Performance Evaluation of High Definition Video Streaming over Mobile Ad Hoc Networks

Muhammad Usman[a,*], Mian Ahmad Jan[b], Xiangjian He[a], Muhammad Alam[c]

[a] *Global Big Data Technologies Center, School of Electrical and Data Engineering, University of Technology Sydney, Australia*
[b] *Department of Computer Science, Abdul Wali Khan University, Mardan, Pakistan*
[c] *Instituto Telecomunicações, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*

Abstract

Video Service Providers (VSPs) can collect and analyze an enormous amount of data from various cloud storage centers using real-time big data systems for supporting various online customers. The infrastructure-less nature of Mobile Ad Hoc Networks (MANETs) and the mobility of devices make the video streaming a challenging task for VSPs. High packet-loss probability in MANETs can create a notable distortion in the received video quality. Existing techniques for modeling the relationship between packet-loss and distortion in MANETs are linear and might not be suitable for multimedia data. In this paper, High Definition (HD) videos are encoded using High Efficiency Video Coding (HEVC) standard, and are streamed over MANETs. First, a model is designed to demonstrate the effect of packet-loss on the received video quality. Based on the proposed model, a smart strategy is designed to efficiently utilize the available bandwidth in MANETs in order to minimize the packet-loss and improve Quality of Service (QoS). Later, an Error Concealment (EC) technique is used to conceal the missing video frames to improve the Quality of Experience (QoE) of users. During experiments, both subjective and objective evaluations are performed to evaluate the perceptual quality of the concealed video data.

Keywords: VSPs, MANETs, packet-loss, HEVC, QoS, QoE

*Corresponding author
Email address: `Muhammad.Usman@uts.edu.au` (Muhammad Usman)

1. Introduction

Due to the rapid development in electronics and telecommunication industries, online video streaming is becoming popular day-by-day. Video Service Providers (VSPs) need to assure smooth data-driven video streaming services to exploit the surging demand for premium services. In a recent survey, more than 80% viewers suggested that it is extremely important to have a TV-like quality experience on their mobile devices [1, 2]. Due to an increasing number of customers, it is critical for VSPs to improve the overall quality of data-driven video streaming services to hold their customers.

Mobile Ad Hoc Networks (MANETs) allow mobile devices to connect with each other and with the base station, without requiring any particular infrastructure in a mesh style [3]. On the other hand, due to the varying channel conditions and mobility of nodes, links are frequently established and terminated in MANETs. Such a frequent establishment and termination of the transmission links open up doors for various network impairments such as network congestion, buffer overloading, packet-loss and network loops [4]. In order to ensure a smooth High Definition (HD) video streaming, the availability of sufficient bandwidth and a minimum end-to-end delay are highly important factors [5, 6]. Due to the large size, HD videos are always streamed in compressed forms. During a video streaming session, a data packet may contain either some part of a video frame or an entire video frame. Without an efficient routing scheme, video streaming over MANETs becomes vulnerable to packet-loss problem, thus degrading the quality of the received videos.

In data-driven video streaming applications, Quality of Service (QoS) plays an important role to ensure smooth video streaming [7]. Different metrics such as encoding bitrate, startup time, delay at buffers and availability of service have a direct impact on QoS [8]. In the case of MANETs and the Internet, maintaining QoS and providing a smooth end-to-end video streaming is a challenging task for data-driven video streaming systems. Through real-time performance analysis and adaptive video streaming, VSPs can monitor the performance of data-driven video streaming systems, improve streaming services and grow the number of audience through better service [9].

Data-driven video streaming systems can minimize the business costs of VSPs through advanced streaming techniques such as adaptive video streaming. The adaptive video streaming can help in reducing the buffering problem and in providing an interruption-free video streaming through efficient routing techniques in real-time. However, it is unable to resist against other network impairments such as power failure of the networking devices. Along with the QoS, Quality of Experience (QoE) also plays an important role in maintaining/improving the reputation of the VSPs. QoE of the data-driven video streaming systems can be improved by assessing the quality of received video through objective and subjective evaluations. The objective evaluation can be performed using various mathematical techniques such as, Mean Square Error (MSE), Peak-Signal-to-Noise-Ratio (PSNR), Structural SIMilarity index (SSIM), Information Weighted SSIM (IW-SSIM), Feature Similarity Index (FSI), Visual Quality Model (VQM) and Visual Signal-to-Noise-Ratio (VSNR) [10]. On the other hand, subjective evaluation is based on human assessment, i.e., a group of people, also known as subjects, are selected randomly from different age range. They rate the quality of the received video, usually on a scale of 1 to 10. Later on, the obtained score is analyzed using statistical analysis and graphs [11].

In this paper, we propose an optimum bandwidth allocation scheme for data-driven HD video streaming applications in MANETs. The proposed scheme is based on an estimation model to predict the effect of packet-loss on the received video quality. This model helps in managing the available bandwidth through efficient data scheduling, thus improving the QoS of the system. To further improve the quality of the data-driven video streaming system, an Error Concealment (EC) technique is used to conceal/recover the lost video frames, thus improving the QoE. To prove the efficiency of the applied EC technique, both objective and subjective evaluations are performed on the recovered video frames. The main contributions of our proposed scheme are as follow:

- A bandwidth allocation scheme is proposed to minimize the loss of data packets during HD video streaming over MANETs. The proposed bandwidth allocation scheme can help in adjusting the bit-rate of the compressed videos during video streaming.

- A lightweight and efficient EC technique is applied to recover the lost video frames. The EC technique is based on multi-threading-based parallel processing and can recover the lost video frames in real-time.

- To test the performance of the applied EC technique, objective evaluation is performed on the recovered video frames using various metrics.

- To maintain the QoE, subjective analysis is combined with the objective evaluation using subjective scores and statistical analysis.

The rest of this paper is organized as follows. Transmission and distortion models are described in Section 2. Proposed streaming framework is explained in Section 3. Experimental setup is presented in Section 4 and experimental results are analyzed in Section 5. Finally, the paper is concluded in Section 6.

2. System Model

In this section, first, we discuss the transmission model in subsection 2.1 followed by distortion model in the subsection 2.2.

2.1. Transmission Model

Fig. 1 depicts a typical MANET. In this figure, node B is the destination node and nodes A_i ($1 \leq i \leq I$) are the source nodes. All A_i generate HD video data. In a MANET, there can be multiple disjoint transmission paths between A_i and B. Let's denote the transmission path by l, where $1 \leq l \leq L$. Multiple relay nodes (i.e., C_i) can be involved in a transmission path between A_i and B. Each l has a specific transmission rate (i.e., T). The transmission rate T can be divided into two sub-categories, i.e., x and y. The category x represents the transmission rate allocated to a specific source node at time t and category y represents the transmission rate of the background traffic. Thus, the transmission rate at time t (i.e., T_t) can be computed by the following equation.

$$T_t = x_t + y_t. \tag{1}$$

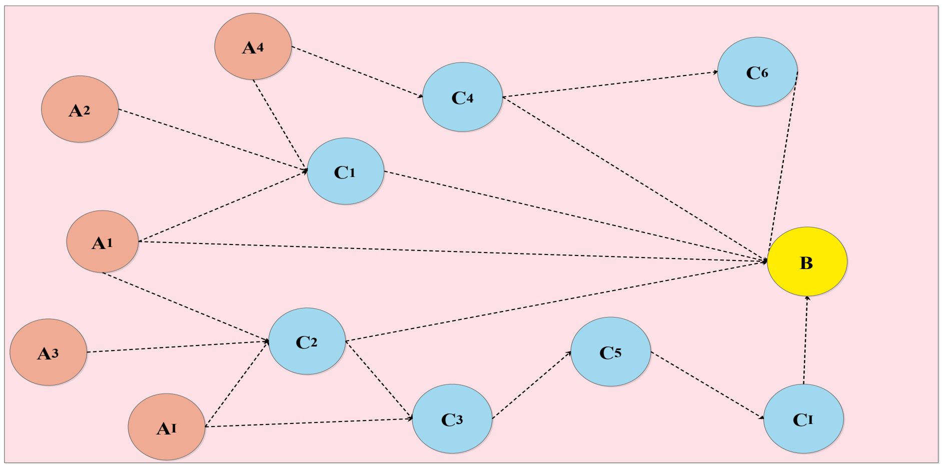

Figure 1: Traditional Mobile Ad Hoc Network

In this model, we assume that the process to transmit the video packets is based on Poisson distribution and the nodes are distributed in a small geographical area (i.e., R) with radius r. Let's denote the node density by α, which is computed by the following equation.

$$\alpha = \frac{\pi \times r^2}{R \times K}, \tag{2}$$

where K represents the total number of mobile nodes in R.

Due to the node mobility, multi-path fading is a common problem in MANETs [12]. In the presence of node mobility and multi-path fading, the bit error rate (i.e., β) of the transmission link from source A to destination B can be approximated by the following equation.

$$\beta_{A,B} = \frac{1}{2}\left[1 - \sqrt{\frac{\alpha \times \gamma}{(\overline{n} \times \overline{B} \times \gamma_o \times T_t) + P + (\alpha \times \gamma)}}\right], \tag{3}$$

where γ is the transmission power of the nodes, $\overline{n}$ is the background noise, $\overline{B}$ is the Boltzman constant, γ_o is the environment temperature and P is the average interference power.

If the packet size is represented by p and load of video traffic is represented by v in R, then the average interference power (i.e., P) between two nodes can be approximated by Eq. 4

$$P = \frac{\alpha \times \gamma \times v \times K \times p}{\beta_{A,B}}. \tag{4}$$

2.2. Distortion Model

In the case of video streaming, the distortion can be divided into two subcategories, i.e., Source Distortion (D_s) and Network Distortion (D_n). The D_s occurs due to encoding at the transmitter side and can be controlled by various encoding parameters such as Quantization Parameters (QPs) and encoding modes (Intra/Inter). On the other hand, many factors such as packet-drop, bit errors and network loops, are possibilities for the D_n. In our proposed distortion model, we are focusing on the D_n only and have paid attention to packet-loss factor.

Before transmission, it is assumed that the videos are encoded in Intra and Unsliced modes. This means that each data packet contains a whole video frame. For a group of video frames (i.e., F), the distortion can be estimated using Eq. 5

$$D_f = \frac{e \times (F - f) \times G}{F \times g \times (1 - e)}, \tag{5}$$

where e is the packet error probability in an end-to-end transmission, f is the lost video frame, g is the selected bit-rate for the entire F, G is the average concealment distortion of the video frame f and D_f is the distortion introduced in F due to the loss of f. G can also be defined as the average channel distortion per pixel of the video frame f when recovered by a suitable EC technique and the value of G is dependent on the characteristics of the applied EC technique.

The packet error probability e can be computed by using Eq. 6.

$$e = \frac{\sum_{l=1}^{L}(p_l \times r_{l,t})}{\sum_{l=1}^{L} r_{l,t}}, \tag{6}$$

where $r_{l,t}$ is the transmission rate permitted by l at time t and p_l is the total transmission rate of l, where $0 \leq r_l \leq p_l$.

If the frame loss probability of f is represented by E_f and transmission coefficient by H_f, then D_n can be computed using Eq. 7.

$$D_n = \sum_{f=1}^{F} D_f \times E_f, \tag{7}$$

where

$$E_f = (1-e)[e^{-H_{(f-1)} \times e} - e^{-H_f \times e}], \tag{8}$$

and

$$H_f = \theta_f . \frac{J}{j}. \tag{9}$$

Here, θ_f is the average relative size of f after encoding, J is the total amount of time required to transmit F from the sender side and j, also known as channel decorrelation time, is the total amount of time after which the characteristics of the selected transmission channel will be changed. The value of j can be estimated using Eq. 10.

$$j = \frac{0.4\lambda}{V}, \tag{10}$$

where λ is the wavelength and V is the velocity of the mobile node.

3. Proposed Streaming Framework

Our proposed framework, i.e. Data-driven-based Video Streaming over MANETs (DVSM), is designed to evaluate the performance of real-time HD video streaming over MANETs and to examine the effect of packet-loss on the quality of received videos. DVSM consists of three major phases, i.e., encoding, streaming and decoding. The main purpose of DVSM is to set a benchmark for the evaluation of real-time HD video streaming over lossy wireless networks.

3.1. Encoding

In this phase, the test video sequences are encoded using the standard test conditions, as specified in the JCT-VC document [13]. The selected version of HEVC encoder is modified to generate a log file. The generated log file consists of eight different fields, i.e., OFFSET, TYPE, NUM_BYTES, FRAME_NO, T-ID, DECODE_TIME, PRIORITY and TIMESTAMP. These fields are used to trace various Network Abstraction Layer (NAL) units, present within the encoded video bit-streams. The encoded video bit-streams consist of various NAL units.

The OFFSET field uses memory offsets to track various byte locations in the encoded video bit-stream such as the first byte of any NAL unit. This field helps to find and extract specific NAL units within the encoded video bit-streams or can help the prioritization schemes to rearrange the priority of NAL units in the encoded video bit-streams. In other words, it plays the role of an identifier for each NAL unit. The TYPE field specifies the type of Raw Byte Sequence Payload (RBSP) data structure contained in the NAL unit. There are 64 different NAL unit types defined within H.265 draft [13]. If the most significant bit of TYPE field is zero, then the NAL unit is a Video Coding Layer (VCL) unit, otherwise it is a non-VCL unit. The NUM_BYTE field is the number of byes of data contained in a NAL unit. Every video frame when it is encoded is split into one or more NAL units depending on the maximum size of a NAL unit. The FRAME_NO is an identifier that the video encoder uses to number video frames during the encoding process. Whenever the video encoder starts encoding a new video sequence, it generates a sequence header and transmits it to the decoder. Afterwards it encodes each incoming video frame and assigns it a new frame number. By reading the FRAME_NO field, the decoder can find out whether a video frame is lost during streaming or not.

The T-ID field specifies the temporal identifier of the NAL unit and cannot be zero. It is used to enable independent considerations of start code emulations in the NAL unit header and in the NAL unit payload data. The DECODE_TIME field defines the amount of time (in milliseconds) by which the current NAL unit should arrive at the decoder side. This time is always considered in relation to the first NAL unit in the video bit-stream, successfully arrived at the decoder side. The NAL units arriving after this specific time are not considered useful and are simply discarded by the decoder. The PRIORITY field defines the priority of various NAL units in a video bit-stream. By default, all NAL units have an equal priority. However, this field can be manipulated at the encoder side, especially in the case of Inter encoding to assign higher priority to Intra packets as compared to the Inter packets. This field can also help to test the performance of EC schemes by dropping or delaying certain NAL units before transmission at the encoder side. The TIMESTAMP field is a ticket, created and attached by the encoder when a NAL unit gets ready to be forwarded to the packet scheduler.

3.2. Streaming

This phase is categorized into four sub-phases. The first three sub-phases are extraction, packetization and scheduling of video bit-stream and are performed at the encoder side at the software level. The last sub-phase is streaming which is performed by the encoding device to transmit the packetized video bit-stream over the wireless network.

3.2.1. NAL Unit Extraction

In the extraction sub-phase, the trace file of an NAL unit is parsed line by line by a customized bit-stream parsing tool to find and extract the NAL units from the encoded video bit-stream. The tool seeks offsets of the NAL units in the encoded video bit-stream. The extracted NAL units are forwarded to the packetization sub-phase for further processing.

3.2.2. NAL Unit Packetization

In the packetization sub-phase, the extracted NAL units are transformed into packets for transmission. For this operation, there are two options defined in the HEVC draft [13]. The first option is the default mode while second option is considered as an optional mode. In the default mode, also known as interleaved mode, a Picture Parameter Set (PPS) message is inserted by the encoder before the first VCL NAL unit in the video bit-stream. During the transmission, the PPS message and the first VCL NAL unit are transmitted separately. If the PPS message gets delayed or lost during the transmission, the entire video bit-stream is discarded by the decoder. In the DVSM framework, we have selected the optional mode, also known as non-interleaved mode, with a feedback modification. In this mode, the PPS message is sent along with the first VCL NAL unit in one data packet. The first VCL NAL unit contains the first Intra frame. After receiving the PPS message along with the first VCL NAL unit, the decoder sends an acknowledgment message to the encoder requesting the next NAL unit. With the exception of the first NAL unit which is sent along with the PPS message, the remaining NAL units are transmitted separately with one NAL unit per data packet.

In the DVSM framework, we customized the Real-time Transport Protocol (RTP) header with 12 bytes size. The customized RTP header, known as Pseudo-RTP Header (PRH), contains the following fields from the generated log file, i.e., OFFSET, NUM_BYTES, FRAME_NO, T-ID, PRIORITY and DECODE_TIME. The PRH enables us to drop selective video packets during video streaming without inspecting the NAL units or payloads. Due to its standard minimum size, the PRH does not bring any extra computational overhead. An RTP packetization tool is used to populate PRH, both with the trace files and headers of the NAL units. The PRH packets are further encapsulated by User Datagram Protocol (UDP) for transmission.

3.2.3. NAL Unit Scheduling

The third sub-phase is the scheduling of encapsulated packets. The scheduling can be performed by two ways, i.e., single path and multi-path transmissions. In the former, the data packets are scheduled over a single transmission path from the source to the destination. In the later, the data packets are scheduled over multiple transmission paths from the source to the destination. In the proposed DVSM, we have implemented both single path and multi-path transmissions.

In both these scheduling, the strategy to forward the data packets is based on first-in-first-out policy. Each mobile node maintains a queue of available transmission paths toward the destination. In the single path scheduling, one earliest available transmission path is fixed for the transmission of the entire video bit-stream. On the other hand, the data packets of a video bit-stream are forwarded on the earliest available transmission paths in the case of multi-path scheduling. In both these scheduling, multiple hops may get involved due to mobility of nodes, for forwarding the data packets towards the destination. In the proposed DVSM framework, multiple network constraints such as packet-loss, delays at the hops and variation in the available bandwidths are applied during the simulations.

3.2.4. Data Streaming

In the proposed DVSM framework, the multi-path transmission involves feedback, i.e., if there is any variation in the available bandwidth or the intermediate hops are busy and the processing delay is increasing, the sender is immediately notified through a feedback message. Upon receiving the feedback message, the sender chooses another available transmission path to schedule the data packets.

The data scheduling is also implemented on the intermediate hops. Therefore, an intermediate hop plays a dual role of a sender and forwarder. The feedback messages keep the senders informed about the current status of the network. Based on the current status of the network, the senders can intelligently schedule the data packets on the transmission paths.

Once the data packets arrive at the destination node, it writes the NAL unit file of the decoder after removing the pseudo-RTP header. The trace file of NAL unit contains the details of the received NAL units along with the order and time stamps of NAL units. Finally, the received NAL units are written in a correct order to the received video bit-stream.

3.3. Decoding

This phase is categorized into three sub-phases, i.e., detection of the lost NAL units, concealment of the lost video frames and quality assessment of the recovered video frames.

3.3.1. Lost NAL Unit Detection

There are two techniques to identify the absence of particular NAL units lost during the transmission. In the first technique, upon receiving the NAL unit trace file at the destination, it is compared with the NAL unit trace file of the encoding device. During this comparison, the lost NAL units are identified and their record is written into a log file of the destination device. This log file contains the status of all NAL units transmitted from the sender device. This technique works in situations where the encoding device transmits the encoder's NAL unit trace file along with the video bit-stream.

In the second technique, the decoder starts decoding after receiving all NAL units. When the decoder detects the absence of a video frame during the decoding process, it pauses the decoding process and calls the EC module to conceal the missing video frame. After the missing video frame is recovered and forwarded to the decoder by the EC module, the decoder adjusts the recovered video frame at its accurate time position and resumes the decoding process until either another missing video frame is detected or the end of file is detected. This technique is suitable for those scenarios where the batch processing is involved, i.e., processing of a group of video frames.

In the proposed DVSM framework, the first technique is adopted based on the following reasons:

- Unlike the second technique, it detects all the errors before the actual decoding starts and helps in smooth decoding of the video bit-stream.
- It not only detects the absence of specific NAL units, but also helps in identifying the modified NAL units with missing fragments.
- If undetected, the modified NAL units can disturb the normal operation of the decoder and can even crash it.
- It also helps in identifying the NAL units arrived after the DECODE_TIME and can help in rearranging the NAL units.

In the selected technique, the decoder is modified to adjust the concealment of lost video frames. Instead of getting crashed due to the missing NAL unit, the modified decoder continues to decode the remaining NAL units. In the meantime, the EC module performs its job by recovering the missing video frame and forwards it to the decoder. Upon reception, the decoder adjusts the recovered video frame at its right time position in the decoded video frames.

3.3.2. Error Concealment

To conceal the missing video frames lost during the video streaming, an EC technique proposed in [14] is applied. This EC technique is designed to conceal HD/UHD video frames and is based on multi-threading-based parallel processing. The main reasons behind choosing this technique are its fast processing and requirement for fewer computational resources. The missing NAL units are recovered by utilizing the current and previous video frames. It is a block-based EC technique in which the blocks of the missing video frames are recovered by estimating the motion vectors. The estimated motion vectors direct the movement of pixels from previous to the current video frames. Finally, the concealed video frames are compared with the original video frames using objective and subjective quality assessment metrics.

3.3.3. Quality Assessment

There are different metrics to perform objective evaluation [10]. On the other hand, the subjective evaluation is based on user's opinion and the users rating of the videos from bad to good quality. In the proposed DVSM framework, we adopt both objective and subjective evaluations to test the quality of the concealed video frames. Mean Opinion Score (MOS) is computed for the subjective evaluation. Original and concealed videos are presented to the subjects to rate the quality of concealed videos on a predefined scale from 0 to 10, where 10 is the rank of highest quality. After obtaining the ratings of all subjects, mean values are computed which are considered as the MOS for presented videos.

4. Experimental Setup

In this section, first we explain the encoding and simulation setup of our proposed DVSM framework. Next, we evaluate the performance of our proposed framework and quality of concealed video sequences using objective and subjective evaluations.

4.1. Encoding Details

There are standard test conditions, specified in the documentation released by JCT-VC to encode videos using HEVC platform [13]. Common encoding modes are summarized in Table 1, from which we select Intra_Main mode. The current HEVC encoder, i.e., HM 16.15 [15] is used to encode the test video sequences.

In our experiments, we select a subset of different test video sequences from the list of test video sequences recommended by JCT-VC. Details of selected test video sequences are shown in Table 2. The details of selected QPs and the generated bit-rates are shown in Table 3. Different values of QPs help us in generating different values of bit-rates to match with the available bandwidth.

The videos are encoded in an Unsliced mode, i.e., one data packet contains an entire video frame and the size of group of pictures is set to 1.

Complexity Configuration	Encoding Mode
All Intra (AI)	Intra_High_Throughput_Rext Intra_Main Intra_Main10 Intra_Main_Rext
Low Delay (LD)	Lowdelay_P_Main Lowdelay_P_Main10 Lowdelay_Main Lowdelay_Main10 Lowdelay_Main_Rext
Random Access (RA)	Randomaccess_Main Randomaccess_Main10 Randomaccess_Main_Rext

Table 1: Encoding modes

Sequence	Resolution	Total Number of Frames	Frame Rate
PeopleOnstreet	2560×1600	150	30
BasketballDrive	1920×1080	500	50
BQTerrace	1920×1080	600	60
Cactus	1920×1080	500	50
Kimono	1920×1080	240	24
ParkScene	1920×1080	240	24

Table 2: Test video sequences

QPs	Test Sequences					
	People On Street	**Basket ball Drive**	**BQ Terrace**	**Cactus**	**Kimono**	**Park Scene**
20	278.128	112.225	162.407	193.524	200.241	230.879
22	288.052	113.666	185.977	294.813	446.911	293.890
24	473.410	144.794	395.298	470.201	450.363	446.098
26	582.405	339.103	544.468	535.997	588.199	502.951
28	648.780	559.067	645.472	655.114	594.499	551.014
30	707.764	829.226	768.785	810.517	655.252	616.100
32	708.505	863.013	923.787	891.769	660.331	633.915
34	726.483	915.518	927.506	955.950	875.172	738.323
36	972.522	932.734	978.381	961.627	928.663	752.869
38	984.595	955.878	982.253	986.038	972.158	909.942

Table 3: QP pairs and video sequences' bit-rates (Kbps)

4.2. Simulation Details

We use Matlab R2017a to setup our underlying network and to execute the simulation environment. We deploy 1000 mobile nodes in an area of 3000×3000 m^2. The mobile nodes are moving with a speed of $1m/s$ and change their positions after every minute. The transmission range of each mobile node is considered as $100m$ and the maximum hop size is set to 15.

The simulations are executed on a desktop computer. In our simulations, we fix the total number of data generating sources, i.e. 300. The simulations run for three times to monitor and test the performance of our proposed DVSM framework. Due to the large network size and desktop environment, each round of simulation ran for more than 10 hours.

For streaming, we implement both single and multi-path transmission modes with time-varying bandwidths. Various percentages of packet-loss are introduced. Packet-loss is the most common form of network impairment and in the real-world scenarios, it can occur due to multiple factors such as fluctuations in the available bandwidth, failure of one or more hops, failure of network devices, network congestion, buffer overloading and end-to-end delay.

4.3. Quality Assessment Details

To assess the quality of concealed videos, we consider both objective and subjective evaluations. We use PSNR, IW-SSIM, FSI, VQM and VSNR metrics to perform objective evaluation in our experiments. Original video frames are used as references for the concealed video frames.

For subjective evaluation, a controlled laboratory space with environmental conditions recommended in [16] is used. Original test videos and the concealed videos are played on a 55-inch LED TV screen (Sony KD55X8500C) with a resolution of 3840×2160 pixels. To support video playback, a high performance server (TVS) is used. The server consists of 6^{th} generation Intel processors, high speed Solid State Drives (SSDs) and a powerful NVIDIA graphics card. A total of 16 subjects (eight males and eight females) participated in the experiments. The ages of the subjects range between 20 and 30. The subjects are mostly students with no experience of video quality assessment and watching videos on such a large screen. During the experiments, the subjects watch both the original and concealed test video sequences and are requested to rate the quality of concealed video sequences on a provided score sheet. A discrete rating scale from 0 to 10 is used to score the video quality, where 0 represents the worst and 10 represents the best quality. In the provided score sheet, five different levels of descriptions such as poor, bad, fair, good and excellent, are also provided. During the experiments, the test video sequences are shuffled for each subject for a fair judgment.

The subjective scores of the individual subjects are converted into difference scores, i.e., $S_{z,b}$ by using Eq. 11.

$$S_{z,b} = s_{z,ref(b)} - s_{z,b}, \tag{11}$$

where $s_{z,b}$ is the score of subject z for the test video sequence b and $s_{z,ref(b)}$ is the score of subject z for the reference video sequence of a test video sequence b.

Next, Differential Mean Opinion Score (DMOS), i.e., X_b, is computed in the range between 0 and 10 for the test video sequence b, using Eq. 12.

$$X_b = \frac{1}{Z}\sum_{z=1}^{Z} S_{z,b}, \tag{12}$$

where Z represents the total number of subjects, which is 16 in this case.

The performances of the selected objective metrics are compared with the subjective evaluation scores, based on the recommendations of ITU-T P.1401 [17]. To compare the performances, we use four different measures, i.e., Outlier Ratio (OR), Spearman's Rank Ordered Correlation Coefficient (SROCC), Root-Mean Square Error (RMSE) and Pearson's Linear Correlation Coefficient (PLCC). OR and SROCC are used to test the consistency of the quality prediction while RMSE and PLCC are used to analyze the accuracy of the quality prediction.

5. Result Analysis

5.1. Performance Evaluation of Video Streaming

In the experiments, both the single and multi-path transmission modes are based on multi-hop transmission. As shown in Fig. 2, the packet-loss ratio is lower in the multi-path transmission as compared to the single-path transmission. In the real-world transmission scenario, the single-path transmission is never preferred for video streaming. It can only be adopted if the source and destination nodes belong to the same local network and the network resources can be guaranteed for end-to-end video streaming.

In the case of single-path transmission, if one relay node stops working, the entire transmission path needs to be rescheduled. This rescheduling requires

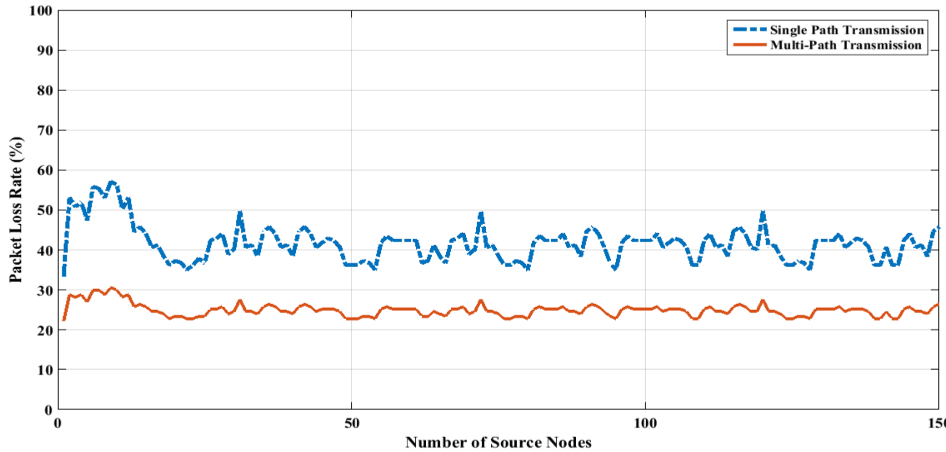

Figure 2: Packet loss rate vs. Number of source nodes

extra computational time and may incur excessive delay in an end-to-end video streaming. An increasing number of relay nodes between the source and destination nodes is another reason for end-to-end delay. The transmission paths keep changing in a multi-path transmission, therefore, on average, the end-to-end delay remains approximately lower as compared to single path transmission, as shown in Fig. 3.

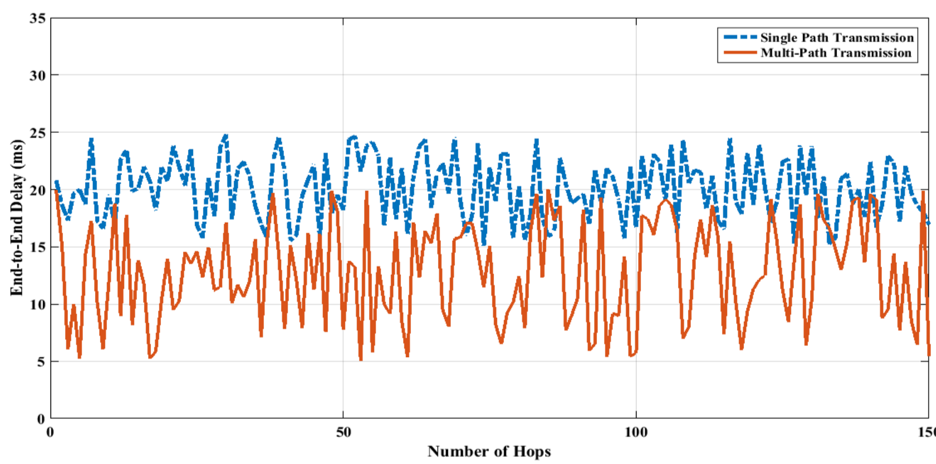

Figure 3: End-to-end delay vs. Number of hops

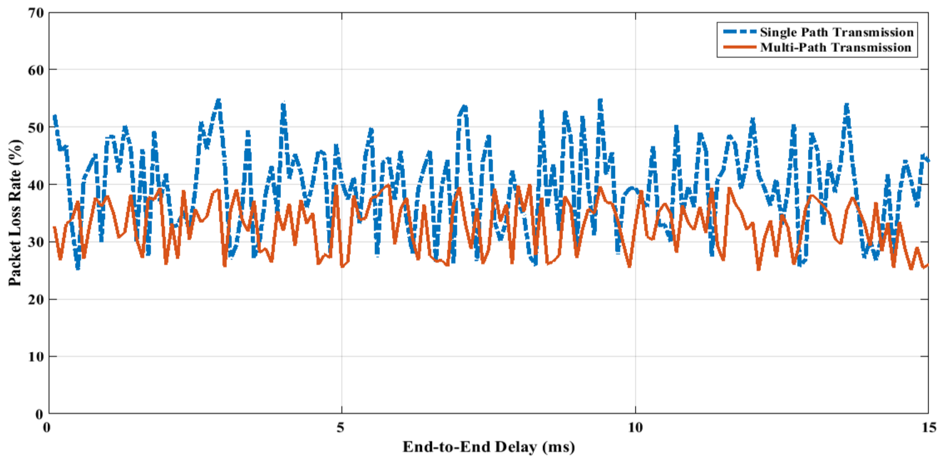

Figure 4: Packet-loss rate vs. End-to-end delay

The packet-loss has a direct relationship with the end-to-end delay as shown in Fig. 4. Due to an indirect communication between the source and destination nodes, the packet-loss problem will always be there. The packet-loss problem can occur due to many reasons such as buffer overloading on the relay nodes, failure of the relay nodes, network loops and unavailability of the bandwidth. Due to these reasons, the data packets either arrive out of time on the destination or get lost during the transmission and need to be retransmitted. In either case, a smooth video streaming is disturbed smooth video streaming and causes end-to-end delay. Finally, a comparison between single and multi-path transmissions is performed based on overhead packets, as shown in Fig. 5. These overhead packets are considered as an extra amount of traffic and increase with an increase in the network size. As a result, an enormous amount of network bandwidth is consumed and become a major disadvantage. On the other hand, these overhead packets contain the current status of the network, thus keeping the nodes updated about various critical facts of the network such as unavailability of the bandwidth on certain transmission paths and status of the relay nodes.

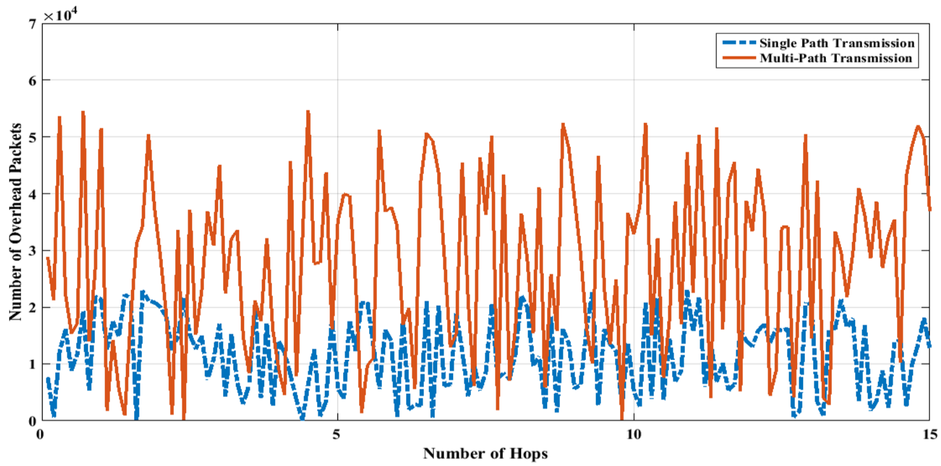

Figure 5: Overhead packets vs. Number of hops

5.2. *Evaluation of Received Video Quality*

The performance of selected metrics for objective evaluation is shown in Table 4. Values in Table 4, are presented in terms of OR, SROCC, RMSE and PLCC measures, respectively. A higher or increasing value in SROCC and PLCC and a lower value in OR and RMSE measures means better or improved performance. As shown in Table 4, column 6 (i.e., PLCC) is showing the best performance on average and column 2 (i.e., RMSE) is showing the worst performance on average, respectively, as compared to the subjective scores. The purpose of this comparison is to find out which metric has statistically performed better than the other metrics.

Rate-Distortion (RD) curves are used to summarize the results of subjective evaluation. The RD curves are used to demonstrate the relationship between the perceived video quality, i.e., DMOS, and the encoding bit-rate. In experiments, it is observed that the encoding bit-rate has a direct relationship with the perceived quality while the DMOS has an inverse relationship with the perceived quality. To show this relationship, a comparison based on RD curves is performed between the average coding efficiency in terms of bit-rate and the obtained subjective quality scores as shown in Fig. 6. To fit the subjective

Metric	Subjective Score	OR	RMSE	SROCC	PLCC
PSNR	0.89662	0.64166	1.57770	0.82328	0.80395
IW-SSIM	0.80065	0.46666	0.86236	0.94593	0.94530
FSI	0.79292	0.37916	0.66030	0.95585	0.96855
VQM	0.77880	0.25833	0.53896	0.96216	0.97895
VSNR	0.88437	0.51666	1.26718	0.90520	0.87866

Table 4: Comparison between objective and subjective scores

evaluation scores in the RD curves, a confidence interval (i.e., 95%) is used due to the statistical property of the subjective evaluation score. As shown in Fig. 6, the DMOS value decreases as the encoding bit-rate increases. A variation in the obtained bit-rate over different values of QPs is demonstrated in Fig. 7. As shown, the bit-rate increases with an increasing value of QP.

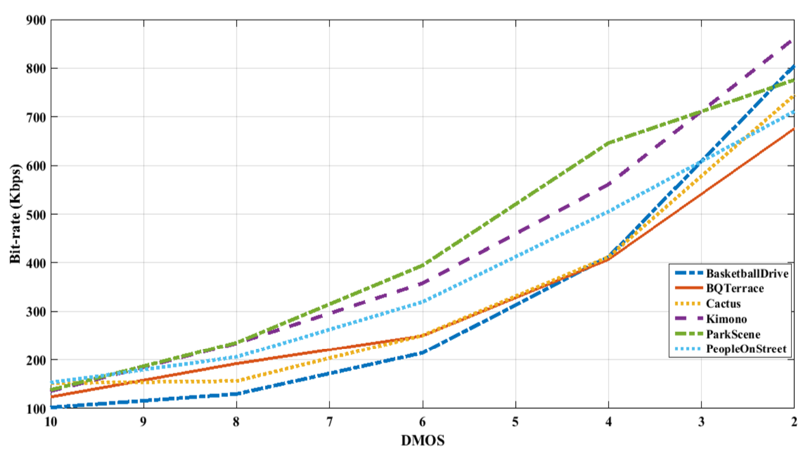

Figure 6: DMOS score vs. Bit-rate (Kbps)

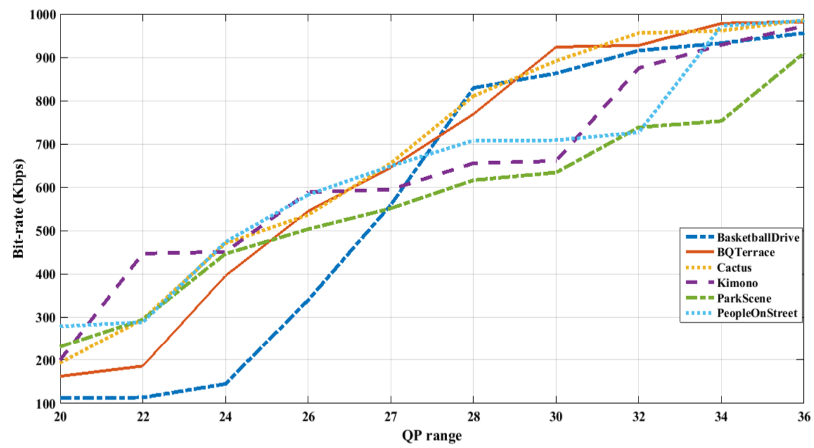

Figure 7: QP values vs. Bit-rate (Kbps)

6. Conclusion

In this work, we have presented a framework to simulate data-driven HD video streaming over MANETs and to evaluate the quality of the concealed video frames through subjective and objective analysis. In this framework, first, a streaming model is proposed for end-to-end video streaming over MANETs and to study the impact of network impairments such as packet-loss, on the transmitted video bitstreams. Later, an EC technique is applied to recover the lost video frames in real-time. The perceptual quality of the recovered video frames is investigated through various objective metrics and subjective scores. The effect of packet-loss on the streamed videos is reported through experimental results under various streaming and coding conditions. The future work includes the implementation of the proposed framework with modifications for scalable HD video streaming over MANETs and VANETs.

References

[1] C. V. N. Index, Global mobile data traffic forecast update, 2016–2021 white paper, link: http://goo. gl/ylTuVx.

[2] E. Consumerlab, Experience shapes mobile customer loyalty, An Ericsson Consumer and Industry Insight Report.

[3] S. Basagni, M. Conti, S. Giordano, I. Stojmenovic, Mobile ad hoc networking, John Wiley & Sons, 2004.

[4] J. Friginal, D. de Andrés, J.-C. Ruiz, M. Martínez, A survey of evaluation platforms for ad hoc routing protocols: A resilience perspective, Computer Networks 75 (2014) 395–413.

[5] P. Juluri, V. Tamarapalli, D. Medhi, Measurement of quality of experience of video-on-demand services: A survey, IEEE Communications Surveys & Tutorials 18 (1) (2016) 401–418.

[6] Y. Chen, K. Wu, Q. Zhang, From qos to qoe: A tutorial on video quality assessment, IEEE Communications Surveys & Tutorials 17 (2) (2015) 1126–1165.

[7] X. M. Zhang, Y. Zhang, F. Yan, A. V. Vasilakos, Interference-based topology control algorithm for delay-constrained mobile ad hoc networks, IEEE Transactions on Mobile Computing 14 (4) (2015) 742–754.

[8] N. Kumar, S. Zeadally, J. J. Rodrigues, Qos-aware hierarchical web caching scheme for online video streaming applications in internet-based vehicular ad hoc networks, IEEE Transactions on Industrial Electronics 62 (12) (2015) 7892–7900.

[9] W. Castellanos, J. C. Guerri, P. Arce, Performance evaluation of scalable video streaming in mobile ad hoc networks, IEEE Latin America Transactions 14 (1) (2016) 122–129.

[10] S. Chikkerur, V. Sundaram, M. Reisslein, L. J. Karam, Objective video quality assessment methods: A classification, review, and performance comparison, IEEE transactions on broadcasting 57 (2) (2011) 165–182.

[11] S. H. Bae, T. N. Pappas, B.-H. Juang, Subjective evaluation of spatial resolution and quantization noise tradeoffs, IEEE Transactions on Image Processing 18 (3) (2009) 495–508.

[12] J. Mullen, H. Huang, Impact of multipath fading in wireless ad hoc networks, in: Proceedings of the 2nd ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks, ACM, 2005, pp. 181–188.

[13] F. H. H. Institute, Jct-vc hevc (2017).
URL https://hevc.hhi.fraunhofer.de/

[14] M. Usman, X. He, K.-M. Lam, M. Xu, S. M. M. Bokhari, J. Chen, Frame interpolation for cloud-based mobile video streaming, IEEE Transactions on Multimedia 18 (5) (2016) 831–839.

[15] Hm-16.15 (2017).
URL https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[16] I. Recommendation, Bt.500,"methodology for the subjective assessment of the quality of television pictures," recommendation itu-r bt. 500, ITU Telecom. Standardization Sector of ITU.

[17] I. Recommendation, 1401: Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models, int'l telecom, Union Std.