Middle-East Journal of Scientific Research 25 (2): 274-281, 2017
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2017.274.281

Energy-Efficient QoE-Aware Video Adaptation and Resource Allocation for Video Streaming

[1]R. Jayavadivel and [2]J. Sundararajan

[1]Department of Information Technology, Paavai Engineering College,
Pachal, Namakkal, Tamilnadu, India 637 018
[2]Principal, Pavai College of Technology, Pachal, Namakkal, Tamilnadu, India 637 018

Abstract: In this paper, QoE-aware resource allocation and video adaptation method is proposed for energy efficient video streaming over OFDMA downlink. The proposed scheme for adaptation drops the packets selectively and that produces lower bitrate in video streaming services under delay and QoE constraints. This leads to load reduction and increased bandwidth capacity in video streaming wireless networks. The main aim of the proposed resource allocation strategy minimizes the transmission power that considers the delay constraints of identified video streams during adaptation phase in video streams. Simulation results shown considerable performance improvement in reducing end-to-end delay and energy efficiency by satisfying the requirements of QoE.

Key words: Video adaptation · QoE mapping · Dynamic Adaptive Streaming in HTTP (DASH) · Energy Efficiency · Resource Allocation

INTRODUCTION

Due to rapid advancement in embracing the tablets, smart phones, etc. has continuously increased the traffic in mobile video streaming services. Such an activity over time insisted the mobile operators for redesigning the wireless networks and that should support more concurrent video streaming. During this time, the several delay constraints and guaranteed QoS levels for each individual users is maintained. The provisions over varied video bitrate sources for an individual video using adaptive bitrate streaming (ABR) probably increases the capacity of the network. This helps in serving the additional concurrent video requests [1]. Moreover, various researches had offered caching in video services that maximizes the capacity of video in wireless streaming networks and further enhances the observed QoE [2-4].

Depending on ABR streaming, the individual video is split into many chunks with varied bitrates. Consequently, in order to serve the whole video from the cache, entire rate variants are cached. Nevertheless, this considerably increases storage requirements and backhaul requirements, since; the video gets encoded into 40+ different variants. This meets the network conditions and device heterogeneity [5]. Moreover, the available rate of transmission in network channels is considered time-variant and hence, it is very hard to predict the rate. Henceforth, the nominated bit stream, which is transmitted by the server, does not match the transmission features of the users [3]. Instead, the video with best quality is cached and the resource that is processed us allowed in performing the transrating [3]. Though, still it consumes the resources and computation that encodes the quality videos to be converted to various bitrates for storing the bitrates in an encoded stream in a real-time environment [4].

The backhaul and storage constraints in radio access network (RAN) are reduced by caching the varied bitrates of a video while downloading. The architecture of which is shown in Figure 1. Further, the RAN is enhanced using queued video adaptation module.

The module adopts the resource allocation strategy and the delay constrained DASH system with active queue management trans-rates the video stream from its lower bitrate and hence leading it to an energy-efficient resource allocation. Here in proposed module, the packet

Correspondin Author: R. Jayavadivel, Department of Information Technology, Paavai Engineering College, Pachal, Namakkal, Tamilnadu, India - 637 018.

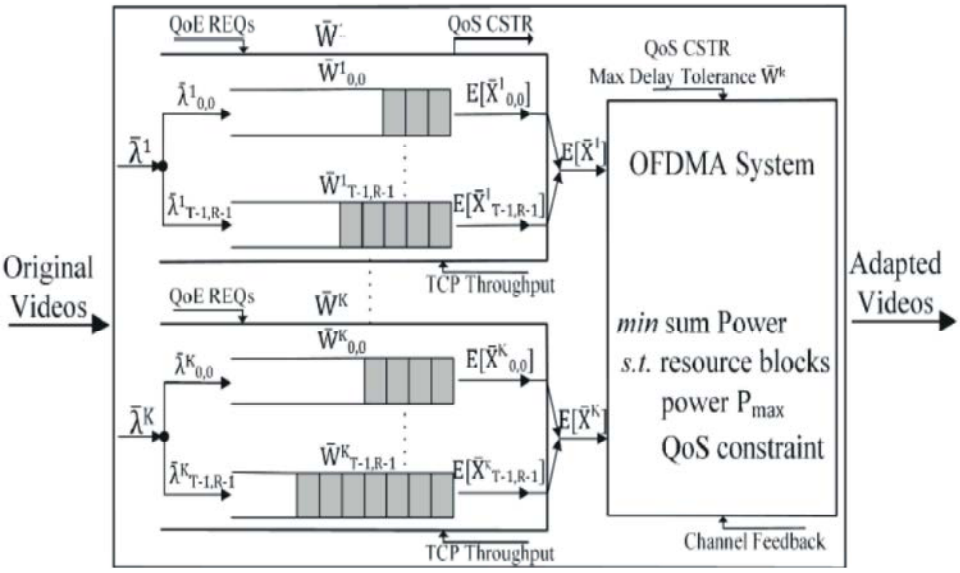

Fig. 1: Proposed Architecture

is dropped, which possess nominal negative impact on QoE of the users to provide QoE satisfaction to a considerable extent. This reduces further the delay and network load with an increase in the network capacity for serving the concurrent additional streams.

System Model: Consider a LTE downlink networks with a multiple user – single cell scenario, as shown in Figure 1. The downlink system holds K video streams/users, which is indexed using the □ sets ≙{1, …, k, …, K}, resource sharing blocks, L, which is indexed using L sets ≙{1, …, l, …, L}. The channel is considered as a frequency-selective wireless channel with Rayleigh fading and within individual resource block flat fading is used in downlink mobile streams. The H.264/DASH streams, k holds the temporal layers, T, which is indexed by the _ sets ≙{1, …, t, …, T^k-1} and the quality layer, R, which is indexed by the R sets ≙{1, …, r, …, R^k-1}.

Hence, a statistical queuing model is deployed for expressing the limitation in streaming delay using corresponding constraints on cross-layer. Thus, similar to [6], here, the arrival of video stream packets for each individual user k with a buffer q^k follows strictly the Poisson arrival process. In q^k, the video streaming system arranges the streamed packets over the quality layer r in a temporal layer t that holds k sequences. The virtual queue (VQ) with k sequences $q^k_{r,t}$ follows the $M/G/1$ queuing [7]. The rates of arrival in $M/G/1$ are hence declared as Poisson processes that are exceedingly suitable for modeling video streaming [8] traffic over DASH system. In addition to this, the service time follows statistical distribution, since; fading channel makes hard the service process modeling [6].

From the Figure 1, the parameters q^k possess a featured tuple of $\left[\bar{\lambda}^k, E\left[X^k\right], \bar{W}^k\right]$ and $q^k_{r,t}$ holds the featured tuple of $\left[\bar{\lambda}^k_{r,t}, E\left[X^k_{r,t}\right], \bar{W}^k_{r,t}\right]$

where, $\bar{\lambda}^k$, $\bar{\lambda}^k_{r,t}$ is the arrival rate of q^k and $q^k_{r,t}$ $E[X^k]$, $E[X^k_{r,t}]$ is the service time of q^k and $q^k_{r,t}$ $\bar{W}^k$, $\bar{W}^k_{r,t}$ is the waiting time of q^k and $q^k_{r,t}$.

The delay in queuing and network congestion is reduced from VQs using packet dropping, which increases the network capacity. The dropping packets, however, reduces the QoE experienced by the user depending the video layer possessing the packets dropped. The metrics for QoE is estimated in next section. Hence, the proposed system drops the video streaming packets across various layers of the video streams and hence it produces lowered bit stream that satisfies the QoE requirements of the users.

The problem of video adaptation in streaming networks in formulated as a function of minimizing the queuing delay in video streams of users. This is achieved consequently under QoE provisioning by dropping the unwanted packets that does not affects the QoE much. Hence, a power-efficient resource allocation – orthogonal frequency division multiple access (RA-OFDMA) module is used for calculating the optimal delay in $M/G/1$ queuing. This takes into account the important consideration of user QoE requirements and timeline of

decoding process. This specifies the maximum delay tolerant value for the videos in streaming network. The delay constraint is transformed into cross-layered constraint using RA-OFDMA module, which is discussed later. This selects the optimal resource blocks and energy allocation policies like P^* and x^* to satisfy the constraints associated with the streaming system.

The proposed QoE model for determining the user QoE's is modeled in the following section. Here, the relationship between the packet reduction and packet loss under QoE constraint is studied. Hence, the relationship between the queuing delay and packet loss ratio in the queue is formulated and that creates an association between the packet loss ratio and QoE of the user. Then, the requirements of the queuing delay are transformed into cross-layered constraint. The relationship between the delay threshold and streaming data rate is used to formulate the transformed cross-layered constraint. Finally, the resource allocation problem and video adaptation in streaming networks in formulated.

QoE Metric Model: This technique uses multi-scale structural similarity index as similar to [9] that helps in providing a better user perceived QOE approximation. Here, the relative QoE scores is calculated between the references and distorted video frames. Also, a mapping of QoE-QoS method is used in this metric model as in [10]. Here, the packet loss ratio is been interpreted to form QoE system level metrics. The QoE degradation is calculated based on the packets dropped at each quality and enhancement layer. For the proposed scheme in DASH video streaming, Monte Carlo simulations are performed. Here, a fixed set or a percentage of video packets are selected uniformly $\rho^k_{r,t}$ from the quality and temporal layer using random distribution. Hence, the average QoE that can be achieved is estimated $E[q(\rho^k_{r,t})]$. This is calculated in the quality or in enhancement layer, when the packet loss ratio is: $\rho^k_{r,t}$. During each iteration, the measurement of quality index and decoded video is done. Each time, $\rho^k_{r,t}$ is performed with different test sets or instance to estimate the average quality value between $0 \le \rho^k_{r,t} \le 1$. The empirical mapping is obtained, when the measurement is taken place repeatedly over temporal and quality layers.

Proposition 1. The reduction of QoE at video stream k is shown in [10]:

$$D^k_{overall} = -q^k_{\max}(\alpha^k - 1) = \sum_{r=0}^{R^k-1}\sum_{t=0}^{T^k-1} D\left(\rho^k_{r,t}\right) \quad (1)$$

where,

$q^k_{\max}$ - Quality of video stream, k and the negative value denotes the absence of losses
α^k - Fractional degradation in quality associated with packet loss

The the QoE degradation caused by $\rho^k_{r,t}$ in r and t is:

$$D\left(\rho^k_{r,t}\right) = -\left(E\left[q\left(\rho^k_{r,t}\right)\right] - q^k_{\max}\right)$$

$\rho^k_{r,t}$ - Packet loss ratio
t - $\rho^k_{r,t}$ in temporal layer
r - $\rho^k_{r,t}$ in quality layer.

Proof. For the preposition is obtainable in [10].

In TCP-ABR system, the packet loss visibility over the individual video layers w.r.t time is quantified through ACK [10]. Once the group of pictures gets transmitted and the entire ACK history is fed to the transmitter, then a prototype of group of pictures that is decoded will be reconstructed and that considers losses over each layer. The packet loss is then directly computed using ACK history and that estimates the effects of distortion arises channel on individual video layer.

Modeling MAC-Layer Through Cross-Layer Perspective: The overall average length of the M/G/1 $q^k_{r,t}$ queue [7] is:

$$\bar{L} = \frac{\bar{\lambda}^2 E\left[X^2\right]}{2\left(1-\bar{\lambda}E[X]\right)} \quad (2)$$

were,

$\bar{L}$, $E[X]$, $E[X^2]$ and $\bar{\lambda}$ denotes $\bar{L}^k_{r,t}$, $E[X^k_{r,t}]$, $E[X^k_{r,t}{}^2]$ and $\bar{\lambda}^k_{r,t}$, respectively.

$E[X]$ - First moment of service time at $q^k_{r,t}$
$E[X^2]$ - Second moment of service time at $q^k_{r,t}$.

The average arrival rate is thus estimated using:

$$\bar{\lambda}^k_{r,t} = \bar{s}^k_{r,t} f^k \left(\frac{n^k_{r,t}}{N^k}\right) \quad (3)$$

where,

f^k - Streaming rate of the frame, k
$\bar{s}^k_{r,t}$ - Average size of temporal layer video frame, t of quality layer r.
N^k - Total frames in GoP
$n^k_{r,t}$ - Total frames in t temporal layers at individual quality layer [11].

$$n_{r,t}^{k}=\begin{cases}2^{t-1} & \text{if } 2\le t\le \log_2 N^k\\ 1 & \text{if } \in [0,1]\end{cases} \tag{4}$$

Form [7], it is found that
Average waiting time is $\bar{W}_{r,t}^{k}=\frac{\bar{L}'^{k}_{r,t}}{\bar{\lambda}_{r,t}^{k}}$ and

Average queue length is $\bar{L}'^{k}_{r,t}=\left(1-\rho_{r,t}^{k}\right)\bar{L}_{r,t}^{k}$.

The packet loss ratio $\rho_{r,t}^{k}$ is taken into consideration. Hence, the average waiting time after substituting $\bar{L}'^{k}_{r,t}$ and (2) in $\bar{W}_{r,t}^{k}$ over M/G/1 queue is given as:

$$\bar{W}_{r,t}^{k}=\frac{\left(1-\rho_{r,t}^{k}\right)\bar{\lambda}^{2}E\left[X^{2}\right]}{2\left(1-\bar{\lambda}E\left[X\right]\right)} \tag{5}$$

Delay Transformation: The maximum tolerance in delay factor $\bar{W}_{\max}^{k}$ is estimated by substituting the upper-bound delay over video stream *k*. Hence, the QoS is transformed into cross-layer constraint through M/G/1 queue. The relationship between the average data rate effectively scheduled between *k* and $\bar{W}_{\max}^{k}$ is modeled using Proposition 2.

Proposition 2: The required and specified condition for maximum delay $\bar{W}_{\max}^{k}$ constraint over a video stream *k* is [6]:

$$e\left[\sum_{l=1}^{L}x_l^k\cdot R_{k,l}\right]\ge\frac{S}{2t_sB\bar{W}_{\max}^{k}}\left(\sqrt{\bar{\lambda}^{k}\bar{W}_{\max}^{k}\left(\bar{\lambda}^{k}\bar{W}_{\max}^{k}+2\left(1-\rho^{k}\right)\right)}+\bar{\lambda}^{k}\bar{W}_{\max}^{k}\right) \tag{7}$$

where,
t_s - Slot duration for scheduling
B - Bandwidth of the resource block and
S - Packet size.
$\bar{\lambda}^{k}$ - Average arrival rate at q^k
ρ^{k} - Packet loss ratio at q^k

The service rate attainable at upper bound with *k* user over the resource block *l* is given as:

$$R_{k,l}=B\log_2\left(1+\frac{P_l^k\left|h_l^k\right|^2}{\sigma^2}\right) \tag{8}$$

where,
σ^2 - noise power and
h_l^k - channel fading coefficient.

Video Adaptation and Resource Allocation: Initially the video adaptation problem is formulated in the form of delay minimization problem associated with QoS queuing constraints. An optimal service rate and packet loss ratio is derived to minimize the delay in queue and thus adapting the QoE constraint video streaming. Finally, cross-layer resource allocation problem is formulated in the form of energy minimization problem over video adaptation phase that considers delay constraint.

Video Adaptation Optimization: The main aim is maximizing the bandwidth capacity that defines the synchronized video streams served during each video stream's delay and QoE requirements. The objective is attained through minimization of the queuing delay over each media streams and it reduces the buffer length of the queue. Hence, a lower bit-rate video streaming version is attained through packet drops that minimizes the QoE and maximizes the deadline of decoding constraints under entire VQ.

$$\min_{\rho,E[X]}E\left[\sum_{r=0}^{R^k-1}\sum_{t=0}^{T^k-1}\bar{W}_{r,t}^{k}\right] \tag{9}$$

$$s.t\quad \bar{W}_{r,t}^{k}\le\bar{W}_{r,t,\max}^{k}\quad \forall k=K,\forall t=T,\forall r=T$$
$$D_{\max}^{k}\ge D_{overall}^{k}$$
$$\sum_{r=0}^{R^k-1}\sum_{t=0}^{T^k-1}E\left[X_{r,t}^{k}\right]\le C^{k}\quad \forall k\in K$$

The (6) defines the objective function, which helps in minimizing the video stream queuing delay. The average waiting time over VQ do not exceed the expiry time or the decoding time $\bar{W}^k_{r,t,\max}$. The decoding deadline of q^t_r is $\bar{W}^k_{r,t,\max} \approx f^{-k}$. $D^k_{\max} \geq D^k_{overall}$ represents the reduction in QoS over at video streams k, which do not exceeds $D^k_{\max}$. This represents maximum degradation in QoE streams, which is decided by the mobile operator. The $\sum_{r=0}^{R^k-1}\sum_{t=0}^{T^k-1} E\left[X^k_{r,t}\right] \leq C^k$ denotes the overall service rates of video stream k with VQ is upper bounded by the end user throughput C^k.

Resource Allocation: The cross layer resource allocation is deployed like [6] and this minimizes the transmitted power from base station to k users. During this time, the delay constraint of each video stream is satisfied and the resource allocation is formulated using:

$$\min_{P,x} E\left[\frac{1}{L}\sum_{k=1}^{K}\sum_{l=1}^{L} x^k_l P^k_l\right] \tag{10}$$

$$s.t.E\left[\frac{1}{L}\sum_{k=1}^{K}\sum_{l=1}^{L} x^k_l P^k_l\right] \leq P_{\max}$$

$$\sum_{k=1}^{K} x^k_l \leq 1$$

$$x^k_l \in \{0,1\}, P^k_l \geq 0$$

$$\bar{W}^k \approx \bar{W}^k_{\max}$$

The optimization problem is shown in (10) that holds the objective of energy and resource allocation for minimizing the transmission power in downlink OFDMA. P_{max} defines the upper limit on available power at the base stations. The $\sum_{k=1}^{K} x^k_l \leq 1$ and $x^k_l \in \{0,1\}$. $P^k_l \geq 0$ denotes the resource block allocated to a single receiver, where x^k_l represents the resource block allocation, which is a binary variable. $\bar{W}^k \approx \bar{W}^k_{\max}$ represents the delay limitations in video streams and $\bar{W}^k_{\max}$ -maximum delay tolerance over stream k.

$$\bar{W}^k_{\max} = E\left[\sum_{r=0}^{R^k-1}\sum_{t=0}^{T^k-1} \bar{W}^{*k}_{r,t}\right] \forall k \tag{11}$$

(11) Provides the optimal solution for resource allocation over video streams k.

Numerical and Simulation Results: The video coding is done through JSVM 9 for deliberately acquiring the results. Here, two video sequences are taken into consideration: *city* video sequence of bitrate 540kbps and *foreman* video sequence of 400kbps. The simulations are performed with highest frame rate, i.e 30 frames per second or fps and total temporal layers is set as 4 and finally the quality enhancement layer is considered as 4. Here, the packet loss visibility over each individual video layer is estimated using proposed constraints.

From the Figure 2, it is found that there is a significant QoE reduction over the city video sequence that possesses more background motions. The rate of reduction is calculated when a uniform packet loss is coded against each layer. Taking into consideration the layer identifier which is set as 0 or r = 0, the losses in each layer affects the video quality and leads to degradation with reduced QoE.

Hence, with scalability in packets, the degradation in video quality has lowered considerably, since the losses occur in upper quality layer or temporal layer [10]. Then the downlink system of each single cell OFDMA is now taken int consideration with the bandwidth of 10 MHz or 50 resource blocks/TTI. The channel is modeled such that it considers log-normal shadow, path loss at larger scale and Rayleigh fading with a noise power of -170 dBm/Hz. Hence, consider that the 8 users are distributed randomly over a 50m distance using uniform distributions from eNodeB.

This research takes into consideration a two scenarios with varied QoE requirements, each of which is discusses as follows:

Scenario A: The considered video sequence is *foreman* video sequence, which is transmitted to the 8 users. The video is dynamically adapted with maximum QoE degradation, $D^k_{\max}$ is set as 0.3.

Scenario B: The considered video sequence is *city* video sequence, which is transmitted to the 8 users. The video is dynamically adapted with maximum QoE degradation, $D^k_{\max}$ is set as 0.1. The degradation factor, $D^k_{\max}$ is set to 0.1, since the scenario considers higher QoE requirements into consideration.

The proposed method is compared over conventional VAWS [3] and WSPmin [12] resource allocation schemes.

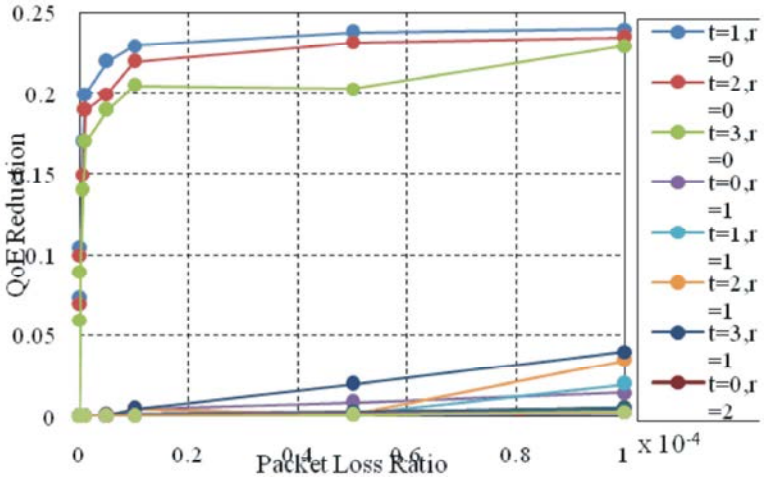

Fig. 2(a): QoE reduction vs. packet loss ratio in *Scenario A* for base layers (t = 1-3 and r=0) and enhancement layers (t=0-3 and r=1-3)

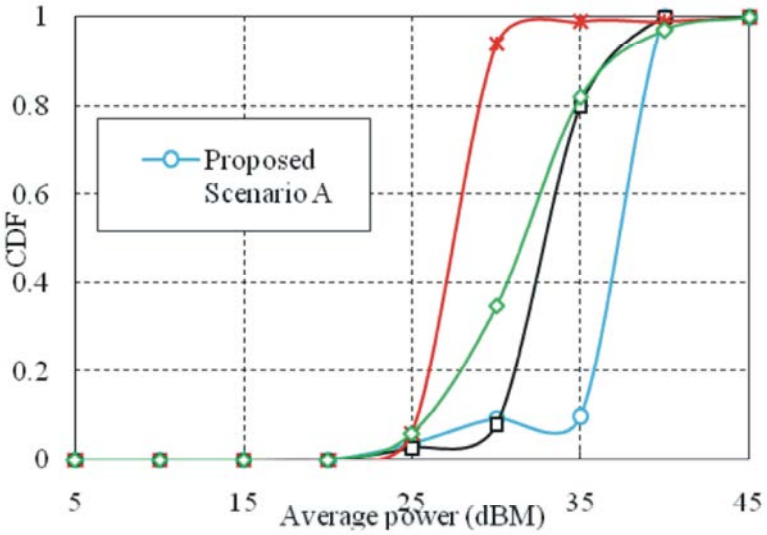

Fig. 3: CDF vs. overall power.

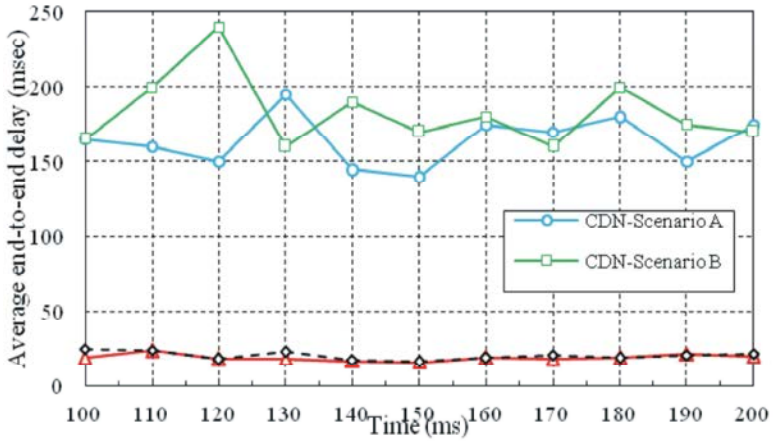

Fig. 4: Comparative analysis of end-to-end delay in various scenarios

The WSPmin technique is designed to minimize the transmit power at a reduced constrained rate. However, in VAWS technique, the resource block is assigned for satisfying the reduced rate constraint that takes into account constant allocation power per resource block. The initial uniform power allocation strategy is refined, such that minimizations of constrained rates are met. The process is continued to the prior phase for refining the power allocation and that meets the minimum rate constraint. The data rate of the conventional WSPmin and VAWS techniques over the served users vary randomly as a 50kbps multiples in between 100-400kbps.

The Figure 3 validates the cumulative distribution power (CDF) of the entire power allocated to various resource allocation schemes, which is iterated over 100 epochs using Matlab. Thus in *Scenario A*, the proposed RA-OFDMA performs well when compared with conventional VAWS and WSPmin algorithms. The power efficiency of the RA-OFDMA is ~12% better than WSPmin and ~20% better than VAWS, over 90% of the time. Similarly, in *Scenario A*, the proposed RA-OFDMA performs well when compared with conventional VAWS and WSPmin algorithms. This proves that the *scenario B* works effectively than *scenario A* that holds more background motions.

The proposed method is further compared with the content delivery networks (CDN) that uses ABR streaming capability to reduce the end-to-end delay. The simulation using OPNET proved that CDN-based streaming transmitted with *scenario A* and *Scenario B* based videos performs less better than the proposed scheme, since it possess a reduction in delay performance by ~89% and ~86%, respectively for each scenarios.

CONCLUSION

A video adaptation scheme using queuing models is designed along with resource allocation strategy in video streaming network edges. The system is implemented over DASH system to prove the effectiveness of the proposed model over its video streams. This method selects the non-optimal packets to be dropped from the video streams. This creates a lowered bitrate, which in turn reduces considerably the delay and provides satisfaction over user QoE requirements. Then the resource is allocated finally to meet the delay constraint in lowered bitrate video streams. The results proved that the proposed method achieves considerable QoE performance in terms of energy efficiency and reduced delay.

REFERENCES

1. Ahlehagh, H. and S. Dey, 2013. April. Adaptive bit rate capable video caching and scheduling. In Wireless Communications and Networking Conference (WCNC), 2013 IEEE (pp: 1357-1362). IEEE.
2. Golrezaei, N., A.F. Molisch, A.G. Dimakis and G. Caire, 2013. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. IEEE Communications Magazine, 51(4): 142-149.
3. Ahlehagh, H. and S. Dey, 2014. Video-aware scheduling and caching in the radio access network. IEEE/ACM Transactions on Networking (TON), 22(5): 1444-1462.
4. Gao, G., W. Zhang, Y. Wen, Z. Wang and W. Zhu, 2015. Towards cost-efficient video transcoding in media cloud: Insights learned from user viewing patterns. IEEE Transactions on Multimedia, 17(8): 1286-1296.
5. Liu, Y., F. Li, L. Guo, B. Shen and S. Chen, 2012. March. A server's perspective of internet streaming delivery to mobile devices. In INFOCOM, 2012 Proceedings IEEE (pp: 1332-1340). IEEE.
6. Zarakovitis, C.C., Q. Ni, D.E. Skordoulis and M.G. Hadjinicolaou, 2012. Power-efficient cross-layer design for OFDMA systems with heterogeneous QoS, imperfect CSI and outage considerations. IEEE Transactions on Vehicular Technology, 61(2): 781-798.
7. Bertsekas, D.P., R.G. Gallager and P. Humblet, 1992. Data networks (Vol. 2). New Jersey: Prentice-Hall International.
8. Sousa-Vieira, M.E., 2011. June. Suitability of the M/G/8 process for modeling scalable H. 264 video traffic. In International Conference on Analytical and Stochastic Modeling Techniques and Applications (pp: 149-158). Springer Berlin Heidelberg.
9. Wang, Z., E.P. Simoncelli and A.C. Bovik, 2003. November. Multiscale structural similarity for image quality assessment. In Signals, Systems and Computers, 2004. Conference Record ofthe Thirty-Seventh Asilomar Conference on, (2: 1398-1402). IEEE.
10. Khalek, A.A., C. Caramanis and R.W. Heath, 2012. A cross-layer design for perceptual optimization of H. 264/SVC with unequal error protection. IEEE Journal on selected areas in Communications, 30(7): 1157-1171.

11. Gupta, R., A. Pulipaka, P. Seeling, L.J. Karam and M. Reisslein, 2012. H. 264 coarse grain scalable (CGS) and medium grain scalable (MGS) encoded video: A trace based traffic and quality evaluation. IEEE Transactions on Broadcasting, 58(3): 428-439.
12. Seong, K., M. Mohseni and J.M. Cioffi, 2006. July. Optimal resource allocation for OFDMA downlinks systems. In Information Theory, 2006 IEEE International Symposium on (pp: 1394-1398). IEEE.