edu.lb) (S. Sharafeddine).

<http://dx.doi.org/10.1016/j.jnca.2017.01.008>

Received 14 June 2016; Received in revised form 9 December 2016; Accepted 15 January 2017

Available online 24 January 2017

1084-8045/ © 2017 Elsevier Ltd. All rights reserved.

of experience with a trade off cost in terms of energy consumption. Since the energy of the mobile devices is limited and due to the urgent need of satisfying user expectations, a balance between energy consumption, user experience and throughput is needed.

In this work, we focus on real-time traffic splitting decisions in cellular/WiFi HetNets to provide the user with a balance between high quality of experience, low energy consumption and delay while using video on demand streaming applications. We aim at developing an optimized multi-objective traffic splitting solution as a function of the dynamic variation of various system parameters. The proposed approach does not only focus on the energy consumption and throughput, however, it considers user satisfaction as a main objective based on QoE metric that is derived from ITU-T P.1201 standard for video streaming applications. Real-time traffic splitting decisions are performed based on Lyapunov optimization utility functions aiming at achieving high quality end-user experience and minimizing energy consumption while stabilizing network queues. The proposed approach traffic splitting with delay-power-QoE balance (TS-PQ) is user-centric and runs in the background at the user side without any intervention from the network or the server, and without performing any changes to the cellular/WiFi standards. The performance is evaluated using simulations based on parameters determined via experimental measurements on mobile devices for video streaming and using our own test bed implementation under realistic conditions.

This paper is organized as follows. Related work is presented in Section 2. The contributions are presented in Section 3. The potential gains of traffic splitting are presented in Section 4. The system model is presented in Section 5. The proposed traffic splitting approach is detailed in Section 6. Performance results are presented and explained in Section 7. Test bed implementation and results are presented in Section 8. Finally, conclusions are drawn in Section 9.

## 2. Related work

Network selection and traffic splitting in cellular/WiFi HetNets have been widely addressed recently in the literature.

### 2.1. Network selection in heterogeneous networks

The authors in the following research works focused on proposing approaches for the user to select one network in HetNets. The authors in Gustafsson and Jonsson (2003) defined the best network depending on the coverage, cost, bandwidth and application QoS requirements, capacity, as well as personal preferences. The authors in Abbas et al., Abbas and Saade, Naghavi et al. (2016), Hasan et al. (2016), and Liu et al. (2011) used machine learning to solve network selection in HetNets. In Abbas et al., the authors proposed a new learning-based approach based on decision tree for performing 3G/WiFi network selection considering different features such as signal strength, data size, location, and type of application to maximize throughput or maximize energy efficiency or reduce energy consumption based on the context. In Abbas and Saade, a fuzzy logic based network selection approach was proposed considering signal strength, network load and mobile movement speed providing the best throughput. In Naghavi et al. (2016), the authors proposed a Q-learning and reinforcement learning-based algorithm for radio access technologies selection game where the throughput is the main objective function. The authors in Hasan et al. (2016) used particle swarm optimization and a modified version of the genetic algorithm to solve network selection and channel allocation providing users with a target quality of service at a low price, subject to the interference constraints. In Liu et al. (2011), the authors proposed a cooperative vertical handoff decision algorithm based on game theory to achieve the load balancing and meet the quality of service (QoS) requirements of various applications. The authors in Ra et al., presented a network selection approach for energy-delay tradeoff using the Lyapunov optimization framework for video upload con-

sidering queue stability, power consumption and throughput. The authors in El Helou et al. (2015) proposed a network selection scheme with network assistance satisfying operator objectives while meeting user preferences and requirements. In Trestian et al. (2013), the authors proposed a handover scheme based on a utility function providing balance between energy, throughput and cost based on the user preferences. In Singh and Andrews (2014), Tsao et al. and Cai et al. the authors considered traffic offloading in HetNets from the cellular network to WiFi to balance the load, accommodate new requests and reduce network congestion. The authors in Lai et al. (2016) addressed power and admission control in small cells deployment aiming at maximizing user admission, network spectral and energy efficiency while satisfying users minimum rate requirements. In Ma and Ma (2014), the authors proposed a joint network selection and resource allocation for multicast in HetNets aiming at minimizing the overall bandwidth cost. The authors in Dinga et al. (2015) proposed automatic energy efficient application-aware multimedia delivery solutions in cooperative HetNets where a device can download content from the neighboring device with the same interest on the content and providing lower energy consumption.

Despite the enhancements provided by the intelligent network selection approaches, their performance was limited to the selection of one interface, and did not consider the simultaneous use of multiple interfaces. Furthermore, they did not consider user quality of experience.

### 2.2. Traffic splitting in heterogeneous networks

The authors in the following research works considered traffic splitting in HetNets. In Kim et al. (2008), Abbas et al., Kim (2010); Yang et al. and Gelabert et al. (2011), the user receives data over different interfaces consecutively according to a specific ratio. The authors in Kim et al. (2008) aimed to split the traffic periodically according to the ratio of the time required to send the data using one link over the sum of time required to send it via both links. The authors in Abbas et al., presented a multi-objective approach for traffic splitting that splits the data file between WiFi and cellular links capturing tradeoffs between throughput maximization and energy consumption minimization. In Kim (2010), the authors worked on minimizing the time required to send the data over WiMAX and WiFi to determine the traffic splitting ratio. In Yang et al., Yang et al. proposed a traffic splitting approach based on a split ratio dynamically adjusted based on the network channel quality and load to enhance throughput. The authors in Gelabert et al. (2011) presented a multipath packet transmission scheme that reduces packet reordering, improves throughput and maximizes utilization of the links. The faster link is assigned more packets than the slower one. In Luo et al. (2003), Lian et al., Dimatteo et al., Stadler and Pospischil, and Yang et al. (2016), the authors aimed to split the traffic based on the service characteristics. In Luo et al. (2003), the control or important information such as base layer in video are sent over 3G while others are sent via WiFi. The authors in Lian et al., proposed layered video streaming allocation based on traffic characteristics to increase system capacity. The authors in Dimatteo et al., proposed a delay tolerant approach in heterogeneous networks where the user sends a request via 3G to the base station which replies by forwarding the requested content via 3G or WiFi based on links' availability. In Stadler and Pospischil, the I-frames, containing the full information of the video, are sent over 3G with a guaranteed level of quality of service while other frames are sent via WiFi. The authors in Yang et al. (2016) proposed a centralized multi-RAT bandwidth aggregation where LTE and WiFi networks are used to transfer different services simultaneously taking into consideration networks congestion. The authors in Ju et al. (2015) proposed a traffic splitting approach performed at the network level, through jointly optimizing traffic control and radio resource allocation of multiple radio access networks. The authors in Li et al., and Song et al.,

considered uplink traffic splitting and scheduling. The authors in Li et al., proposed a packet scheduling algorithm based on parallel aggregation of radio nodes transmission schemes to improve the delay performance. The authors in Song et al., proposed resource allocation for uplink traffic splitting while considering a particular aspect of user QoE, where the QoE metric focuses on: (1) link reward function reflecting the achieved throughput as quality of service, and (2) resource cost function representing the cost required to use the allocated resources per unit bandwidth. They aimed at illustrating the trade-off between the link throughput and associated cost without considering actual user experience as perceived by the user end.

In summary, previous network selection and traffic splitting approaches considered load balancing, system capacity enhancement, bandwidth allocation, throughput maximization and power consumption reduction. However, they did not consider simultaneously user quality of experience. Some work addressed system performance enhancement, and based their results on simulations or arbitrarily generated values for factors affecting network selection decisions such as signal quality, network load and achievable data rate. Some proposed approaches were not dynamic, the decision is static and made only once. Traffic splitting decisions in some works were based on traffic characteristics, and may require network assistance.

### 3. Contributions

This paper focuses on meeting the real-time demands of video streaming applications where higher bandwidth and QoE are required. This can be achieved by allowing the use of multiple wireless interfaces simultaneously, which would lead to higher energy consumption. Alternatively, we propose a QoE-aware resource management approach providing the user with higher throughput, good quality while keeping low bounds on energy consumption and delay.

The main contributions of this work can be presented as follows:

1. Developing an optimized multi-objective traffic splitting solution that minimizes delay, stabilizes network queue while reducing energy consumption and achieving high quality end-user experience. In contrast to the literature, our approach does not focus only on throughput and energy consumption, but also considers user quality of experience based on ITU-T P.1201 standard to better capture the video quality as perceived by the end user.
2. Providing real-time traffic splitting decisions as a function of the dynamic variation of various system parameters. The formulation customizes the Lyapunov drift plus penalty optimization approach to meet the desired objectives of capturing user satisfaction, achieving a balance between high quality end-user experience, low energy consumption and delay, and allowing the use of multiple interfaces simultaneously to split the traffic at a specific time slot into different links.
3. Allowing the proposed approach to function in the background at the user side without any intervention from the network or the server, and without performing any changes to the cellular/WiFi standards.
4. Evaluating and validating our proposed approach using our own test bed implementation under realistic conditions.

### 4. Motivating the benefits of traffic splitting in HetNets

To demonstrate the potential gains of traffic splitting in heterogeneous networks, a realistic toy example for video on demand transmission is presented in Fig. 1. Simulations are conducted using MATLAB to stream a video using different strategies: (1) WiFi only (WO), (2) cellular only (CO), and (3) using both links simultaneously (TS-S), and their performance in terms of average throughput, total energy consumption, frequency of stalls and length, and satisfaction metric evaluated based on ITU-T P.1201 (2013) QoE metric (15) (details are presented in Section 6.2). Fig. 1 shows the three different

data transmission decision strategies for three consecutive 117 KB data download time slots. The video has a size of 7 MBytes, duration of 60 s, and frame rate of 25 fps. The arrival rate will be 117 KBytes every second. At each time slot of duration 1 s, the mobile device will make decision on the links to use for downloading data based on the selected strategy. If the download rate is less than the video arrival rate, the user will experience buffering events. The video data is not lost and the frames are not skipped. Instead, they are delayed when stalls happen. In the top-most plot, the data is always sent over WiFi. The average throughput over WiFi was higher than the average throughput provided by cellular link while the total energy consumed is the lowest comparing to the other two scenarios where data is sent over cellular only or split between the two links.

In the lowest plot, splitting the traffic by using both links simultaneously provides higher throughput while consuming more energy compared to using WiFi and cellular alone. The results showed that the user experience stalls and freezing frames while watching over WiFi and cellular only, however, no stalls are experienced when traffic splitting is used. The estimated QoE is 5 when using traffic splitting while it is lower 4.3979 and 4.0229 when using WiFi only and cellular only, respectively.

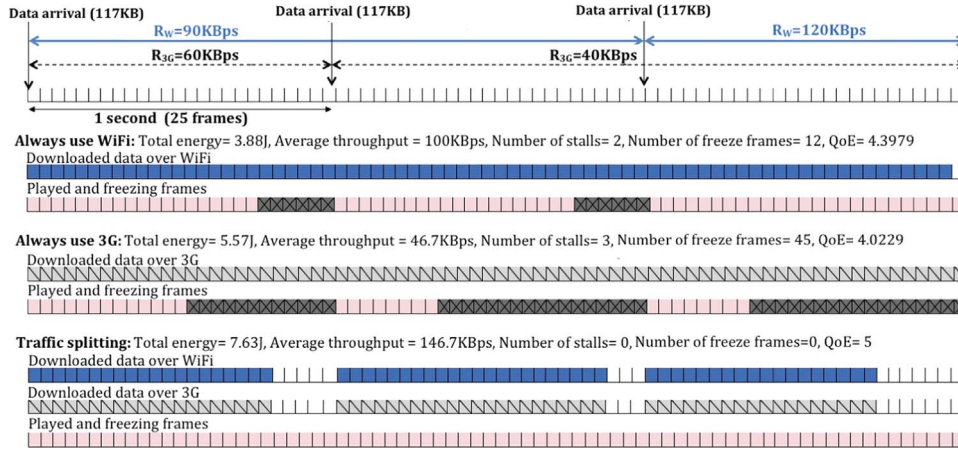
This demonstrates the potential gains of traffic splitting in HetNets and proves its ability to provide high performance in terms of throughput and user satisfaction with a trade of in energy consumption. The results emphasize the need for an optimized decision making approach to achieve target performance tradeoff and motivate our work for proposing better solutions for designing an optimized approach to determine the best cellular/WiFi resource management strategy considering traffic splitting decisions in every time slot. Many questions arise for selecting the best suitable decision at each time slot: (1) What are the available interfaces and transmission strategies? (2) What are the effects of using each strategy on the device power consumption, queue length and user satisfaction? (3) What is the best strategy and the amount of data to be transmitted over each interface that reduces power consumption, queue backlog length, number and length of stalls, and maximizes user QoE? (4) Is it possible to provide a device centric approach performing autonomously without any intervention or change in the standards? (5) What will be the gains under practical implementation and operational conditions?

### 5. System model

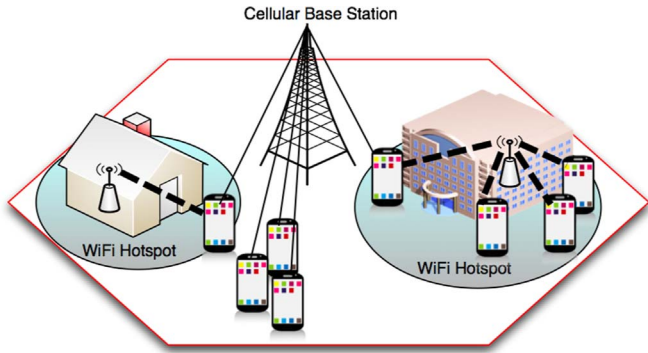
A cellular/WiFi heterogeneous network is typically composed of areas covered via 3G/4G cellular macro-cells and WiFi hotspots (see illustration in Fig. 2). Individual users can take advantage of the coexistence of the different technologies for enhancing their perceived quality of service and experiences. Our proposed method will decide on behalf of the user to use one link or multiple links simultaneously based on system parameters such as throughput, energy consumption, quality of experience and cellular/WiFi links characteristics. Therefore, deciding on the best download strategy needs to be dynamic to provide the user with the best performance at discrete time sample points, represented by time slots. The notion of time slots is introduced to handle discretization of the real-time aspect of the system and handle the queuing theory operation of the proposed approach. It provides practical feasibility to make decisions periodically every time slot based on data collected and previous actions. Accordingly, at every time slot  $t$ , when a mobile device has both cellular and WiFi connections available, the best link for data download or the best split ratio of data over the two links should be determined to optimize performance and provide the best balance between QoE, energy consumption and delay.

In general, the main system parameters are:

- **Time slot duration ( $T_s$ ):**  $T_s$  is the time slot duration, in seconds, representing how often the decision is taken.
- **Resource management solution ( $L[t]$ ):**  $L[t]$  is the possible traffic



**Fig. 1.** Toy example illustrating the benefits of traffic splitting by comparing three scenarios of: (i) always use WiFi, (ii) always use cellular, (iii) traffic splitting while using the two links simultaneously, for downloading three consecutive 117 KB data blocks. The Figures show the performance in terms of average throughput, total energy consumption, frequency of stalls and length (freeze frames are marked with X), and QoE metric evaluated based on (15).



**Fig. 2.** Heterogeneous network formed by cellular macro-cell and WiFi hotspots.

splitting decision at time slot  $t$ , which can be one of the following: (1) WiFi only, (2) cellular only, (3) both links simultaneously, and (4) no transmission.

- **Strategy ( $\ell$ ):** The index  $\ell$  represents one of the possible resource management strategies represented by  $L[t]$ .  $\ell$  represents the index  $W$  when WiFi only is selected,  $C$  when cellular only is selected and  $WC$  when both links are used simultaneously. Based on the selected strategy  $\ell$ , the amount of data to be sent over WiFi and cellular links, respectively during time slot duration  $T_s$ , can be determined.
- **Transmission data rate ( $R_\ell[t]$ ):**  $R_\ell[t]$  represents  $R_W[t]$  the estimated data rate over WiFi link only,  $R_C[t]$  over cellular link only, and  $R_{WC}[t]$  when using both links simultaneously at time slot  $t$ , expressed in bits/second. Note that  $R_{WC}[t] = R_W[t] + R_C[t]$ .
- **Power consumption ( $P_\ell[t]$ ):**  $P_\ell[t]$  represents the estimated power consumed by the mobile device while receiving via WiFi, cellular or both links simultaneously,  $P_W$ ,  $P_C$  and  $P_{WC}$  in Watts, respectively. Note that  $P_{WC}[t] = P_W[t] + P_C[t]$ .
- **Arrival rate ( $A[t]$ ):**  $A[t]$  represents the amount of data, in bits, that arrives to the user's queue at the server from the application layer within time slot  $t$ .
- **Cost ( $C_\ell[t]$ ):**  $C_\ell[t]$  represents the estimated penalty function and cost in terms of power consumption and QoE degradation when choosing resource management strategy  $\ell$  at time slot  $t$ . The selected resource management strategy will determine the cost of the decision at time slot  $t$ .
- **QoE metric ( $\phi_\ell[t]$ ):**  $\phi_\ell[t]$  represents the expected quality of experience and user satisfaction when choosing download strategy  $\ell$  at time slot  $t$ .  $\phi_\ell[t]$  represents  $\phi_W[t]$ ,  $\phi_C[t]$ ,  $\phi_{WC}[t]$  when WiFi only, cellular only, and simultaneous use of both links are selected, respectively.  $\phi_\ell[t]$  is expressed as mean opinion score ranging from 1 to 5.

- **Number of stalls ( $N_\ell[t]$ ):**  $N_\ell[t]$  represents the predicted number of stalls or rebuffering events that the user is expected to experience if resource management strategy  $\ell$  is used at time slot  $t$ . If the transmission rate is less than the required amount of data to be played, the user will experience a re-buffering event during time slot  $t$ ; in this case,  $N_\ell[t] = N[t - 1] + 1$ .
- **Average stalls length ( $L_\ell[t]$ ):**  $L_\ell[t]$  represents the average length of stalls that the user is expected to experience if download strategy  $\ell$  is used at time slot  $t$ .  $L_\ell[t]$  considers the length of all the previous stalls experienced by the user in addition to the expected stall length at time slot  $t$ .

The performance parameters are:

- **Transmission data ( $\mu_\ell[t]$ ):**  $\mu_\ell[t]$  represents the amount of data that has been transmitted in time slot  $t$  over WiFi and cellular links  $\mu_W[t]$  and  $\mu_C[t]$ , respectively, and on both links simultaneously  $\mu_{WC}[t]$ , expressed in bits.
- **Transmission data ( $\mu[t]$ ):**  $\mu[t]$  represents the total amount of data that has been transmitted till time slot  $t$ , expressed in bits.
- **Queue backlog ( $Q[t]$ ):** The queue backlog  $Q[t]$  represents the amount, in bits, of unfinished work as data not being downloaded yet at the beginning of time slot  $t$  and can be expressed as follows:
 
$$Q[t + 1] = Q[t] - \mu_\ell[t] + A[t] \quad (1)$$
- **Video data played ( $Y[t]$ ):**  $Y[t]$  represents the amount of video data played till time slot  $t$ , expressed in bits.
- **Video data downloaded but not yet played ( $D[t]$ ):**  $D[t]$  represents the amount of video data downloaded but not yet played till time slot  $t$ , expressed in bits.  $D[t]$  can be computed as follows:
 
$$D[t] = \max(\mu[t] - Y[t], 0).$$
- **QoE ( $\phi[t]$ ):**  $\phi[t]$  represents the quality of experience metric reflecting the user satisfaction at time slot  $t$ . QoE metric does not capture subjective quality of experience, however,  $\phi[t]$  will reflect the user satisfaction based on the video stalling length and frequency objective measures.
- **Number of stalls ( $N[t]$ ):**  $N[t]$  represents the number of stalls that the user experiences till time slot  $t$ .
- **Stalls length ( $W[t]$ ):**  $W[t]$  represents the length of stalls that the user experiences till time slot  $t$ .
- **Instantaneous throughput ( $\mathcal{J}[t]$ ):**  $\mathcal{J}[t]$  represents the instantaneous download rate obtained at every time slot  $t$ , expressed in bits per second.
- **Energy consumption ( $\mathcal{E}[t]$ ):**  $\mathcal{E}[t]$  represents the energy consumed to download data at every time slot  $t$ , expressed in Joules.

- **Actual streaming time** ( $\mathcal{S}[t]$ ): Due to the channel condition variations, the estimated rate may be different from the actual transmission rate at time slot  $t$ . The data may be downloaded in less time if the actual transmission rate is higher than the estimated.  $\mathcal{S}[t]$  represents the actual amount of time needed to download the video data at every time slot  $t$ , expressed in seconds.

## 6. QoE-aware traffic splitting optimization

This paper presents a QoE-aware resource management approach for video on demand streaming applications. Our main aim is to solve traffic splitting problem capturing the balance between user QoE, delay bounds and energy consumption for video streaming applications. To achieve high quality of experience with our target application of video streaming, we want to minimize, if not eliminate, video stalls for the users. As a result, the goal for QoE is to keep the network queue backlog from building up and causing video stalls and delays. Therefore, we aim to find the best traffic splitting solution at every time slot  $t$  minimizing the delay and stabilizing network queues while reducing the average power consumption and achieving high quality of experience.

The queue length will grow infinitely when the download rate is less than the video arrival rate; the user will then experience stalling events. The queue backlog length is thus directly related to the system parameters and channel quality such as download rates over each interface. Under queue stability, all requested bits are delivered within an acceptable limited delay experienced by the user, such that, all video chunks will be delivered within their playback deadline (Bethanabhotla et al.). To ensure queue stability, decisions need to be made at every time slot  $t$  based on the current queue state and system parameters, to control the change in a function at every step. This process will allow controlling the ending value of the queue backlog size from growing infinitely.

The traffic splitting problem can be formulated as a multi-objective optimization function leading to high QoE with a controlled tradeoff in energy consumption. We use Lyapunov optimization framework from queuing theory, which also provides low computational complexity and enables real-time decisions on network transmission. The Lyapunov optimization guarantees queue stability and achieves near-optimal performance for the chosen optimization objective (Ra et al.; Neely, 2010).

Lyapunov-based utility functions are derived to provide solution for the multi-objective optimization providing a balance between QoE, energy consumption and delay. These utility functions are computed for the set of possible download strategies which are in our case WiFi link alone, cellular link alone, and both links simultaneously. The strategy providing the maximum utility function will be selected for transmission at time slot  $t$ .

In this section, we present: (1) the Lyapunov drift-plus-penalty optimization formulation of the multi-objective function minimizing the Lyapunov drift and cost penalty function, (2) penalty cost function capturing the balance between the power consumption and QoE in addition to queue stability, and (3) the proposed solution and utility functions derivation from the Lyapunov-based multi-objective function.

### 6.1. Lyapunov drift-plus-penalty optimization formulation

The Lyapunov optimization considers controlling and minimizing the change in the user download queue backlog size  $Q[t]$  at every time slot  $t$  resulting in a scheduling algorithm that reduces delay bounds, stabilizes the queue over time and enhances QoE.

**Definition 1.** The Lyapunov function is a scalar measure of the network congestion.

$$\zeta(Q[t]) = \frac{1}{2}(Q[t])^2 \quad (2)$$

**Definition 2.** The Lyapunov drift function  $\Delta(Q[t])$  measures the difference in the Lyapunov function between two consecutive time slots.

$$\Delta(Q[t]) = \mathbb{E}\{\zeta(Q[t+1]) - \zeta(Q[t])|Q[t]\} \quad (3)$$

The function grows large when the system moves towards undesirable states. Therefore, system stability is achieved by taking control actions that minimize the Lyapunov function drift function  $\Delta(Q[t])$ . If control decisions are made every slot  $t$  to greedily minimize  $\Delta(Q[t])$ , then backlogs are consistently pushed towards a lower congestion state, which maintains network stability (Bethanabhotla et al.; Neely, 2010).

Lyapunov drift-plus-penalty method is used as an extension to the base Lyapunov optimization by adding a penalty  $C[t]$  term weighted by a positive coefficient  $V[t]$  that determines the significance of the penalty cost function. In our case, the penalty cost function is expressed in function of power consumption and quality of experience. The Lyapunov drift-plus-penalty approach uses drift steering technique for achieving real-time near-optimal performance-delay tradeoffs for dynamic resource management (Ra et al.; Neely, 2010).

The Lyapunov drift-plus-penalty method is used to capture queue backlog stability in real-time network systems while optimizing the penalty objective metric which allows a balance between delay, QoE and energy in our case. The advantages of using this approach is that (1) it converges to  $[O(1/V), O(V)]$  performance-delay tradeoff; it results in a time average penalty that is within  $O(1/V)$  of optimality, with a corresponding  $O(V)$  tradeoff in average queue size, (2) it provides local optimum guarantees even for non-convex functions, and (3) it provides simple and fast solutions by making decisions based on the current queue states and system parameters without requiring knowledge of the probabilities associated with future random events such as arrival rates and channel variation (Neely, 2010).

The objective function of the Lyapunov drift-plus-penalty approach will be:

$$\underset{\ell \in L[t]}{\operatorname{argmin}} \Delta(Q[t]) + V[t] \cdot \mathbb{E}\{C_\ell[t]|Q[t]\} \quad (4)$$

The objective function aims at (1) minimizing the Lyapunov drift to ensure queue stability and prevent the queue to grow large, and (2) minimizing the cost function at every time slot  $t$  that is in our case expressed in terms of power consumption and quality of experience. The goal is to find the best traffic splitting strategy  $\ell$  that minimizes the Lyapunov drift to ensure minimum delay and queue stability while reducing the transmission cost at every time slot  $t$ . Since the system is dynamic and the channel conditions and rates estimation vary over time, the positive weight  $V[t]$  and cost  $C_\ell[t]$  are time dependent and may vary based on the link selected for transmission at each time slot  $t$ .

For real-time traffic splitting decisions, the multi-objective function will be used to derive utility functions computed at every time slot  $t$  to choose the most efficient traffic splitting strategy  $\ell$  among  $L[t]$  represented by the following: (1) WiFi only (W), (2) cellular only (C), (3) both links simultaneously (WC), and (4) no transmission (0). These utility functions can be obtained by developing the objective function as follows.

Combining (1), (2) and (3), the Lyapunov drift function will be:

$$\Delta(Q[t]) = \frac{1}{2} \mathbb{E}\{(Q[t] - \mu[t] + A[t])^2 - Q[t]^2|Q[t]\} \quad (5)$$

$$\leq \frac{1}{2} \mathbb{E}\{\mu[t]^2 + A[t]^2|Q[t]\} - Q[t] \cdot \mathbb{E}\{\mu[t] - A[t]|Q[t]\} \quad (6)$$

Therefore, the multi-objective function in (4) can be upper bounded as follows:

$$\begin{aligned} \Delta(Q[t]) + V[t] \cdot \mathbb{E}\{C_\ell[t]|Q[t]\} &\leq \frac{1}{2} \mathbb{E}\{\mu_\ell[t]^2 + A[t]^2|Q[t]\} - Q[t] \cdot \mathbb{E}\{\mu_\ell[t] \\ &|Q[t]\} + Q[t] \cdot \mathbb{E}\{A[t]|Q[t]\} + V[t] \cdot \mathbb{E}\{C_\ell[t] \\ &|Q[t]\} \end{aligned} \quad (7)$$

The amount of data  $\mu_\ell[t]$  to be sent over link  $\ell$  can be estimated at the user side based on the transmission data rate  $R_\ell[t]$  representing  $R_W[t]$ ,  $R_C[t]$  or  $R_{WC}[t]$ . Therefore,  $\mu_\ell[t]$  is replaced by its estimate amount of data transfer when using resource management strategy  $\ell$  at time slot  $t$  expressed as follows:  $\mathbb{E}\{\mu_\ell[t]|R_\ell[t]\}$ .

Minimizing our target multi-objective function (4) can thus be achieved by minimizing the upper bound of the objective function in (7). Define  $B[t]$  and  $\lambda$  as follows:

$$B[t] = \frac{1}{2} \mathbb{E}\{\mu_\ell[t]^2 + A[t]^2|Q[t]\} \quad (8)$$

$$\lambda = \mathbb{E}\{A[t]|Q[t]\} = \mathbb{E}\{A[t]\} \quad (9)$$

$B[t]$  and  $\lambda$  are non controllable parameters.  $\lambda$  represents the expected data arrival rate  $A[t]$  defined by the application which is in our case the video arrival rate.  $\lambda$  cannot be controlled by the user and is independent of the current queue back  $Q[t]$ .  $B[t]$  is the sum of the variances of the transmission rate and the arrival rate, which are non controllable parameters. In addition,  $B[t]$  is assumed to be bounded by a fixed value  $B$  (Ra et al., ). Thus, minimizing the Lyapunov drift and penalty will result in minimizing the controllable part of the upper bound in (7)  $-Q[t] \cdot \mathbb{E}\{\mu_\ell[t]|Q[t]\} + V[t] \cdot \mathbb{E}\{C_\ell[t]|Q[t]\}$  which is equivalent to  $-\mathbb{E}\{Q[t] \cdot \mu_\ell[t] - V[t] \cdot C_\ell[t]|Q[t]\}$ .

Using the concept of opportunistically maximizing an expectation, the upper bound expression is maximized by choosing the traffic splitting strategy  $\ell$  every time slot  $t$  as follows (Neely, 2010):

$$\operatorname{argmax}_{\ell \in L[t]} Q[t] \cdot \mathbb{E}\{\mu_\ell[t]|R_\ell[t]\} - V[t] \cdot C_\ell[t] \quad (10)$$

$L[t]$  is the set of possible traffic splitting decision at time slot  $t$ . The solution of the optimization problem is to find the best download strategy  $\ell$  providing the highest performance gains as the best balance between QoE, delay, and energy consumption. The selected strategy  $\ell$  will determine the amount of data  $\mu_\ell[t]$  to be downloaded during time slot  $t$  as  $\mu_W[t]$  if WiFi link only is selected,  $\mu_C[t]$  if cellular link only, and  $\mu_{WC}[t]$  if both links are used simultaneously.

## 6.2. Cost function in terms of power consumption and QoE metric for video streaming

We define the cost  $C_\ell[t]$  of using a transmission link  $\ell$  at time slot  $t$  as a function of power consumption and QoE parameters as follows:

$$C_\ell[t] = f(P_\ell[t], \phi_\ell[t]) = w_1 P_\ell[t] - w_2 \phi_\ell[t] \quad (11)$$

where  $w_1$  and  $w_2$  are positive weights that define the relative importance of the power consumption and QoE metrics.  $P_\ell[t]$  is the power expenditure at time slot  $t$  when using strategy  $\ell$  and  $\phi_\ell[t]$  is a metric representing the quality of experience.

$P_\ell[t]$  represents the average power consumed by the mobile device while receiving data over different interfaces over time slot following time instance  $t$ . When WiFi link is selected, the mobile device will consume  $P_W[t]$  to download data during time slot  $t$ . Similarly, the device will consume  $P_C[t]$  while receiving data over cellular link only. When both links are used simultaneously for data download, the device will use both interfaces in parallel and will consume  $P_{WC}[t] = P_W[t] + P_C[t]$ .

The quality of experience metric  $\phi_\ell[t]$  is based on both objective and subjective psychological measures of using an information and communication technology service (Recommendation ITU-T P10/G100, 2016; Rjaibi et al.; Pinson and Wolf, 2004). Several factors affect quality of the video experienced by the user end such as: (1) network parameters including transmission rate, packet loss, delay, and jitter resulting in stalls and freeze frames (2) application type and char-

acteristics, for instance, video characteristics such as size, frame rate and resolution, and (3) user characteristics such as user's age, and interests. The satisfaction of the user when using the application can be measured by Mean Opinion Score (MOS) (Kilkkki, 2008; Recommendation ITU-T P, 1202, 2012; European Telecommunications Standards Institute). The MOS ranges between 1 (bad) and 5 (excellent) (Recommendation ITU-T P10/G100, 2016).

For video on demand streaming, the video bit rate, frame rate, compression parameters, codec and resolution are non-adaptive and fixed. Video streaming is characterized by playing synchronized media streams in a continuous way while those streams are being downloaded from the application server without having to wait for the entire video to be delivered. Once the playout phase starts, the player fetches video frame from the buffer at a constant speed defined by the video characteristics. When the service transmission rate is less than the arrival data rate, the playing buffer becomes empty. In this case the player pauses and the user will experience stalling and re-buffering events. The video streaming data is not lost, instead the frames are delayed, rather than being skipped. The last received frame freezes and is displayed until the data for the next frame is being downloaded. Therefore, the main distraction for the user and quality satisfaction degradation factors are stalling events, frequency and length. In our model, compression artifacts and losses impairments are not considered since the compression rate is non adaptive and the frame is only displayed when its data is completely downloaded. As a result, common QoE metrics relying on frame by frame and pixel analysis such as peak-signal-to-noise ratio (PSNR) (Recommendation ITU-T J340, 2010), structural similarity (SSIM) (Wang et al., 2004) and video quality metric (VQM) (Pinson and Wolf, 2004) are not suitable for our model and cannot be considered (Szilágyi and Vulkán).

In our work, we estimated the MOS values based on a QoE metric that is derived from standards to better capture the video quality in our optimization and performance assessment. We used re-buffering artifact QoE metric presented in Recommendation ITU-T P.1201 (2013) considering stalling and initial buffering for several reasons: (1) it assesses the effect of perceptual buffering-related indicator to the overall media session quality score, (2) the model predicts mean opinion scores on a 5-point scale as defined in ITU-T P.911, (3) it does not consider the effects of audio level, noise, delay and other impairments related to the payload, and (4) it can be applied for non-adaptive, progressive download type media streaming such as YouTube and operator managed video services over Transmission Control Protocol (TCP) (Recommendation ITU-T P, 1201, 2013). For these reasons, ITU-T P.1201 QoE metric is found to be valid and suitable to our model and consistent with our considerations. In addition, the ITU-T P.1201 QoE metric  $\Phi_\ell[t]$  can be customized to make real time decisions every time slot  $t$ .  $\Phi_\ell[t]$ , presented in Recommendation ITU-T P. 1201 (2013), can be expressed as follows:

$$\Phi_\ell[t] = 5 - \max(\min(\Omega_\ell[t] + \Gamma_\ell), 4), 0) \quad (12)$$

where  $\Omega_\ell[t]$  and  $\Gamma_\ell$  are the expected degradation caused by stalls and initial buffering till time  $t$ , respectively, when using link  $\ell$ . They are defined as follows:

$$\Omega_\ell[t] = \max(\min(s_4 + s_1 \cdot \exp((s_2 \cdot L_\ell[t] + s_3) N_\ell[t]), 4), 0) \quad (13)$$

$$\Gamma_\ell = \begin{cases} \max(\min(d_1 \cdot \log(T_0 + d_2), 4), 0), & \text{if } T_0 \geq 1 - d_2 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

where  $T_0$  is the initial loading time in seconds,  $L_\ell[t]$  is the averaged stalling duration in seconds and  $N_\ell[t]$  is the number of stalling events excluding initial buffering happening till time slot  $t$  when using link  $\ell$ . The coefficients  $s_1, s_2, s_3, s_4, d_1$  and  $d_2$  have the following values  $-1.72, -0.04, -0.36, 1.66, 0.29$  and  $-3.29$ , respectively (Recommendation ITU-T P.1201, 2013).

In our model, we assume the QoE is not affected by the initial buffering degradation; the video is either played without initial

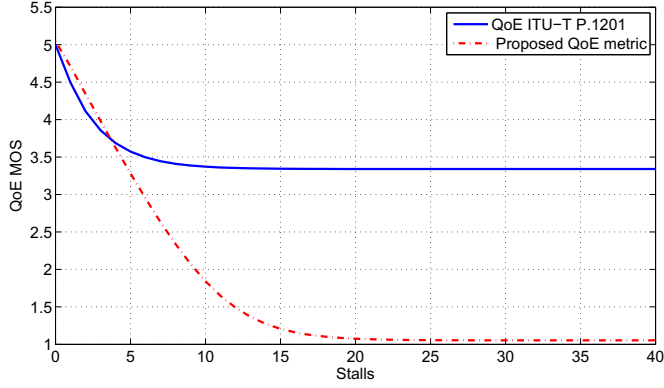


Fig. 3. QoE ITU-T P.1201 metric and proposed QoE metric variations with respect to the number of stalls considering average stalling length of 1 s.

buffering or the initial buffering is lower than  $4.71(1 - d_2)$  seconds. Accordingly, the main degradation of the QoE is caused by the re-buffering and stalling artifacts. To study the effect of the lengths and frequency of occurrence of stalling events on the QoE, Fig. 3 shows the QoE MOS while varying the number of stalls. In the considered case scenario, the average stalling length is considered 1 s. The results show that the QoE varies from 5 (excellent) where the user experience no stalls, to 3.34 where the number of stalls is high. The QoE values are limited to 3.34 since the initial buffering artifacts are not considered. In addition, the re-buffering degradation score  $\Omega_\ell[t]$ , expressed in (13), ranges between 0 and 1.66.

In general, the MOS needs to vary between 1 and 5, however, the QoE metric presented by ITU-T P.1201 is limited to 3.34. For this reason, we utilize an updated metric inspired from the ITU-T P.1201 QoE metric to better emphasize the impact of stalls on MOS. To this end, we use a logarithmic transformation curve fitting to transform the ITU-T P.1201 QoE metric scale from 3.34 to 5 to 1–5. Accordingly, the new metric  $\phi_\ell[t]$  can be obtained from the ITU-T P.1201 QoE metric  $\Phi_\ell[t]$  as follows:

$$\phi_\ell[t] = a \cdot \log(b \cdot \Phi_\ell[t] + c) \quad (15)$$

where  $a$ ,  $b$  and  $c$  are found to be 0.9377, 128.9 and  $-427.6$ , respectively. As shown in Fig. 3, the proposed QoE metric values range between 1 and 5. For instance, when the number of stalls is 9 and 16, the MOS score provided using ITU-T P.1201 QoE metric was 3.3715 and 3.3429, respectively. The video is then considered fair, perceptible but not annoying for both cases. However, when using the proposed QoE metric in (15), the scores were 2.0741 and 1.1590, respectively. The results emphasized the impact of stalls; accordingly, when the user experience 16 stalls, the video will be bad and very annoying, and in case of 9 stalls, the video will be poor and annoying.

The QoE metric  $\phi_\ell[t]$  will be integrated in the Lyapunov drift-plus-penalty objective to capture QoE-related tradeoffs.

### 6.3. Solution approach: traffic splitting with delay-power-QoE balance (TS-PQ)

The Lyapunov drift-plus-penalty problem formulated in (10) can be

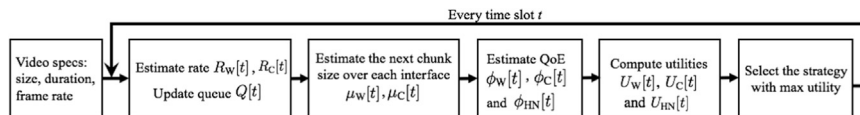


Fig. 4. Diagram representing the proposed TS-PQ approach for near real-time optimized traffic selection.

developed by expressing the cost function in terms of power consumption and QoE. The weighted cost function in (11) will be:

$$V[t] \cdot C_\ell[t] = V[t](w_1 \cdot P_\ell[t] - w_2 \cdot \phi_\ell[t]) \quad (16)$$

$$= V_1[t] \cdot P_\ell[t] - V_2[t] \cdot \phi_\ell[t] \quad (17)$$

Therefore, the objective function will be:

$$\operatorname{argmax}_{\ell \in L[t]} Q[t] \cdot \mathbb{E}\{\mu_\ell[t] | R_\ell[t]\} - V_1[t] \cdot P_\ell[t] + V_2[t] \cdot \phi_\ell[t] \quad (18)$$

where  $V_1[t]$  and  $V_2[t]$  are positive weights that define the relative importance of the power and QoE metrics in the objective function.

In our problem, the decision variable is the strategy  $\ell$  to be selected at time slot  $t$  providing the best performance. One of four possible download strategies can be selected. The decision will be either selecting WiFi link only, cellular link only, both links simultaneously or no transmission at time slot  $t$ . Therefore, the solution of the problem can be achieved by computing the following utility function for the possible resource management strategies considered at every time slot  $t$ . The optimal solution is given by the decision that maximizes the utility per time slot (Neely, 2010).

$$U_\ell[t] = Q[t] \cdot \mathbb{E}\{\mu_\ell[t] | R_\ell[t]\} - V_1[t] \cdot P_\ell[t] + V_2[t] \cdot \phi_\ell[t] \quad (19)$$

$U_\ell[t]$  is computed for different download strategies as follows:

$$U_W[t] = Q[t] \cdot \mathbb{E}\{\mu_W[t] | R_W[t]\} - V_1[t] \cdot P_W[t] + V_2[t] \cdot \phi_W[t] \quad (20)$$

$$U_C[t] = Q[t] \cdot \mathbb{E}\{\mu_C[t] | R_C[t]\} - V_1[t] \cdot P_C[t] + V_2[t] \cdot \phi_C[t] \quad (21)$$

$$U_{WC}[t] = Q[t] \cdot (\mathbb{E}\{\mu_W[t] | R_W[t]\} + \mathbb{E}\{\mu_C[t] | R_C[t]\}) - V_1[t] \cdot (P_W[t] + P_C[t]) + V_2[t] \cdot \phi_{WC}[t] \quad (22)$$

where  $U_W[t]$ ,  $U_C[t]$  and  $U_{WC}[t]$  are the utility functions of using WiFi link alone, cellular link alone, and both links simultaneously, respectively, at time slot  $t$ . The strategy providing the maximum utility function will be selected for transmission at time slot  $t$  (Neely, 2010). However, when the three utility functions are negative, there is no benefit of sending over the links since the device will be consuming more power than benefiting from downloading the data in terms of throughput and QoE; no transmission is recommended in this case. The proposed TS-PQ approach performs in real-time, autonomously at the user end following the steps shown in Fig. 4. The video specifications such as video size, duration and frame rate, are obtained from the server before the start of the video download. The cycle starts with estimating the rate provided by each link and updating the queue based on the arrival rate and data received. Based on the rate estimation, the data for the next time slot is estimated, and is considered for data download over each interface. The quality of experience is estimated over each link using (15) based on the estimated number of stalls and length over each interface. The utility for each strategy is then computed using (20), (21), and (22). The strategy providing the maximum overall utility function is selected at every time slot  $t$  for data download. The cycle is then repeated after updating statistics such as transmission time, data rates, queue length, number and duration of stalls, until video data is downloaded. The steps of the method are shown in Algorithm 1 along with the needed parameters and computations for the proposed approach.

**Algorithm 1.** The proposed multi-objective traffic splitting with delay-power-QoE balance (TS-PQ).

**Input:**

- Video specifications: video size, duration, frame rate, arrival rate  $A[t]$
- Time slot duration:  $T_s$
- Available interfaces (cellular/WiFi) and set of possible download strategies solutions:  $L[t]$
- Power consumption and QoE metric weights:  $V_1[t]$  and  $V_2[t]$ , respectively
- Initial queue backlog size:  $Q[0] = 0$
- Initial video data downloaded:  $\mu[0] = 0$
- Initial video data played:  $Y[0] = 0$
- Initial video data downloaded but not yet played:  $D[0] = 0$
- Initial number of stalls:  $N[0] = 0$
- Initial stalls length:  $W[0] = 0$

**Output:**

The download strategy  $\ell[t] \in L[t] = \{0, W, C, WC\}$

1: **Estimate** WiFi and cellular transmission rates  $R_W[t]$  and  $R_C[t]$ , respectively, by dividing the data downloaded over the time required for download

2: **Estimate** next chunk data size using each strategy  $\ell$ ,

$\mathbb{E}\{\mu_\ell[t]|R_\ell[t]\}$ : (1)  $\mathbb{E}\{\mu_W[t]|R_W[t]\}$ , (2)  $\mathbb{E}\{\mu_C[t]|R_C[t]\}$ , and (3)

$\mathbb{E}\{\mu_{WC}[t]|R_{WC}[t]\}$  as follows:  $\mathbb{E}\{\mu_\ell[t]|R_\ell[t]\} = R_\ell[t] \cdot T_s$

3: **Estimate** quality of experience  $\phi_\ell[t]$  based on (15) for every strategy:  $\phi_W[t]$ ,  $\phi_C[t]$ ,  $\phi_{WC}[t]$  by estimating the number of stalls  $N_\ell[t]$  and average length of stalls  $L_\ell[t]$  as follows:

- **if**  $D[t-1] + \mathbb{E}\{\mu_\ell[t]|R_\ell[t]\} < A[t]$  **then**
- **Update**  $N_\ell[t] = N[t-1] + 1$
- **Estimate** stalling length in time slot  $t$  as

$$\frac{A[t] - (D[t-1] + \mathbb{E}\{\mu_\ell[t]|R_\ell[t]\})}{A[t]} \cdot T_s$$

- **Compute**  $L_\ell[t] = \frac{W[t-1] + \frac{A[t] - (D[t-1] + \mathbb{E}\{\mu_\ell[t]|R_\ell[t]\})}{A[t]} \cdot T_s}{N_\ell[t]}$

• **else**

• **Update**  $N_\ell[t] = N[t-1]$

• **Compute**  $L_\ell[t] = \frac{W[t-1]}{N_\ell[t]}$

• **end if**

4: **Compute** the utility functions  $U_W$ ,  $U_C$  and  $U_{WC}$  as described in (20), (21), and (22)

5: **Select** the strategy  $\ell$  providing the higher utility function. No transmission when utility functions are all negative

6: **Send request** to download the chunk from the server

7: **Compute** the actual data downloaded  $\mu[t]$  and time needed to download the data

8: **Update** queue backlog  $Q[t] = Q[t-1] - \mu[t] + A[t]$

9: **Update** data played  $Y[t] = \min(D[t-1] + \mu[t], A[t])$

10: **Update** data downloaded but not played

$$D[t] = \max(\mu[t] - Y[t], 0)$$

11: **Update** QoE  $\phi[t]$  based on (15), number of stalls  $N[t]$ , stalls length  $W[t]$ , instantaneous throughput  $\mathcal{J}[t]$ , and energy consumption  $\mathcal{E}[t]$

In regards to complexity of our method, the proposed approach is scalable since the traffic splitting decision is made independently at the user end. The method can accommodate for multi-users and multi-interfaces. In the case of multi-user scenario, every user is responsible for her or his own decisions based on its system parameters. If  $n$  interfaces are available for a user,  $(2^n - 1)$  utility functions are computed. In our case study, we consider the coexistence of WiFi and cellular networks. For this case, as previously described, three utility functions are computed at each user end:  $U_W[t]$ ,  $U_C[t]$  and  $U_{WC}[t]$

corresponding to the following transmission strategies: WiFi only, cellular only and both interfaces simultaneously, respectively.

## 7. Results and discussion

To validate the proposed QoE-aware traffic splitting approach under realistic conditions, experimental measurements are used to determine WiFi and cellular key link parameters, such as effective download rate and energy consumed per second during data reception. The obtained link parameter values are then used to quantify and analyze the performance of the proposed QoE-aware Lyapunov-based approach for HetNet resource management.

### 7.1. Experimental energy measurements

Experimental measurements were conducted to capture the effect of signal strength and traffic load on the effective download rate and energy consumption. An Android application was developed on the Samsung Galaxy SIII device to download data of different sizes ranging between 100 KB and 1.2 MB from an HTTP server via WiFi (802.11b) and 3G/4G cellular links. The collection was repeated at different locations where we could have varying traffic loads and signal strengths, which included home (low load) and library (high load) network environments, and different signal strengths such as close and far from the WiFi access point, indoor and outdoor scenarios. In each scenario, the data rate was obtained from the application while power consumption was measured using a data acquisition device (DAQ) from National Instruments monitored via a LABVIEW application. The results showed that the mobile device consumes more energy when receiving on the 3G interface than when receiving on the WiFi interface. The average power consumed was 1.307 Watts for WiFi and 1.859 Watts for 3G.

### 7.2. Performance evaluation

In order to assess the performance effectiveness of the proposed cellular/WiFi traffic splitting approach, we generated results for the following different strategies including state-of-the-art related work from the literature:

1. *WiFi only* (WO): the user downloads data using WiFi link only.
2. *Cellular only* (CO): the user downloads data using WiFi link only.
3. *Maximum rate network selection* (MaxR-NS): the link providing the higher rate is selected in every time slot.
4. *Minimum energy network selection* (MinE-NS): the link proving the lower energy consumption is selected in every time slot.
5. *Stable and adaptive link selection approach* (SALSA): the network providing higher delay-power tradeoff utility function is selected. The authors in Ra et al., presented a network selection approach for energy-delay tradeoff using the Lyapunov optimization framework for video upload. The weight  $V[t]$  of the power metric is considered to be variable over time to adapt the impact of power based on the queue size and delay. We compare our proposed approach to SALSA since it also uses Lyapunov drift-plus-penalty for HetNet resource management, however, our approach is different since it considers traffic splitting in addition to quality of experience.
6. *Traffic splitting using both links simultaneously* (TS-S): the user always uses both links simultaneously to download data.
7. *Traffic splitting with delay-power balance* (TS-P): the user uses the strategy that provides higher utility based on our proposed delay-power tradeoff, with fixed weighing factor  $V[t]$  of the power metric and considering traffic splitting option. Therefore, the cost function in this case only captures power without considering QoE. The objective function (18) is reduced to:

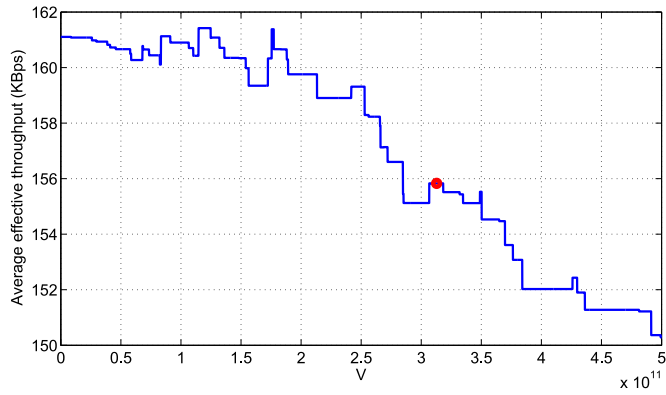


Fig. 5. Average effective throughput in KBps variation with respect to  $V$ . The highlighted value  $V$ , equal to  $3.128 \cdot 10^{11}$ , represents a tradeoff between energy and throughput.

$$\operatorname{argmax}_{\ell \in \mathcal{L}[t]} Q[t] \cdot \mathbb{E}\{\mu_{\ell}[t] | S_{\ell}[t], R_{\ell}[t]\} - V[t] \cdot P_{\ell}[t] \quad (23)$$

where the queue length  $Q[t]$  and  $V[t]$  give weights to the transmission rate and power consumption respectively.

8. *Traffic splitting with delay-power-QoE balance (TS-PQ)*: the user uses the strategy including traffic splitting that provides higher utility based on our proposed QoE-aware resource management approach providing delay-cost tradeoff where cost is a function of both power and QoE with weighing factors  $V_1[t]$  and  $V_2[t]$ , respectively.

### 7.3. Simulations setup

We first evaluated the performance of our proposed approach using simulations conducted using MATLAB to stream a video using different strategies. We used MATLAB for convenience, which allowed us to easily access and modify the physical layer and provided high flexibility in modifying the data requests.

In our model, a mobile device downloads data for video streaming application, where the video specifications, such as video size, duration, and frame rate, are obtained as input from the server before the start of the video download. The chosen video has a size of 7 MBytes, a duration of 60 s, a frame rate of 25 fps, and an arrival rate of 117 KBytes every second. The cellular and WiFi transmission rates were assumed to have exponential distribution with different mean values as presented in the results section below.

As presented in Algorithm 1 and Fig. 4, at each time slot of duration 1 s, the transmission rate, queue size, QoE and power consumption are first estimated. The quality of experience is estimated over each link using Eq. (15) based on the estimated number of stalls and length over each interface. The mobile device makes decision on the link combination that provides the highest utility function, where the options are: (1) WiFi only, (2) cellular only, (3) both links simultaneously, or (4) no transmission. Once data is downloaded based on the selected strategy, the actual parameters are recorded such as queue size, transmission rate, QoE and energy consumption. Recordings become part of the inputs for estimating the parameters of the next time slot. The queue is updated based on the arrival rate and data received. If the transmission rate is lower than the video arrival rate, the mobile device is able to download only a fraction of the requested data. The remaining data that was not downloaded is stored in the queue to be downloaded in the next time slots. If the transmission rate is higher than the video arrival rate, the data is downloaded on time without any delay, stalls, or freeze frames. In this case, the queue is empty with a queue size of zero. The process is then repeated, statistics are updated every time slot until video data is completely downloaded.

### 7.4. Simulations results and analysis

To compare the performance of the various strategies mentioned in Section 7.2, we evaluated the queue size, the average throughput, total energy consumption, delay and QoE for three case scenarios. In the first two case scenarios, symmetric rate for both WiFi and cellular is considered. In the first case, WiFi and cellular transmission rate follow exponential distribution with average rate of 110 KBps. In the second case, WiFi and cellular transmission rate follow exponential distribution with average rate of 450 KBps. In the third scenario, we considered asymmetric rates; we fixed the value of the average WiFi rate of 200 KBps and varied the average cellular rate from 1 KBps to 600 KBps. In this section, we present analysis for the selected power and QoE weights used in our simulations and performance results for the considered scenarios. In addition, we present a study on the duration of time slot and its effect on the performance of the considered approaches.

#### 7.4.1. Study on the power consumption and QoE weights

As presented in (23), when TS-P is used, the queue length  $Q[t]$  and  $V[t]$  give weights to the transmission rate and power consumption, respectively. Accordingly, when the queue length of unfinished work is high, the transmission rate will have more impact than power and the transmission will occur even if the transmission rate is too small. However, when the queue length is low, the power has more impact; thus, the approach can decide to defer transmission to save energy. To give the power and rate similar impact, the value  $V[t]$  should be estimated based on the observed values for queue length, transmission rate, and power consumption through simulations. Figs. 5 and 6 show the average effective throughput and energy consumption variations with respect to  $V[t]$ . In this considered scenario, the WiFi and cellular transmission rate followed an exponential distribution with average of 110 KBps. To obtain a tradeoff between throughput and energy, we selected the value of  $V$  highlighted in red to be  $3.128 \cdot 10^{11}$ . Similarly, the value of  $V[t]$  was selected to be  $6.256 \cdot 10^{11}$  when the average WiFi and cellular transmission rate was set to 450 KBps.

Using TS-PQ,  $V_1[t]$  and  $V_2[t]$  present the weights for power consumption and QoE, respectively. Using the same analysis,  $V_1[t]$  was chosen to be  $3.128 \cdot 10^{11}$  and  $6.256 \cdot 10^{11}$  when the average link transmission rate was set to 110 KBps and 400 KBps, respectively. Similarly,  $V_2[t]$  is fixed to be  $5.52 \cdot 10^{10}$  and  $11.04 \cdot 10^{12}$  when the average link transmission rate was set to 110 KBps and 450 KBps, respectively.

#### 7.4.2. Results for scenario 1: WiFi and cellular average rate of 110 KBps

In the first scenario, we considered a symmetric rate for WiFi and cellular links. The transmission rates are modeled following an exponential distribution with average rate of 110 KBps. Table 1 pre-

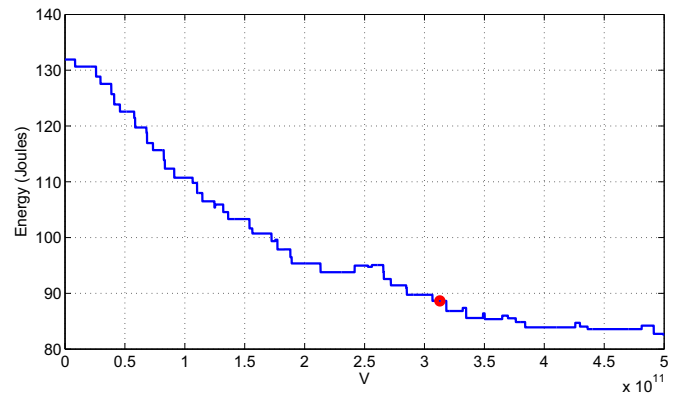


Fig. 6. Energy consumption in Joules variation with respect to  $V$ . The highlighted value  $V$ , equal to  $3.128 \cdot 10^{11}$ , represents a tradeoff between energy and throughput.

**Table 1**  
Simulations results and statistics with  $R_W$  and  $R_C$  average of 110 KBps.

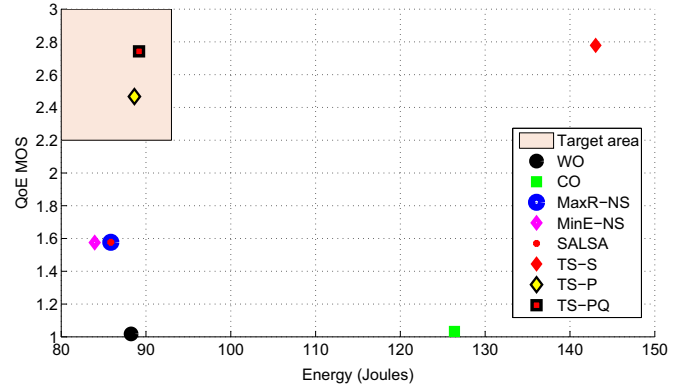
	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	72.4	68.64	64.64	64.76	64.64	62.04	63.2	63.2
Total delay (seconds)	12.4	8.64	4.64	4.76	4.64	2.04	3.2	3.2
Average throughput (KBps)	98.9	104.3	110.8	110.6	110.8	115.4	113.3	113.3
Average effective throughput (KBps)	114.7	108.4	147.1	142.6	147.1	216.2	155.8	155.9
Total energy consumption (Joules)	88.2	126.3	85.8	83.9	85.8	143.1	88.6	90.4
Average queue size (KB)	532.1	394.6	116.6	132.3	116.6	37.3	80.3	76.82
Maximum queue size (KB)	1430.2	972.2	536.4	547.8	536.4	231.8	367.1	367.1
Number of stalls	23	21	12	12	12	7	8	7
Number of freeze frames	310	216	116	119	116	51	80	80
QoE $\phi$ (15)	1.01	1.03	1.57	1.57	1.57	2.77	2.46	2.74

sents the results for scenario 1 and compares the performance of the various approaches presented in Section 7.2 in terms of (1) total streaming time which is the time required to stream and play all the video, (2) delay which is the total duration of stalls, (3) average throughput which is computed by dividing the total amount of data received by the total streaming time, (4) average effective throughput which computed by averaging the throughput achieved in every time slot  $t$ , (5) total energy consumed for downloading the video, (6) average and maximum queue size, (7) number of stalls which represents to rebuffering events experienced by the user, (8) number of freeze frames, and (9) QoE MOS computed based on (15).

The results show low performance for WiFi only, cellular only and network selection strategies. The user will experience more delay, high queue length, freeze frames and stalls when one link is selected instead of traffic splitting on both links. The proposed TS-PQ approach provided the best balance in terms of throughput, queue stability, energy consumption and user satisfaction.

To show performance in terms of queue stability, Fig. 7 shows the queue length variation over time. The queue size varies based on the service transmission rate and the arrival data process. The queue size will go very large if the transmission rate is lower than the arrival data rate. Otherwise, the data will be downloaded on time and the queue size will be 0. The results in Fig. 7 show that traffic splitting approaches TS-S, TS-P and TS-PQ provided higher queue stability, lower queue length and lower delay. TS-S approach showed higher queue stability since the data is always downloaded simultaneously over both links.

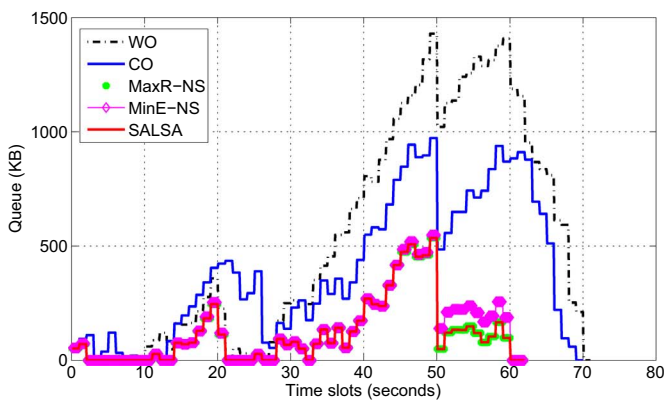
To quantify the tradeoff between QoE and energy consumption, Fig. 8 presents the total energy consumption versus the QoE mean opinion score for each approach. In addition, Fig. 9 presents the tradeoff between total energy consumption in Joules versus the average effective throughput in KBps. The aim is to minimize the energy



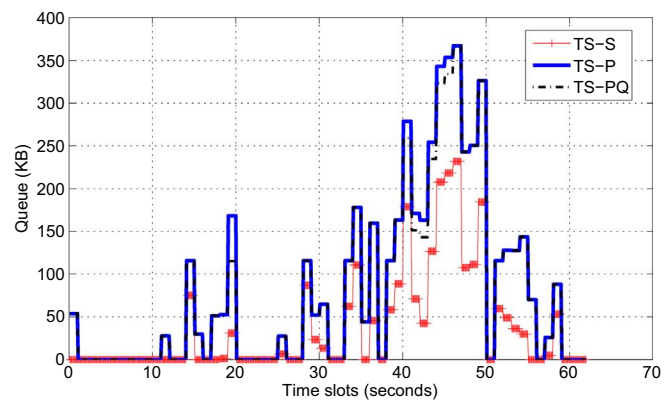
**Fig. 8.** Total energy expenditure (Joules) and QoE MOS for every approach with  $R_W$  and  $R_C$  average of 110 KBps.

expenditure while increasing user quality of experience and throughput. The areas with best performance are highlighted in each figure; these areas lead to high QoE and throughput with reduced energy consumption.

The approaches where traffic splitting is considered provided the best performance in terms of delay, queue length, throughput and user satisfaction while consuming very high energy. TS-P consumed lower energy consumption with a performance reduction in throughput, delay and QoE since it aims to provide a tradeoff between energy-throughput without considering QoE. Our proposed approach TS-PQ aimed to provide a balance between QoE, energy consumption and delay. The results for scenario 1 showed that TS-PQ provided QoE of 2.74 which is higher than the QoE provided by TS-P (2.46) and close to the QoE provided by the TS-S approach (2.77). However, the proposed



(a): Network selection approaches



(b): Traffic splitting approaches

**Fig. 7.** Queue size in KB variation over time with  $R_W$  and  $R_C$  average of 110 KBps.

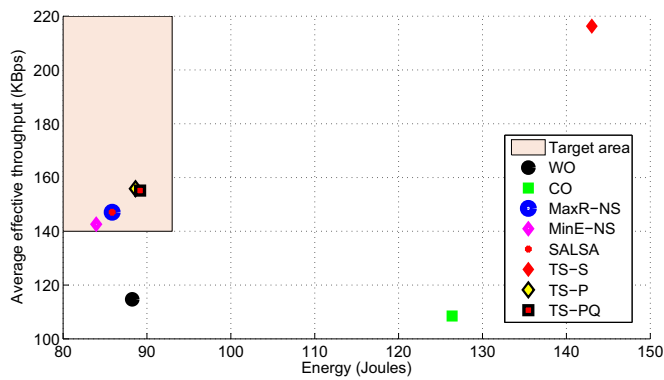


Fig. 9. Total energy expenditure (Joules) and average effective throughput (KBps) for every approach with  $R_W$  and  $R_C$  average of 110 KBps.

approach TS-PQ provided lower energy cost of 89.1 Joules which is 37.7% less than the energy consumed when using both links simultaneously TS-S. The number of freeze frames and stalls were the lowest with 7 stalls and 80 freeze frames which leads to a delay of 3.2 s.

Considering the same scenario 1 where the WiFi and cellular rates are symmetric and modeled following an exponential distribution with average rate of 110 KBps, initial buffering is considered before starting video playback. We assumed the video playback starts after downloading specific amount of data corresponding to 2 s of video streaming. Table 2 presents the performance results for scenario 1 with initial buffering considerations and compares the performance of the various approaches presented in Section 7.2 in terms of (1) total streaming time which is the time required to stream and play all the video, (2) delay which is the total duration of stalls, (3) average throughput which is computed by dividing the total amount of data received by the total streaming time, (4) average effective throughput which computed by averaging the throughput achieved in every time slot  $t$ , (5) total energy consumed for downloading the video, (6) average and maximum queue size, (7) number of stalls which represents to rebuffering events experienced by the user, (8) number of freeze frames, and (9) QoE MOS computed based on Eq. (15). As presented in Table 2, adding initial buffering enhanced the quality of experience of all the compared strategies. The number of stalls is reduced since the initial buffer allowed to download more data before starting the video streaming playback, which enhanced the QoE metric. The network selection strategies were able to provide the user with higher QoE of 2.75 when initial buffering is considered instead of 1.57 without initial buffering. Similarly, the TS-PQ proposed approach was able to provide higher QoE of 3.37. The initial buffering does not have an effect on the transmission parameters and performance in terms of download time, throughput and energy consumption of the algorithms. The mobile is able to download data within the same duration, however, the playback is affected. Initial buffering allows the video playback to start after

Table 2

Simulations results and statistics with  $R_W$  and  $R_C$  average of 110 KBps with initial buffering of 2 s of video data.

	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	72.4	68.64	64.64	64.76	64.64	62.04	63.2	63.2
Total delay (seconds)	12.4	8.64	4.64	4.76	4.64	2.04	3.2	3.2
Average throughput (KBps)	98.9	104.3	110.8	110.6	110.8	115.4	113.3	113.3
Average effective throughput (KBps)	114.7	108.4	147.1	142.6	147.1	216.2	155.8	156.2
Total energy consumption (Joules)	88.2	126.3	85.8	83.9	85.8	143.1	88.6	89.1
Average queue size (KB)	532.1	394.6	116.6	132.3	116.6	37.3	80.3	75.7
Maximum queue size (KB)	1430.2	972.2	536.4	547.8	536.4	231.8	367.1	367.1
Number of stalls	17	15	7	7	7	4	6	5
Number of freeze frames	260	166	66	69	66	26	55	54
QoE $\phi$ (15)	1.1	1.22	2.75	2.75	2.75	3.72	3.06	3.37

downloading 2 s of video, which results in a reduction in the number of freeze frames and stalls duration. This, in turn, directly affects the quality perceived by the end user and enhance the QoE.

### 7.4.3. Results for scenario 2: WiFi and cellular average rate of 450 KBps

In the second scenario, we considered WiFi and cellular links with transmission rates following an exponential distribution with average rate of 450 KBps. Table 3 presents the performance results for scenario 2 and compares the performance of the various approaches and parameters presented in Section 7.2. The results in Table 3 show that traffic splitting approaches provide higher performance with lower delay, freeze frames and stalls. In addition, all the approaches were able to provide better performance when the transmission rates increased from 110 KBps in scenario 1 (see Table 1) to 450 KBps in scenario 2 (see Table 3). For instance, when WiFi is only used, the delay was reduced from 12.4 in scenario 1 where the average rate is 110 KBps to 2.96 s where the average rate is 450 KBps; the number of stalls was reduced from 23 (scenario 1) to 6 (scenario 2). The proposed approach showed an excellent performance in terms of QoE without any stalls or buffering events similar to the performance of using both links simultaneously TS-S while consuming less energy.

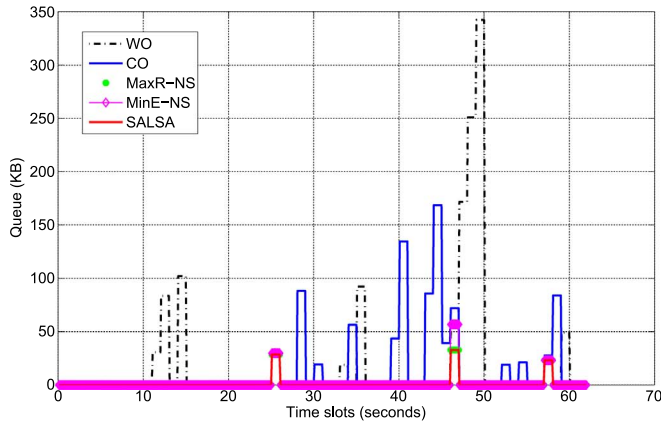
Fig. 10 shows the queue length variation over time when the WiFi and cellular average transmission rates are 450 KBps. Same analysis obtained from Fig. 7 can be drawn, except that when the transmission rate is higher, more data can be downloaded which makes the queue length smaller. The proposed traffic splitting approach TS-PQ and using both links simultaneously TS-S showed an empty queue, which indicates an excellent quality of experience. The provided transmission rate was higher than the arrival rate. All the data is downloaded on time without any stalls or delay which reflects an excellent QoE performance. However, the traffic splitting with delay-power balance TS-P approach causes stalls as shown in Fig. 10(b). These results can also be reflected in Table 3.

Figs. 11 and 12 show QoE-energy consumption and throughput-energy consumption tradeoffs, respectively. The proposed approach was able to perform perfectly without any delay, stalls or freeze frames and provide the user with excellent quality of experience MOS of 5. Similar performance analysis can be remarked when evaluating QoE by metrics presented in Mok et al., and Khan et al. (2010). Using both links simultaneously always (TS-S) approach was able to provide similar performance in terms of quality of experience, however, the proposed approach consumes 26.6% less energy.

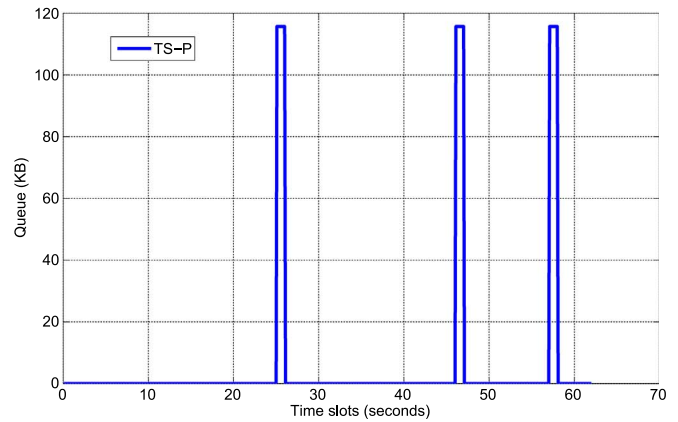
Adding initial buffering enhanced the quality of experience of all the compared strategies. Table 4 presents the results for scenario 2 with initial buffering considerations and compares the performance of the various approaches presented in Section 7.2. We assumed the video playback will start after downloading specific amount of data corresponding to 2 s of video streaming. The compared strategies showed higher user quality of experience. The network selection and traffic splitting strategies were able to provide QoE of 5 without any stalls or buffering events.

**Table 3**  
Simulations results and statistics with  $R_W$  and  $R_C$  average of 450 KBps.

	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	62.96	61.48	60.32	60.52	60.32	60	61	60
Total delay (seconds)	2.96	1.48	0.32	0.52	0.32	0	1	0
Average throughput (KBps)	111.1	113.7	115.9	115.5	115.9	116.5	114.6	116.5
Average effective throughput (KBps)	469.1	448.1	655.1	638.3	655.1	917.3	634.5	642.5
Total energy consumption (Joules)	42.5	56.3	31.8	30.2	31.8	44.2	26.7	32.4
Average queue size (KB)	20.1	14.3	1.3	1.7	1.3	0	5.6	0
Maximum queue size (KB)	342.1	168.6	32.7	56.7	32.7	0	115.6	0
Number of stalls	6	4	2	2	2	0	1	0
Number of freeze frames	74	37	8	13	8	0	25	0
QoE $\phi$ (15)	3.04	3.71	4.40	4.39	4.40	5	4.70	5



(a): Network selection approaches



(b): Traffic splitting approaches

Fig. 10. Queue size in KB variation over time with  $R_W$  and  $R_C$  average of 450 KBps.

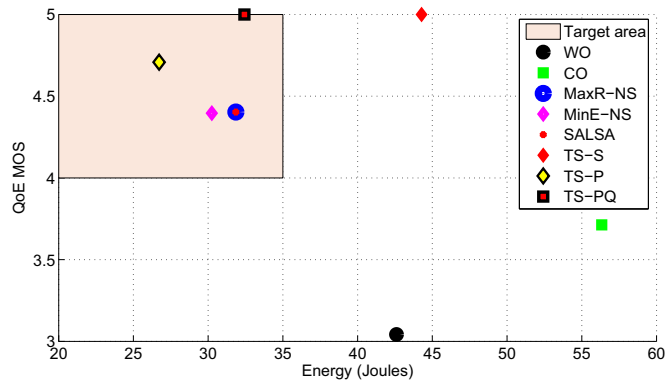


Fig. 11. Total energy expenditure (Joules) and QoE MOS for every approach with  $R_W$  and  $R_C$  average of 450 KBps.

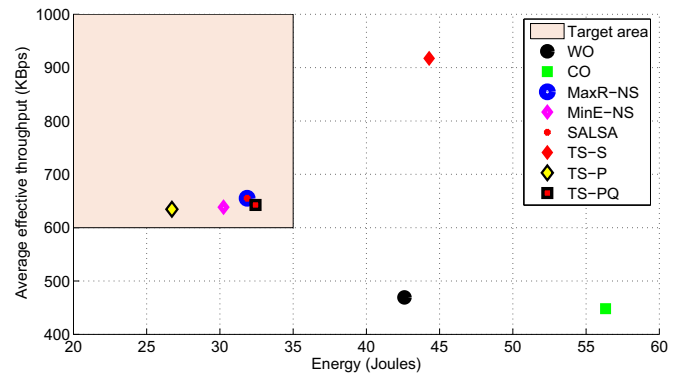


Fig. 12. Total energy expenditure (Joules) and average effective throughput (KBps) for every approach with  $R_W$  and  $R_C$  average of 450 KBps.

7.4.4. Results for Scenario 3: WiFi average rate of 200 KBps with different cellular average rates

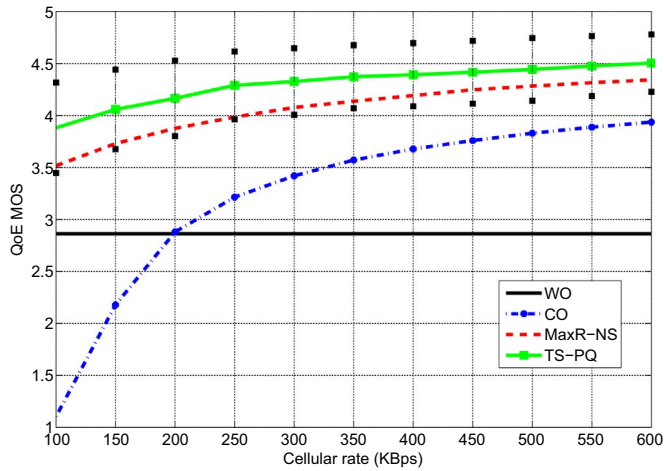
In the previous scenarios, we used symmetrical rate for both WiFi and cellular links. To show the performance of more realistic scenarios, we compared the performance of asymmetrical WiFi and cellular transmission rates considering video streaming for 17 h duration which corresponds to 1020 runs of 60 s videos. The performance metrics such as QoE and energy consumption are measured every 60 s and the overall metrics represent the average over the 1020 runs. In our simulations, the WiFi transmission rate follows an exponential distribution with average of 200 KBps. The cellular rate followed the exponential distribution with average transmission rate varying from 100 KBps to 600 KBps. The performance of the proposed QoE-aware Lyapunov based approach (TS-PQ) was compared to the maximum rate network selection approach (MaxR-NS), WiFi only (WO) and cellular

only (CO) approaches. Figs. 13 and 14 show the performance of the mentioned approaches in terms of QoE and total energy consumption, respectively. Fig. 13 shows the QoE MOS for the different approaches with respect to the variation of average  $R_C$  rate. The ITU-T P.1201 (2013) QoE metric is trained and validated for sequences having duration between 30 and 60 s, where the user does not interact with the player such as stop, play, rewind and fast forward. Accordingly, the QoE for 17 h duration is measured every 60 s and the overall QoE metric represents the average QoE over the whole duration.

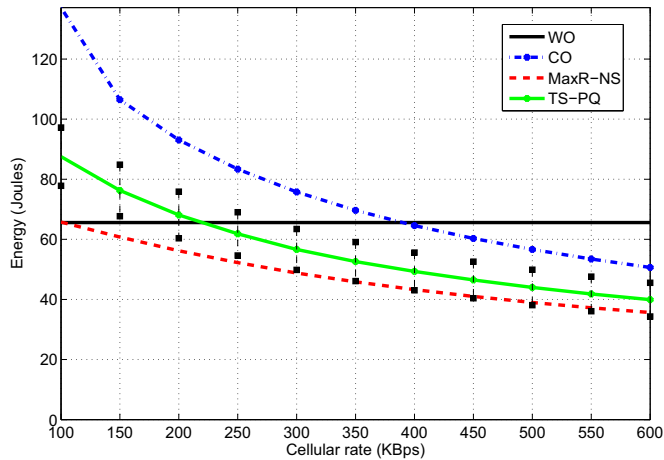
The proposed approach was able to provide a MOS of 4.5 when the average cellular rate was higher than 550 KBps, and a MOS of 3.9 when the cellular rate was 100 KBps. CO performance was enhanced with the increase of average transmission rate to achieve MOS of 2.8, similar to WO, when the WiFi and cellular rates have equal average rates, and a maximum MOS of 3.9 when the average rate is 600 KBps. The QoE

**Table 4**  
Simulations results and statistics with  $R_w$  and  $R_c$  average of 450 KBps with initial buffering of 2 s of video data.

	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	62.96	61.48	61	61	61	61	62	61
Total delay (seconds)	2.96	1.48	1	1	1	1	2	1
Average throughput (KBps)	111.1	113.7	114.6	114.6	114.6	114.6	112.8	114.6
Average effective throughput (KBps)	469.1	448.1	655.1	638.3	655.1	917.3	634.5	642.5
Total energy consumption (Joules)	42.5	56.3	31.8	30.2	31.8	44.2	26.7	32.4
Average queue size (KB)	20.1	14.3	1.3	1.7	1.3	0	5.6	0
Maximum queue size (KB)	342.1	168.6	32.78	56.7	32.7	0	115.6	0
Number of stalls	3	2	0	0	0	0	0	0
Number of freeze frames	49	12	0	0	0	0	0	0
QoE $\phi$ (15)	4.01	4.39	5	5	5	5	5	5



**Fig. 13.** QoE MOS variation with respect to cellular average rate  $R_c$ .



**Fig. 14.** Total energy consumption in Joules with respect to cellular average rate  $R_c$ .

performance of TS-PQ and MaxR-NS approaches also increased with the increase of the average  $R_c$  rate since the mobile device will take advantage of the better channel quality to make better decisions reducing energy while achieving high quality end-user experience.

Fig. 14 shows that MaxR-NS provided lower energy consumption since the approach decides on the link providing higher rate every time slot. When the average  $R_c$  is low, MaxR-NS tends to select the WiFi link more often; MaxR-NS and WO have similar energy consumption. When  $R_w$  and  $R_c$  average rates are equal, WO and CO presents similar QoE, however, energy consumption is higher since mobile device consumes more power while receiving over cellular link.

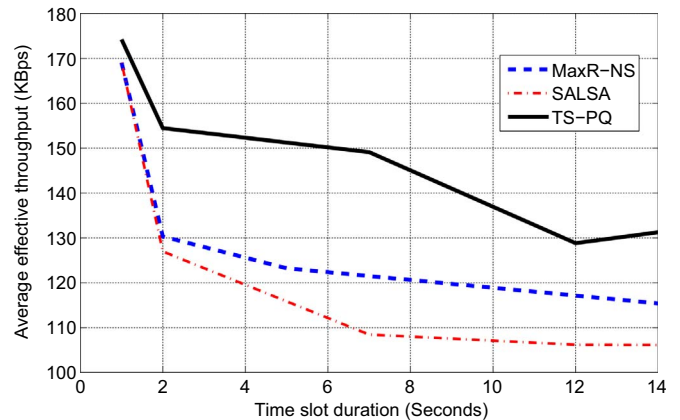
TS-PQ consumed more energy than MaxR-NS and WO when the cellular average rate was small since the proposed approach will tend to

use of both links simultaneously to download more data, reduce delay and maximize QoE. When the average  $R_c$  increases, TS-PQ consumes less energy since it took advantage of the intelligence of the proposed approach; the data in the queue is lower when the transmission rate is higher, thus, the impact of delay is reduced and the power has more impact on the decision.

In addition to the average values for QoE and energy consumption, the variations in each of the measures over 1020 runs (17 h) are captured to show the accuracy of the proposed approach. As presented in Fig. 13, the standard deviation for QoE metric ranged between 0.2 and 0.5 which corresponds to a range of 4.44–14%. The standard deviation for energy consumption ranged between 5.6 and 11.2 Joules (Fig. 14), which corresponds to a range of 10.85–14%. This indicates a relatively acceptable variation level in the performance of each run and validates the reliability and accuracy of our TS-PQ proposed approach.

#### 7.4.5. Study on the time slot duration

In our work, we used the notion of time slot to allow the feasibility of making decisions periodically. To evaluate the effect of the duration of the time slot on the performance of the compared approaches, simulations were conducted for different time slot durations varying from 1 s to 16 s. WiFi and cellular transmission rates follow the exponential distribution with average rate of 110 KBps. The results in Figs. 15, 16 and 17 show the average effective throughput, energy consumption and QoE, respectively, for maximum rate network selection (MaxR-NS), SALSA and the proposed TS-PQ approach. The duration of the time slot will decide how often the transmission decision is made. In general, very small time slot duration (less than 1 s) will not be practical due to the overhead of establishing connection between the server and the mobile device, sending the data request and receiving the data. When the time slot has longer duration, the system will not be able to adapt with fast channel variations, and take advantage of better strategies and transmission decisions, which affects



**Fig. 15.** Average effective throughput in KBps variation with respect to time slot duration  $T_s$ .

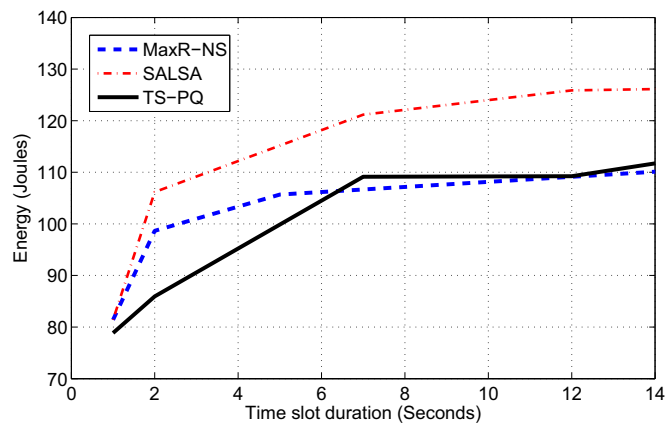


Fig. 16. Total energy consumption in Joules variation with respect to time slot duration  $T_s$ .

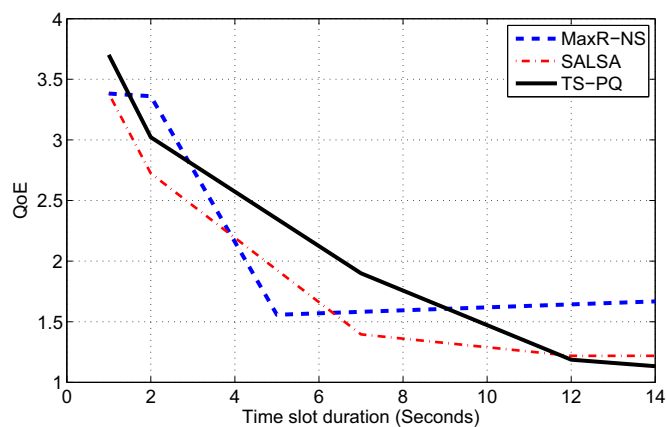


Fig. 17. QoE MOS variation with respect to time slot duration  $T_s$ .

the quality perceived at the user end. Therefore, a large time slot duration may lead to only one decision if the video size is small. As shown in Figs. 15, 16 and 17, the performance decreased with the increase of the time slot duration. Our proposed approach TS-PQ decides every time slot duration on the transmission strategy based on the WiFi and cellular rates estimation. The accuracy of the rates estimation decreases when the time slot duration is longer. This results in inaccurate and inefficient transmission decisions affecting the overall performance of the approach. Accordingly, the solution needs to be real-time, adapt to the fast variation of the channel conditions to provide accurate rate estimation; for these reasons, we chose to use 1 s time slot duration in our work.

### 8. Test bed implementation for cellular/WiFi traffic splitting under realistic operational conditions

In the previous section, the proposed real-time QoE-aware resource management approach for video streaming applications was evaluated using simulations. In this section, the approach is tested under realistic operational conditions for accurate evaluation and validation using our own test bed implementation.

#### 8.1. Test bed setup

In our test bed implementation, we considered the different components of the HetNet architecture; they can be represented by the following three levels: (i) the application service provider layer, (ii) the network wireless interfaces and operator level, and (iii) the user

end as shown in Fig. 18. The test bed is implemented using a modular approach which facilitated enhancements and extensions to implement and test various protocols, design alternatives, or intelligence options.

- **The application service provider layer:** Video streaming applications are deployed on an HTTP server acting as application service provider, and the source for the video files. The server receives data requests from the mobile device over WiFi and cellular networks, with specific data size and offset to be downloaded.
- **The network wireless interfaces and operator level:** In our conducted experiments, we used two wireless interfaces WiFi (802.11b) and 3G cellular networks.
- **The user end:** The client application is implemented using Java programming language on an Android mobile device. The decision on the amount of data to be downloaded over each interface every time slot is made, as described in Section 7.2, based on the one of the following selected transmission strategy: (1) WiFi only, (2) cellular only, (3) both links simultaneously, or (4) no transmission. Accordingly, a specific amount of data is then requested at time slot  $t$  from the server using cellular and WiFi links, respectively. The requests are sent to the HTTP server indicating the offset and the size of the requested data. The data downloaded is reassembled as frames and played on the device.

The main challenge in the test bed implementation is to allow the use of multiple interfaces simultaneously, in addition to implementing the proposed algorithms and test them under real-time conditions. In the current mobile devices, the traffic is offloaded directly to WiFi when WLAN is available. Some new smartphones such as iOS 9 iPhones, Samsung Galaxy S5 and Sony Xperia Z3 introduced auto-switching between WiFi and cellular data networks to avoid poor WiFi connections. However, traffic splitting and the use of multiple networks simultaneously is not yet supported. Our test bed design addresses this issue by supporting both techniques and giving the opportunity to the device to be connected to the best network for data download or using both links simultaneously to achieve performance gains. This can be achieved by allowing parallel transmission using the concept of multi-threading on a rooted Android device.

The user downloads a video with specific size, duration and frame rate. For real-time decision making, at each time slot of duration 1 s, the client application makes decision on the links to use for downloading data based on the selected strategy. If the transmission rate is lower than the video arrival rate, the mobile device downloads only a fraction of the requested video data. The remaining data that was not downloaded is stored in the queue to be downloaded in the next time slots. The video data is not lost, the frames are not skipped, they are only delayed when stalls exist. If the transmission rate is higher than the video arrival rate, the data is downloaded on time without any delay, stalls or freeze frames. In this case, the queue is empty with a queue size of zero. To compare the performance of the various strategies mentioned above, we based our evaluation on performance metrics such as the queue size, the average throughput, total energy consumption, delay and QoE.

Since the rates are varying and dependent on the channel conditions, the comparison of the algorithms may not be accurate. For these reasons, we developed an emulator for transmission rate control. The emulator controls the transmission rate and provides the same rate distribution every time for fair comparison between algorithms. To represent realistic scenarios, the WiFi and cellular networks were monitored, the values for WiFi and cellular transmission rates were recorded as traces. These traces were fed into the emulator; which restricts the downloading rates to the values in the traces. Accordingly, the performance of the algorithms was tested under same conditions for accurate evaluation and comparison.

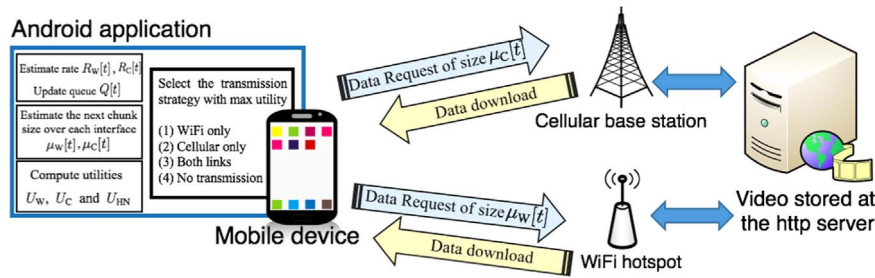


Fig. 18. Test bed implementation composed of three components: (1) mobile device application, (2) network interfaces, and (3) HTTP application server.

Table 5

Test bed results and statistics with  $R_W$  and  $R_C$  average of 42 KBps and 46 KBps, respectively.

	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	169.72	154.96	144.31	169.11	170.11	72.76	117.33	73.26
Total delay (seconds)	109.72	94.92	84.31	109.11	110.11	12.76	57.33	13.26
Average throughput (KBps)	42.21	46.24	49.65	42.37	42.22	100.73	61.06	98.45
Total energy consumption (Joules)	221.82	287.29	214.14	211.02	222.32	224.83	218.51	219.24
Average queue size (KB)	2374.21	2444.66	2367.91	2294.07	2317.01	858.06	1212.97	1106.96
Maximum queue size (KB)	4626.59	4554.67	4569.81	4567.51	4588.73	1524.61	2270.57	7165.45
Number of stalls	173	175	165	174	175	52	99	50
QoE $\phi$ (15)	1	1	1	1	1	1	1	1

Table 6

Test bed results and statistics with  $R_W$  and  $R_C$  average of 130 KBps.

	WO	CO	MaxR-NS	MinE-NS	SALSA	TS-S	TS-P	TS-PQ
Total streaming time (seconds)	62.02	61.87	61.81	61.81	61.97	60	60.58	60
Total delay (seconds)	2.02	1.87	1.81	1.81	1.97	0	0.58	0
Average throughput (KBps)	115.53	115.82	115.92	115.92	115.62	119.42	118.27	119.42
Total energy consumption (Joules)	79.72	112.63	112.03	79.27	80.31	95.92	81.08	84.51
Average queue size (KB)	152.14	137.60	132.00	125.37	147.84	0	12.96	0
Maximum queue size (KB)	241.66	223.08	216.58	215.68	235.84	0	69.82	0
Number of stalls	11	9	8	10	11	0	4	0
QoE $\phi$ (15)	1.80	2.25	2.51	2.01	1.80	5	3.74	5

### 8.2. Test bed experimental results

In our considered scenario, the video has a size of 7 MBytes, duration of 60 s, and frame rate of 25 fps. The arrival rate will be 117 Kbytes every second. The WiFi and cellular transmission rates traces collected in Beirut Lebanon,  $R_W$  and  $R_C$ , presented an average of 42 KBps and 46 KBps, respectively. Using the same analysis presented in Section 7.4,  $V_1[t]$  and  $V_2[t]$  were chosen to be  $1.25 \cdot 10^{11}$  and  $3 \cdot 10^{10}$ , respectively. The results for the different approaches are presented in Table 5. In addition, we considered different scenario with  $R_W$  and  $R_C$  average rates of 130 KBps.  $V_1[t]$  and  $V_2[t]$  chosen to be  $2.5 \cdot 10^{10}$  and  $9 \cdot 10^{12}$ , respectively. The results are shown in Table 6.

Similar to the simulations analysis, the results showed that our proposed approach was able to provide the best balance between QoE, delay and energy consumption. It provided high user satisfaction with an acceptable increase in energy consumption. Note that the number of stalls is high in Table 5 since the WiFi and cellular rates were relatively low in the traces, which led to a poor QoE for all the compared algorithms.

The test bed results presented in Table 6 are highly correlated to the simulations results presented in Tables 1 and Tables 3. Similar analysis presented in Sections 7.4.2 and 7.4.3 can be obtained when comparing the implemented approaches. The results show low performance for WiFi only, cellular only and network selection strategies. The user experienced more delay, freeze frames and stalls when one link is selected. Comparing the approaches considering traffic splitting, TS-P provided the lower energy consumption with a tradeoff cost in terms of

QoE. The proposed TS-PQ approach provided a very high quality of experience similar to TS-S when both links are used simultaneously with 11.8% lower energy consumption. This proves the effectiveness of the proposed approach and demonstrates the feasibility of achieving performance gains in practice using standard mobile devices.

### 8.3. Validation: test bed versus simulation results

To validate our simulation results, we conducted scenarios where data rates are similar to those used in the simulation results (Sections 7.4.2 and 7.4.3). Table 7 presents test bed results for WiFi and cellular average rates of 110 KBps and 450 Kbps using the following approaches: (1) WiFi only (WO), (2) Cellular only (CO), (3) Maximum rate network selection (MaxR-NS), and (4) our proposed traffic splitting approach with delay-power-QoE balance (TS-PQ).

The obtained test bed results are similar to the simulation results presented in Tables 1, 3. For instance, the simulated results for TS-PQ approach showed a total streaming time of 63.2 s with 7 stalls and a QoE of 2.74 when the average rate is 110 KBps (see Table 1). The test bed results for the same scenario showed a total streaming time of 62.85 s with 6 stalls and a QoE of 3.04 (see Table 7). TS-PQ led to high performance gains in both simulations and test bed results for average rates of 450 KBps. The total streaming time is 60 s without any stalls providing an excellent QoE (see Tables 1, 7). The results presented in Table 7 are also coherent with the test bed results presented in Tables 5, 6. Our proposed approach is shown to achieve enhanced performance and a balance between delay, energy consumption and QoE.

**Table 7**

Test bed results and statistics with  $R_W$  and  $R_C$  average of 110 Kbps and 450 Kbps.

	$R_W$ and $R_C$ average 110 Kbps				$R_W$ and $R_C$ average 450 Kbps			
	WO	CO	MaxR-NS	TS-PQ	WO	CO	MaxR-NS	TS-PQ
Total streaming time (seconds)	72.57	68.60	66.92	62.85	62.49	61.89	60.71	60
Total delay (seconds)	12.57	8.60	6.92	2.85	2.49	1.89	0.71	0
Average throughput (KBps)	98.7	104.4	107.1	114.0	114.66	115.78	118.03	119.42
Total energy consumption (Joules)	94.85	125.35	101.55	92.61	75.48	114.03	109.56	72.65
Average queue size (KB)	197.41	241.15	148.78	23.39	86.62	8.24	2.84	0
Maximum queue size (KB)	817.61	851.00	515.83	300.04	297.44	145.79	68.49	0
Number of stalls	22	22	15	6	6	3	2	0
QoE $\phi$ (15)	1.02	1.02	1.22	3.05	3.06	4.02	4.39	5.00

**9. Conclusions and future works**

This paper provided a solution for real-time traffic splitting across cellular and WiFi heterogeneous networks that provides improved QoE while reducing energy consumption and delay. The solution is based on a Lyapunov drift-plus-penalty formulation. The performance of the proposed approach was evaluated using both simulations and our own test bed implementation under realistic operational conditions using video on demand streaming applications. Results for various scenarios demonstrated favorable performance for the proposed traffic splitting approach.

The proposed solution can be extended to handle other types of applications that may exhibit different experience for the user such as live video or large file downloads. In these cases, the QoE metrics need to be adjusted for the specific experience such as data loss or total delays. Additionally, it can be extended towards a more optimized solution by predicting downstream performance in future time slots, capturing the need and cost associated with data re-transmission to recover from losses, or accounting for video multicasting in multiuser scenarios taking into account resource limitations.

**Acknowledgements**

This work was made possible by NPRP [grant 7-1529-2-555] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

**References**

3GPP TR 22.934 version 11.0.0 Release 11, Feasibility Study on 3GPP System to Wireless Local Area Network (WLAN) Interworking.  
 3GPP TR 36.839 version 11.0.0 Release 11, Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks.  
 3GPP TR 36.932 version 12.1.0 Release 12, LTE; Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN.  
 3GPP TS 22.278 version 8.5.0 Release 8, Service requirements for the Evolved Packet System (EPS).  
 Abbas, N., Saade, J.J. A fuzzy logic based approach for network selection in WLAN/3G heterogeneous network. In: Proceedings of the IEEE Consumer Communications and Networking Conference.  
 Abbas, N., Taleb, S., Hajj, H., Dawy, Z. A learning-based approach for network selection in WLAN/3G heterogeneous network. In: Proceedings of the Third International Conference on Communications and Information Technology.  
 Abbas, N., Dawy, Z., Hajj, H., Sharafeddine, S. Energy-throughput tradeoffs in cellular/WiFi heterogeneous networks with traffic splitting. In: Proceedings of the IEEE Wireless Communications and Networking Conference.  
 Andrews, J.G., Buzzi, S., Choi, W., Hanly, S., Lozano, A., Soong, A.C.K., Zhang, J.C. What will 5G be?, IEEE Journal on Selected Areas in Communications, Special Issue on 5G Communication Systems arXiv:1405.2957v1.  
 Bethanabhotla, D., Caire, G., Neely, M.J. Adaptive video streaming for wireless networks with multiple users and helpers. arXiv preprint arXiv:1304.8083v2.  
 Cai, S., Duan, L., Wang, J., Zhou, S., Zhang, R. Incentive mechanism design for delayed WiFi offloading. In: Proceedings of the IEEE International Conference on Communications.  
 Cisco, Cisco Visual Networking Index: Forecast and Methodology, 2015–2020, White Paper, Cisco.  
 Dimatteo, S., Hui, P., Han, B., Li, V.O.K. Cellular traffic offloading through WiFi

networks. In: Proceedings of the 8th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS).  
 Dinga, R., Hava Munteanb, C., Munteana, G.-M., 2015. Energy-efficient device-differentiated cooperative adaptive multimedia delivery solution in wireless networks. J. Netw. Comput. Appl. 58, 194–207.  
 El Helou, M., Ibrahim, M., Lahoud, S., Khawam, K., Mezher, D., Cousin, B., 2015. A network-assisted approach for RAT selection in heterogeneous cellular networks. IEEE J. Sel. Areas Commun. 33 (6), 1055–1067.  
 European Telecommunications Standards Institute, ETSI TR 102 714 V1.1.1, Speech and Multimedia Transmission Quality (STQ); Multimedia Quality Measurement; End-to-End Quality Measurement Framework.  
 Gelabert, X., Sallent, O., Perez-Romero, J., Agusti, R., 2011. Performance evaluation of radio access selection strategies in constrained multi-access/multi-service wireless networks. Comput. Netw. 55 (January (1)), 173–192.  
 Gustafsson, E., Jonsson, A., 2003. Always best connected. IEEE Wirel. Commun. 10 (February (1)), 49–55.  
 Hasan, N., Ejaz, W., Ejaz, N., Kim, H.S., Anpalagan, A., Jo, M., 2016. Network selection and channel allocation for spectrum sharing in 5G heterogeneous networks. IEEE Access PP (99), 1–11. <http://dx.doi.org/10.1109/ACCESS.2016.2533394>.  
 Ju, H., Liang, B., Li, J., Long, Y., Yang, X., 2015. Adaptive cross-network cross-layer design in heterogeneous wireless networks. IEEE Trans. Wirel. Commun. 14 (2), 655–669.  
 Khan, A., Sun, L., Jammeh, E., Ifeachor, E., 2010. Quality of experience-driven adaptation scheme for video applications over wireless networks. IET Commun. Spec. Issue Video Commun. Wirel. Netw. 4 (July (11)), 1337–1347.  
 Kilkki, K., 2008. Quality of experience in communications ecosystem. J. Univers. Comput. Sci. 14 (March (5)), 615–624.  
 Kim, J.-O., Ueda, T., Obana, S., 2008. MAC-level measurement based traffic distribution over IEEE 802.11 multi-radio networks. IEEE Trans. Consum. Electron. 54 (August (3)), 1185–1191.  
 Kim, J.-O., 2010. Feedback-based traffic splitting for wireless terminals with multi-radio devices. IEEE Trans. Consum. Electron. 56 (May (2)), 476–482.  
 Lai, W.-S., Chang, T.-H., Lee, T.-S., 2016. Joint power and admission control for spectral and energy efficiency maximization in heterogeneous OFDMA networks. IEEE Trans. Wirel. Commun. PP (99), 1–16. <http://dx.doi.org/10.1109/TWC.2016.2522958>.  
 Li, J., Zheng, J., Liu, Q., Yang, X. Delay performance optimization of multiaccess for uplink in heterogeneous networks. In: Proceedings of the 79th IEEE Vehicular Technology Conference.  
 Lian, H., Yan, X., Weng, L., Feng, Z., Zhang, Q., Zhang, P. Efficient traffic allocation scheme for multi-flow distribution in heterogeneous networks. In: Proceedings of IEEE Globecom Workshops.  
 Liu, X., Fang, X., Chen, X., Peng, X., 2011. A bidding model and cooperative game-based vertical handoff decision algorithm. J. Netw. Comput. Appl. 34, 1263–1271.  
 Luo, J., Mukerjee, R., Dillinger, M., Mohyeldin, E., Schulz, E., 2003. Investigation of radio resource scheduling in WLANs coupled with 3G cellular network. IEEE Commun. Mag. 41 (June (6)), 108–115.  
 Ma, D., Ma, M., 2014. Network selection and resource allocation for multicast in HetNets. J. Netw. Comput. Appl. 43, 17–26.  
 Mok, R.K.P., Chan, E.W.W., Chang, R.K.C. Measuring the quality of experience of HTTP video streaming. In: Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management.  
 Naghavi, P., Hamed Rastegar, S., Shah-Mansouri, V., Kebriaei, H., 2016. Learning RAT selection game in 5G heterogeneous networks. IEEE Wirel. Commun. Lett. 5 (1), 52–55.  
 Neely, M.J., 2010. Stochastic Network Optimization with Application to Communication and Queueing Systems. Morgan & Claypool Publishers series, SYNTHESIS LECTURES ON COMMUNICATION NETWORKS, USA.  
 Pinson, M.H., Wolf, S., 2004. A new standardized method for objectively measuring video quality. IEEE Trans. Broadcast. 50 (3), 312–322.  
 Ra, M.-R., Paek, J., Sharma, A.B., Govindan, R., Krieger, M.H., Neely, M.J. Energy-delay tradeoffs in smartphone applications. In: Proceedings of the Eight International Conference on Mobile Systems, Applications, and Services.  
 [ITU-T J.340] Recommendation ITU-T J.340 (2010). Reference Algorithm for Computing Peak Signal to Noise Ratio of a Processed Video Sequence with Compensation for Constant Spatial Shifts, Constant Temporal Shift, and Constant Luminance Gain and Offset.  
 [ITU-T P.10/G.100] Recommendation ITU-T P.10/G.100 (2016). Amendment 1: New Appendix I - Definition of Quality of Experience (QoE).

- [ITU-T P.1201] Recommendation ITU-T P.1201 (2013). Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality – Amendment 2: New Appendix III – Use of ITU-T P.1201 for Non-Adaptive, Progressive Download Type Media Streaming.
- [ITU-T P.1202] Recommendation ITU-T P.1202 (2012). Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality.
- Rjaibi, N., Ben Arfa Rabai, L., Limam, M. Modeling the prediction of student's satisfaction in face to face learning: an empirical investigation. In: Proceedings of the International Conference on Education and e-Learning Innovations.
- Singh, S., Andrews, J.G., 2014. Joint resource partitioning and offloading in heterogeneous cellular networks. *IEEE Trans. Wirel. Commun.* 13 (2), 888–901.
- Song, Y., Han, Y., Choi, Y. Radio resource management based on qoe-aware model for uplink multi-radio access in heterogeneous networks. In: Proceedings of the 79th IEEE Vehicular Technology Conference.
- Stadler, J., Pospischil, G. Simultaneous usage of WLAN and UTRAN for improved multimedia and data applications. In: Proceedings of the 11th International Telecommunications Network Strategy and Planning Symposium.
- Szilágyi, P., Vulkán, C. Network side lightweight and scalable YouTube QoE estimation. In: Proceedings of the IEEE International Conference on Communications.
- Trestian, R., Member, Ormond, O., Muntean, G.-M., 2013. Energy-quality-cost tradeoff in a multimedia-based heterogeneous wireless network environment. *IEEE Trans. Broadcast.* 59 (2), 340–357.
- Tsao, S.-L., Wang, Ch.-W., Lin, Y.-C., Cheng, R.-G. A dynamic load-balancing scheme for heterogeneous wireless networks. In: Proceedings of the IEEE Wireless Communications and Networking Conference.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13 (4), 600–612.
- Yang, R., Chang, Y., Sun, J., Yang, D. Traffic split scheme based on common radio resource management in an integrated LTE and HSDPA networks. In: Proceedings of Vehicular Technology Conference.
- Yang, S.-N., Ho, Sh.-W., Lin, Y.-B., Gan, C.-H., 2016. A multi-RAT bandwidth aggregation mechanism with software-defined networking. *J. Netw. Comput. Appl.* 61, 189–198.