

Research Article

Multi-Resolution Multimedia QoE Models for IPTV Applications

Prasad Calyam,¹ Prashanth Chandrasekaran,¹ Gregg Trueb,¹ Nathan Howes,¹
Rajiv Ramnath,¹ Delei Yu,² Ying Liu,² Lixia Xiong,² and Daoyan Yang²

¹ Ohio Supercomputer Center/OARnet, The Ohio State University, Columbus, OH 43210, USA

² Huawei Technologies, Shenzhen 518129, China

Correspondence should be addressed to Prasad Calyam, pcalyam@osc.edu

Received 2 April 2011; Accepted 27 September 2011

Academic Editor: Thomas Magedanz

Copyright © 2012 Prasad Calyam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Internet television (IPTV) is rapidly gaining popularity and is being widely deployed in content delivery networks on the Internet. In order to proactively deliver optimum user quality of experience (QoE) for IPTV, service providers need to identify network bottlenecks in real time. In this paper, we develop psycho-acoustic-visual models that can predict user QoE of multimedia applications in *real time* based on online network status measurements. Our models are neural network based and cater to multi-resolution IPTV applications that include QCIF, QVGA, SD, and HD resolutions encoded using popular audio and video codec combinations. On the network side, our models account for jitter and loss levels, as well as router queuing disciplines: packet-ordered and time-ordered FIFO. We evaluate the performance of our multi-resolution multimedia QoE models in terms of prediction characteristics, accuracy, speed, and consistency. Our evaluation results demonstrate that the models are pertinent for real-time QoE monitoring and resource adaptation in IPTV content delivery networks.

1. Introduction

Internet television (IPTV) is rapidly gaining popularity and is expected to reach over 50 million households in the next two years [1]. The key drivers of IPTV deployment are its cost-savings when offered with VoIP and Internet service bundles, increased accessibility by a variety of mobile devices, and compatibility with modern content distribution channels such as social networks and online movie rentals.

In spite of the best-effort quality of service (QoS) of the Internet, IPTV service providers have yet to deliver the same or better user quality of experience (QoE) than traditional TV technology. Consequently, they need to understand and balance the trade-offs involved with various factors that affect IPTV deployment (shown in Figure 1). The primary factors are: *user* (video content, display device), *application* (codec type, encoding bit rate), and *network* (network health, router queuing discipline) factors.

User factors relate to the temporal and spatial activity level of the video content. For example, a news clip has low activity level, whereas a sports clip has high activity level. Also, based on mobility and user-context considerations, the

display device could support a subset of video resolutions such as quarter common intermediate format (QCIF), quarter video graphics array (QVGA), standard definition (SD), or high definition (HD). Typically QCIF and QVGA with video resolutions of 176×144 and 320×240 , respectively, are suited for hand-held devices, whereas SD and HD with video resolutions of 720×480 and 1280×720 , respectively, are suited for fixed displays at home and business. Application factors relate to the audio (e.g., MP3, AAC, AMR) and video (e.g., MPEG-2, MPEG-4, H.264) codecs and their corresponding peak encoding bit rates, whose selection is influenced by network factors. Network factors relate to the end-to-end network bandwidth available between the head-end and consumer sites, and consequently the network health levels are measured using the delay, jitter and loss QoS metrics.

In order to proactively deliver optimum user Quality of experience (QoE), providers and IPTV application developers need to identify network bottlenecks in real time and assess their impact on the audio and video quality degradation. While doing so, it is impractical for them to rely on actual end users to report their subjective QoE of

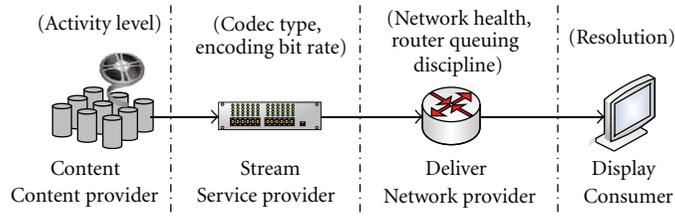


FIGURE 1: IPTV system components.

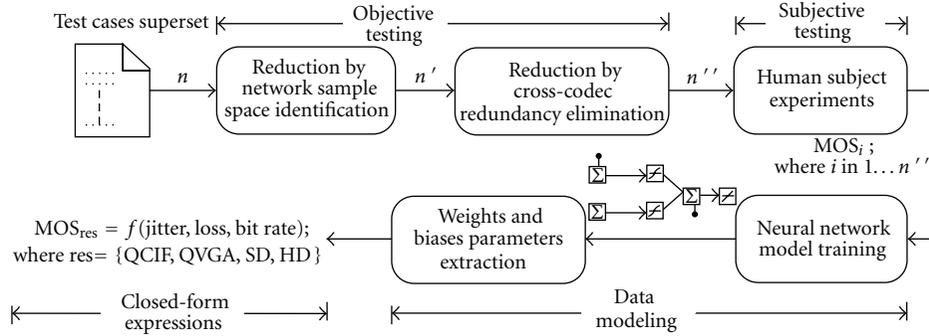


FIGURE 2: Overall architecture of proposed method for real-time multimedia QoE monitoring.

audio-visual quality. Also, they cannot rely on objective techniques such as [2, 3] that involve frame-to-frame peak-signal-to-noise ratio (PSNR) comparisons of original and reconstructed video sequences, which are computationally intensive and time consuming to perform. Hence, there is a dire need for objective techniques that can predict user QoE in *real time* for any given network health condition, while paying due considerations to salient user and application factors.

In this paper, we address this issue and propose psycho-acoustic-visual models that can predict user QoE of multimedia applications in *real time* based on online network status measurements. Our models are “zero-reference” in practice and hence do not require involvement of end users and are non dependent on actual video content. Our models pay attention to salient user and application factors by considering multi-resolution IPTV applications encoded using popular audio-video codec combinations and bit rates. More specifically, our models account for QCIF, QVGA, SD, and HD resolutions and the following codec combinations: MPEG-2 video with MPEG-2 audio, MPEG-4 video with AAC audio, and H.264 video with AAC audio. On the network side, our models account for jitter and loss levels, as well as router queuing disciplines: (i) packet-ordered FIFO (PFIFO), where packet egress of a flow is ordered based on packet sequence numbers, and (ii) time-ordered FIFO (TFIFO), where packet egress is ordered based on packet timestamps. We remark that one of the notable contributions of our study corresponds to our experiments described in this paper that show how the user QoE varies depending on whether the multimedia flows traverse PFIFO- and/or TFIFO- based routers along the network path between the head-end and the consumer sites.

Our proposed solution characteristics are shown in Figure 2. In order to develop the models, we execute a series of well-designed objective and subjective test cases in an offline closed-network IPTV testbed featuring human subject experiments. The ultimate goal of our objective and subjective testing is to come up with a set of user QoE training data from human subject experiments that can be fed to a *neural network* tool. By extracting the weights and biases parameters from the trained neural network, we obtain closed-form expressions for online prediction of user QoE. We use the popular “mean opinion score” (MOS) [4] as the subjective QoE metric during human subject experiments. The QoE prediction for a given resolution (MOS_{res}) is a function of online measurable network parameters (jitter, loss, and bit rate) obtainable from IPTV streams at monitoring points in the network.

The challenge is to significantly minimize human subject experimentation time, without compromising adequate sampling coverage necessary to achieve reasonable prediction accuracy from the neural network models. One of the minimization strategies we use is the objective testing step, in which we develop novel test case reduction schemes, namely, “network sample space identification” and “cross codec redundancy elimination”. These schemes take as input the test cases superset n that includes all of the salient user, application and network factor considerations desired in the models. Using two novel objective QoE metrics at the user-level (“perceptible impairment rate”) and network-level (“frame packet loss”), our schemes notably reduce the test cases set to a manageable number n'' that is suitable for human subject experimentation in compliance with the widely accepted ITU-T P.911 recommendation [4]. We evaluate the prediction performance of our models in terms

of overall characteristics, accuracy, consistency, and speed. More specifically, we systematically evaluate model prediction characteristics for a comprehensive set of control inputs and verify if the prediction behavior follows expected patterns. The accuracy of the model predictions is evaluated using metrics such as the correlation coefficient and root mean square error. The consistency of the model predictions are evaluated using an outlier ratio metric. We use small-to-large scale flows simulation to measure the model prediction speeds.

The remainder of the paper is organized as follows: Section 2 presents related work. Section 3 describes our IPTV testbed setup and component configurations chosen for test cases execution. Section 4 presents the measurement methodology involved in the objective and subjective testing to collect the model training data. Section 5 describes our neural network architecture and model parameterization process. Section 6 presents our models prediction performance evaluation experiments and results. Section 7 concludes the paper.

2. Related Work

Factors that affect multimedia QoE have been extensively studied in earlier works. In [5], the effects of video activity levels on the instantaneous encoded bit rates are shown. Authors of [6, 7] show that higher activity level clips are more affected due to network congestion by 1% to 10% compared to lower activity level clips. The performance of MPEG and H.26x codecs used in IPTV applications are evaluated at various resolutions in [8]. The effects of degraded QoS conditions on multimedia QoE due to network congestion and last-mile access-network bottlenecks is presented in [9–11]. Several objective (e.g., PSNR, SSIM, PEVQ) and subjective (e.g., ACR MOS, DSIS MOS) metrics that quantify user QoE performance are described in [4, 12–14].

There already exist several objective techniques that use psychoacoustic-visual models to predict user QoE for resource adaptation purposes in multimedia content delivery networks. One of the earliest is the E-model [15, 16], which is a psycho-acoustic model developed to address VoIP applications. Recently, works such as [13, 17, 18] have attempted to develop psycho-acoustic-visual models for online multimedia QoE estimation. In [17], video distortion due to packet loss is estimated using a loss-distortion model. The loss-distortion model uses online packet loss measurements and takes into account other inputs such as video codec type, coded bit rate, and packetization to estimate online relative-PSNR values. In [18], a human visual system (HVS) model is proposed that produces video QoE estimates without requiring reconstructed video sequences. The HVS model is primarily targeted for 2.5/3G networks, and consequently it only accounts for PSNR degradation for online measurements of noisy wireless channels with low video encoding bit rates. In [13], a random neural network (RNN) model is proposed that takes video codec type, codec bit rate, and packet loss as well as loss burst size as inputs and produces real-time multimedia QoE estimates. In [19],

authors use weighted values of QoS metrics within quality bounds obtained from other experimental studies and derive a numerical formula for IPTV QoE estimation. In [20], machine learning classifications are used on subjective test data for real-time QoE prediction in mobile networks. In [21], it is shown through simulations how objective QoE monitoring information can be leveraged for resource management via gradient-based routing in IPTV content delivery networks.

In comparison to above works, our novel QoE models encompass multiple video resolutions such as QCIF, QVGA, SD, and HD over a wide range of codec combinations such as MPEG-2 video with MPEG-2 audio, MPEG-4 video with AAC audio, and H.264 video with AAC audio, as well as incorporate variations in jitter, loss and router queuing disciplines on network health side. Our QoE models are able to deal with such a large combination of QoE factors than any of the earlier works due to our innovative methodologies of test-case reduction (i.e., reduction by network sample space identification and reduction by cross-codec redundancy elimination) by objective testing before conducting expensive and time-consuming subjective testing with human subjects. In addition, we have proposed two novel objective QoE metrics at the user-level (“perceptible impairment rate”) and network-level (“frame packet loss”) to aid in the test-case reduction steps. Moreover, we apply a feed-forward neural network [22] with an input layer, hidden layer with 2 neurons, and an output layer. Our multiple layers of neurons with nonlinear transfer functions allow the feed-forward neural network to learn both linear and nonlinear relationships between input (i.e., jitter, loss, bitrate) and output (i.e., MOS) vectors.

In addition to above academia efforts, there have lately been significant efforts in IPTV industry forums such as the Video Quality Experts Group (VQEG), International Telecommunications Union (ITU-T), Alliance for Telecommunications Industry Solutions (ATIS), and Broadband Forum to develop, validate, and standardize online multimedia QoE models [23]. To the best of our knowledge, model requirements being discussed in those forums closely match the modeling considerations in this paper.

3. IPTV Testbed and Experiments Setup

The IPTV testbed setup for our experiments is shown in Figure 3. We used the popular open-source VLC software [24] that has both the streaming server and media player components that allow streaming video content over IP networks to remote users. The Netem network emulator [25] separated two virtual LANS (VLAN 1 and 2) that were abstractions of the stream provider and consumer ends of Figure 1, respectively. For each experiment, we collected both the sender-side and receiver-side packet traces using a sniffer PC with tcpdump utility and stored them in a traffic traces repository. Our Netem was built on Linux Kernel 2.6.15 and had the following hardware: Intel Pentium 4 1.70 GHz CPU, 512 MB RAM, Debian OS, and Intel 10/100 interface card.

We now describe the user, application, and network factor component configurations used in our experiments.

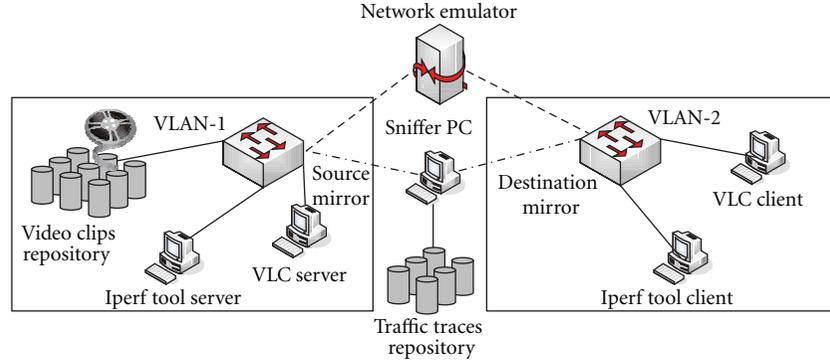


FIGURE 3: IPTV testbed setup.

TABLE 1: Peak encoding bit rates selection.

Resolution	Peak encoding bit rates (bps)
QCIF (177 × 140)	32 K, 64 K, 128 K, 256 K, 512 K
QVGA (340 × 240)	128 K, 192 K, 256 K, 512 K, 768 K
SD (720 × 480)	512 K, 1 M, 2 M, 3 M, 5 M
HD (1280 × 720)	1 M, 2 M, 5 M, 8 M, 12 M

3.1. User Factors. We generated several video clips (with audio tracks) that featured low and high activity level content with clip durations of 16–20 seconds. The video clips were extracted from a variety of sources that included conversation scenes, news telecasts, nature shows, sports telecasts, and movie trailers. Low activity level clips had little primary subject movement and little or no background movement. High activity level clips had high primary subject movement with fast changing backgrounds. Uncompressed video clips were exported to avi file format in QCIF, QVGA, SD, and HD resolutions.

3.2. Application Factors. The uncompressed video clips were then transcoded for streaming and playback with different audio and video codec combinations using the MPEG-TS container format. Specifically, the transcoding involved the following audio and video codec combinations that are popularly used: MPEG-2 video with MPEG-2 audio, MPEG-4 video with AAC audio, and H.264 video with AAC audio. For each resolution, a set of peak video encoding bit rates were chosen as shown in Table 1 based on the bandwidth consumption ranges and step sizes commonly used in practice.

3.3. Network Factors. We configured the network health, that is, delay, jitter and loss values using the corresponding commands found in Netem documentation. We did not constrain the end-to-end bandwidth capacity using Netem. In all our configurations of jitter, we set the delay value to be twice the jitter value because the delay value generally needs to be greater than the configured jitter value for proper jitter emulation. Since we are interested in only the playback quality of the IPTV content and not the channel

change time, we did not use delay as a control parameter in our experiments. We qualified our Netem jitter and loss configurations using Iperf and Ping tools and verified that Netem behavior was as expected before we commenced any of our experiments.

We also configured PFIFO/TFIFO router queuing disciplines in Netem as control parameters in our experiments. Figures 3(a) and 3(b) illustrate the comparison between the TFIFO and PFIFO router queuing disciplines. We can see that the TFIFO router processes packets of a flow F_1 at time t with interpacket jitter J as and when they arrive interleaved with packets of other flows. During the processing time δ , the router scheduler ensures the packets of flow F_1 egress the TFIFO such that the original J is retained. Given that IPTV traffic streams have small interpacket times, they are more likely to experience reordering on the Internet [26], and the consumer end may receive several re-ordered packets. Such packet reordering does not affect general TCP traffic of, for example, web file downloads. However, this seriously affects IPTV application UDP traffic because it burdens the receiver-side media player with packet ordering, and also set-top box decoders do not attempt to order packets due to their processing constraints (Figure 4).

To overcome such re-ordering effects, some network routers support PFIFO queuing that order packets at egress based on their packet sequence numbers. However, in the process, the routers change the inter-packet jitter J to J' . Depending upon the magnitude of the inter-packet jitter change $\Delta J = J - J'$, the end-to-end IPTV traffic stream jitter varies significantly, and consequently, so does the multimedia QoE at the receiver-side.

To illustrate the effects of ΔJ on the end-to-end network jitter experienced by a traffic flow and also to demonstrate the impact of PFIFO and TFIFO router queuing disciplines, we conducted a set of experiments whose results are shown in Tables 2 and 3. The experiments involved measuring end-to-end network jitter using Iperf (10 Mbps UDP mode) and Ping tools between two hosts separated by Netem configured with both PFIFO and TFIFO queuing disciplines. We can observe that both the Iperf and Ping jitter measurements correlate well with the Netem configured jitter values in the TFIFO case. However, in the case of PFIFO, Ping jitter measurements correlate well with the Netem configured jitter

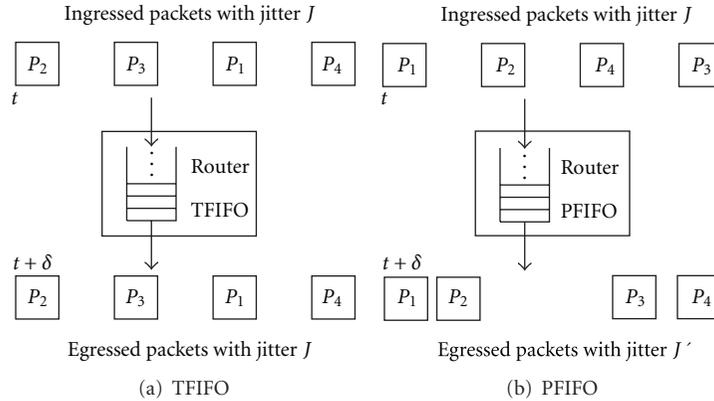


FIGURE 4: Comparison of router queuing disciplines.

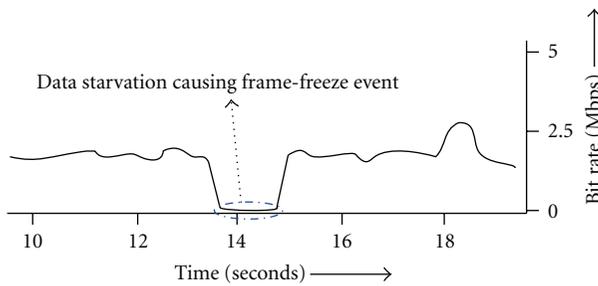


FIGURE 5: Illustration of data starvation due to high interpacket arrival time.

TABLE 2: Jitter measurements under TFIFO.

Netem configured	Iperf measured	Ping measured
6 ms	6 ms	6 ms
18 ms	18 ms	17 ms
75 ms	75 ms	69 ms
120 ms	120 ms	110 ms

TABLE 3: Jitter measurements under PFIFO.

Netem configured	Iperf measured	Ping measured
6 ms	1.861 ms	6 ms
18 ms	1.891 ms	16 ms
75 ms	1.494 ms	61 ms
120 ms	1.056 ms	105 ms

values, and Iperf jitter measurements do not correlate well. This is because the default inter-packet time in Ping is one second, which is much larger than the average PFIFO processing time. Whereas, the inter-packet times of Iperf UDP packets are small enough that the Iperf stream packets experience buffering in the PFIFO router queue and undergo processing to egress in order. The net effect of the PFIFO processing on IPTV application traffic is that the actual network-induced jitter is subsumed and the audio and video impairments perceived at the receiver end are reduced than compared to TFIFO processing.

It is relevant to note that network path processing due to PFIFO and TFIFO could sometimes hold up trains of IPTV application packets causing intermittent high inter-packet arrival times. In such cases, we observed “receiver buffer data starvation” at the client side, which causes frame-freeze impairment events at receiver-side playback. Figure 5 shows an illustration of the data starvation due to a high inter-packet arrival time case. Since data starvation cases could occur due to network processing anomalies such as network load bursts or denial of service attacks, their occurrence is probabilistic in nature on the Internet. To suitably inject frame-freeze impairment events in objective and subjective testing using Netem, we conducted a systematic set of experiments on Netem and concluded that the probability of these events occurring is well correlated with relatively high network jitter conditions under both PFIFO and TFIFO. Hence, we installed a script in Netem to randomly inject one or more frame-freeze events per video sequence in test cases with relatively high jitter settings; specifically we choose >200 ms threshold in PFIFO and >50 ms threshold in TFIFO based on our empirical observations.

4. Measurement Methodology

To systematically analyze the multimedia QoE performance in IPTV content delivery networks under PFIFO and TFIFO, we have to deal with a large sample space with several possible network health conditions. The network health conditions consist of the isolated and combined effects of jitter and loss. Obviously, it is not feasible to analyze multimedia QoE performance for all possible combinations of network jitter and loss on the Internet. Fortunately, earlier empirical studies [10, 27] have shown that multimedia QoE tends to be either in “Good”, “Acceptable”, or “Poor” (GAP) grades of subjective user perception for certain levels of jitter and loss values. The Good grade corresponds to cases where a human subject perceives none or minimal impairments and the application is always usable. The Acceptable grade refers to cases where a human subject perceives intermittent impairments yet the application is mostly usable. Lastly, the Poor grade refers to cases where a human subject perceives

severe and frequent impairments that make the application unusable.

In this section remainder, we first describe the objective testing to reduce the number of test cases to be manageable for subjective testing. Next, we describe the subjective testing and data collection for modeling.

4.1. Objective Testing. The first goal of the objective testing is to perform QoE GAP grades mapping to QoS levels for the different resolutions. This testing enables us to determine the QoS levels that provide definitive GAP sampling ranges of network health conditions. With this knowledge of the QoS levels, we deduce the following 9 network conditions (each denoted by [(jitter grade, loss grade)] pair) that are sufficient for model coverage: [(GG), (GA), (GP), (AG), (AA), (AP), (PG), (PA), (PP)]. The second goal of the objective testing is to perform test case reduction by eliminating test cases that are redundant in the application factors. Specifically, we perform cross-codec elimination when we know that two codecs perform the same under a given network condition.

4.1.1. Objective Metrics. In each test case in the objective testing, a random activity level video clip with the corresponding resolution was streamed from VLAN 1 using the VLC streaming server at a particular bit rate with the corresponding codec. The Netem was configured with a particular network health condition and router queuing discipline. The video clip stream was made to pass through the Netem before playback using the VLC media player in VLAN 2. We collected measurements of two objective QoE metrics, one at the user-level and another at the network-level, namely, “perceptible impairment rate (PIR) events/sec” and “frame packet loss (FPL) %,” respectively.

PIR is the sum of the audio impairment events (e.g., dropouts, echoes) and video impairment events (e.g., tiling, frame freezing, jerkiness, blur) counted by two human observers at the receiver end (one “listener” for audio and one “viewer” for video) divided by the length of the video clip. FPL is the percentage of the number of packets lost (audio and video combined) in a frame and is calculated from the traffic traces as a ratio of number of packets lost to the number of frames in a video clip. We used PIR as our primary metric in test case reduction experiments. To minimize learning effect of the human observers measuring PIR, we randomized the test case execution. Also, we confirmed that the PIR measurements were repeatable across diverse demographics and expertise levels with multimedia applications. From the PIR repeatability studies, we found the PIR threshold values to be: ≤ 0.2 for Good grade, ≤ 1.2 for Acceptable grade, and > 1.2 for Poor grade. Further, we found a direct correlation between PIR and FPL, and hence we used the FPL measurements selectively to verify the sanity of our PIR measurements.

4.1.2. GAP Network Sample Space Identification. We conducted experiments to determine the GAP ranges of jitter and loss for the QCIF, QVGA, SD, and HD resolutions using the MPEG-4 video with AAC audio codecs. The video codec

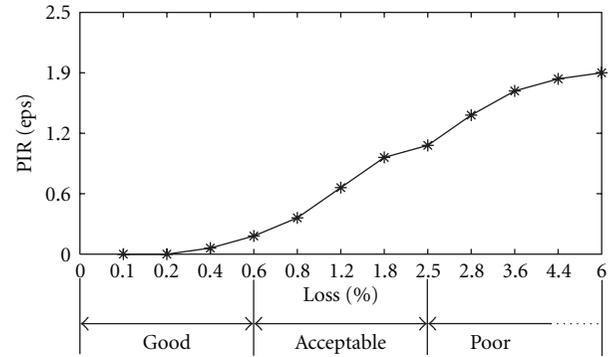


FIGURE 6: Loss GAP ranges for SD resolution under both PFIFO/TFIFO.

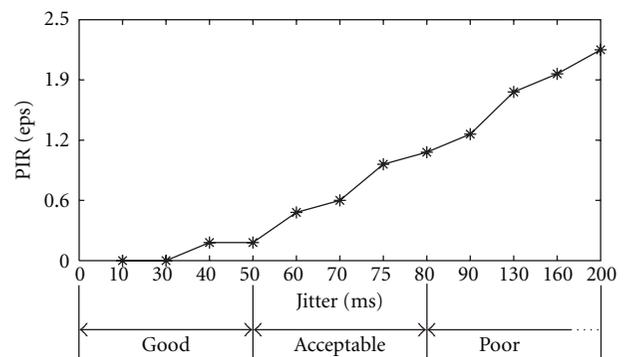


FIGURE 7: Jitter GAP ranges for QCIF resolution under TFIFO.

choice is representative and is motivated by the fact that MPEG-4 (specifically MPEG-4 part 2) is widely used and is known to have better performance than MPEG-2 and typically as good performance as H.264 (specifically MPEG-4 part 10). Also, we selected the median of the peak encoding bit rate step sizes shown in Table 1 for each of the resolutions (i.e., 128 Kbps for QCIF, 256 Kbps for QVGA, 2 Mbps for SD, and 5 M for HD). In each experiment, we gradually increased one of the QoS metric (i.e., jitter or loss) levels till PIR measurements crossed thresholds for GAP QoE grades. We stopped at PIR measurements close to 2, which is the expected human perception capability limit.

Figure 6 shows how the loss GAP boundaries were fixed for SD resolution based on the PIR threshold values. Our experiment results showed that the loss characteristics were independent of the queuing discipline, and hence the loss GAP ranges are the same for PFIFO and TFIFO. Similarly, Figure 7 shows how the jitter GAP boundaries were fixed for the QCIF resolution under TFIFO. The complete list of jitter and loss GAP ranges under PFIFO and TFIFO for the different resolutions are shown in Tables 4 and 5, respectively.

We now discuss salient observations from the above results of GAP ranges for different resolutions and queuing disciplines. We can see that higher resolutions are more sensitive to degrading network QoS conditions as evident from the narrow ranges of jitter and loss when compared to lower resolutions. For example, the loss range for Good grade

TABLE 4: Jitter and loss GAP ranges under PFIFO.

Display	Metric	Good	Acceptable	Poor
QCIF	Jitter (ms)	[0–200]	(200–400)	(> 400]
	Loss (%)	[0–2)	(2–4.4)	(> 4.4]
QVGA	Jitter (ms)	[0–200]	(200–350)	(> 350]
	Loss (%)	[0–1.4)	(1.4–2.8)	(> 2.8]
SD	Jitter (ms)	[0–175)	(175–300)	(> 300]
	Loss (%)	[0–0.6)	(0.6–2.5)	(> 2.5]
HD	Jitter (ms)	[0–125)	(125–225)	(> 225]
	Loss (%)	[0–0.3)	(0.3–1.3)	(> 1.3]

TABLE 5: Jitter and loss GAP ranges under TFIFO.

Display	Metric	Good	Acceptable	Poor
QCIF	Jitter (ms)	[0–50)	(50–80)	(>80]
	Loss (%)	[0–2)	(2–4.4)	(>4.4]
QVGA	Jitter (ms)	[0–40)	(40–70)	(>70]
	Loss (%)	[0–1.4)	(1.4–2.8)	(>2.8]
SD	Jitter (ms)	[0–30)	(30–60)	(>60]
	Loss (%)	[0–0.6)	(0.6–2.5)	(>2.5]
HD	Jitter (ms)	[0–20)	(20–50)	(>50]
	Loss (%)	[0–0.3)	(0.3–1.3)	(>1.3]

is [0–0.3) for HD, whereas the same for QCIF is [0–2). Consequently, we can conclude that higher resolution streams in IPTV deployments demand notably higher QoS levels than lower resolution streams.

Also, we can see that the PFIFO queuing makes the IPTV streams more tolerant to network jitter compared to TFIFO as evident from the higher ranges of jitter at all resolutions. For example, the jitter range for Good grade is [0–175) for SD under PFIFO, whereas the same for TFIFO is [0–30). Thus, we can conclude that having PFIFO queuing disciplines in routers at congestion points in the network or at the edges of access networks can reduce the burden of in-ordering packets for media player playback at the consumer sites. This in turn significantly increases the multimedia QoE resilience at the consumer sites towards higher network jitter levels.

4.1.3. Cross-Codec Redundancy Elimination. After the test case reduction by GAP network sample space identification, we reduced the sample space considerably by focusing on the network jitter and loss ranges that are relevant. However, due to the large number of application factors considered, the number of test cases still remains large. Specifically, owing to the 9 network conditions, 4 resolutions, 5 bit rates, and 3 codecs, we are left with 540 test cases for both PFIFO and TFIFO queuing disciplines. Obviously, it is infeasible to perform subjective testing on 540 test cases. In order to reduce the number of subjective test cases even further, we consider a “cross-codec elimination” scheme.

This scheme compares two test cases under the same network condition, resolution, and bit rate for a given queuing discipline but with different codecs and determines

if they are equivalent or different. The method we used to determine test case equivalence is based on the difference of the PIR objective QoE measurements for the two test cases under consideration. The lesser the difference in the PIR values, the greater is the likelihood that both the test cases have the same user QoE. Consequently, if two test cases are equivalent, we perform subjective testing only for one of the test cases, and assign the same MOS ranking to the other test case. If they are different, we perform subjective testing for both the test cases to obtain MOS rankings.

Since the PIR values increase with increasing severity of the network conditions as evident from Figures 6 and 7, the PIR difference ranges are relatively smaller at Good network conditions and relatively large at Poor network conditions. Hence, we derive “threshold curves” for each of the GAP curves such that the threshold curve values are 0.3 times the GAP curve values. We chose the 0.3 value because we found that differences in PIR below this value are not perceptibly apparent to human observers. If the PIR difference between two cases falls below the threshold curve, we consider the two test cases to be equivalent. Whereas, if the PIR difference of the two test cases falls above the threshold curve, we consider the two test cases to be different. By employing this scheme, we reduced the subjective test cases from 540 to 280 (48% reduction) for the PFIFO queuing discipline, and from 540 to 322 (40% reduction) for the TFIFO queuing discipline.

4.2. Subjective Testing. Our subjective testing was based on online network condition emulation and not based on using presaved clips as done in most other works [13]. This enabled us to inject frame-freeze events intermittently as explained in Section 3.

4.2.1. Subjective Metrics. The collection of the subjective MOS rankings during voting periods of human subject experiments was conducted in compliance with the ITU-T P.911 standard. Note that we measured “relative MOS” and not “absolute MOS” for the different video resolutions. The absolute MOS depends on the resolution, whereas the relative MOS does not. For instance, if the absolute MOS under a close to ideal network condition for a HD resolution is 4.8, then the absolute MOS for QCIF resolution will be 4.2. By measuring relative MOS, under the same network condition, MOS will be 4.8 for both HD and QCIF resolutions. This way, video sequence transcoding to other resolutions does not impact MOS ranking for the different resolutions. The MOS rankings provided by the human subjects during voting periods were on a subjective quality scale of 1 to 5 as shown in Figure 8.

4.2.2. Compliance with ITU-T P.911. For compliance with ITU-T P.911, we developed a test administration application for the human subject experiments. The application uses the Java Socket API and consists of client and server modules that are installed on the VLC server machine and Netem machine, respectively. Upon indication of the start of a test case by the test administrator at the VLC server, the server module selects a random test video sequence for streaming to

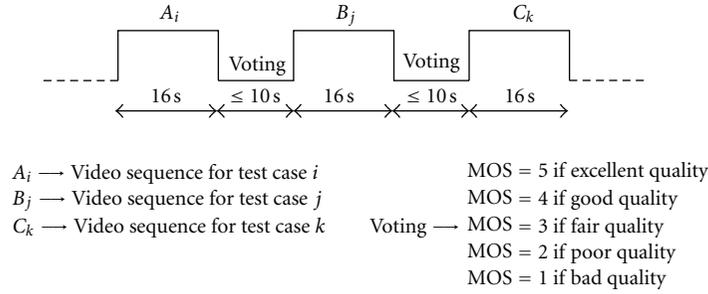


FIGURE 8: Stimulus presentation for ACR MOS collection as per ITU-T P.911.

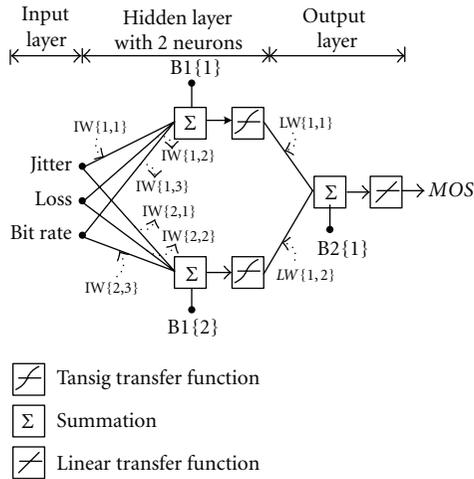


FIGURE 9: Neural network architecture showing model parameters.

the VLC client and also notifies the Netem client to configure the corresponding test case's network condition. We adopted the most commonly used “absolute category rating” (ACR) method specified in the ITU-T P.911, which is the most commonly used method in subjective testing. The ACR method shown in Figure 8, which is a single stimulus method where test video sequences are presented one at a time using the test administration application and the human subjects rate independently. Our test administration application that implemented the ACR method had a time pattern for video sequence stimulus presentation to the human subjects.

We divided human subjects into two groups for model construction: (i) 10 human subjects for QCIF and QVGA resolution test cases and (ii) 10 human subjects for SD and HD resolution test cases. Note that ITU-T recommends 4 as a minimum number of human subjects needed for statistical soundness [13] in human subject experiments. Each of the human subjects were provided with participation instructions that included purpose, procedure, potential risks, expected duration, confidentiality protection, and legal rights. The human subjects were given generous breaks to avoid fatigue, and the testing time per human subject varied between 60–80 minutes.

5. Neural Network Modeling

We now present our modeling methodology based on neural network principles to derive closed-form expressions of multimedia QoE. After populating the human subjects MOS rankings database, we use the database to train our psycho-acoustic-visual model. We employ neural networks as our modeling technique using the Matlab Neural Network Toolbox [22]. Neural networks are essentially a system of adjustable parameters called “Weights” (i.e., IW—Initialization Weight, LW—Load Weight) and “Biases” (i.e., B). The parameters are adjusted during the training of the neural network, which results in specific weights and biases. A scalar input is multiplied by the effective weight and added to the bias to produce the target output. The network is adjusted based on a comparison of the output and the target, until the network output matches the target.

The type of neural network used in our study is a “feed-forward network”, whose architecture is shown in Figure 9. The neural network consists of an input layer, hidden layer with 2 neurons, and an output layer. The output layer acts as a normalizing segment and is of the `Purelin` type, which covers the entire output range. Our feed-forward modeling schema allowed us to approximate both linear and nonlinear data functions. Two hidden neurons are used to perform computation of weights and biases paying due considerations to the trade-off between speed and accuracy. It is known that greater the number of hidden layers, greater is the time taken by the model to compute the weights and biases, and greater is the accuracy. The `Tansig` transfer function (covers range $-1-1$) is used to facilitate accurate modeling of nonlinear aspects of the data process. The type of training function we used is the `Trainlm` (Levenberg-Marquardt backpropagation algorithm), which produces suitable accuracy for the size of the network in our problem. It cuts off the model training (i.e., the number of epochs) before the model becomes overtrained with a given data set, and learns much faster than the traditional training functions such as `Traingd` (Gradient Descent backpropagation algorithm).

The input parameters we modeled are: (i) jitter, (ii) loss, and (iii) bit rate. The model output is the MOS ranking in the range (1–5). We developed multi-resolution neural network models for each codec combination and queuing disciplines. We developed a total of 12 models corresponding to the 3 codecs for each of the 4 resolutions

Model parameters:

IW			B1	LW		B2
IW{1, 1}	IW{1, 2}	IW{1, 3}	B1{1}	LW{1, 1}	LW{1, 2}	B2{1}
IW{2, 1}	IW{2, 2}	IW{2, 3}	B1{2}			

MOS calculation:

$$\alpha = \text{loss} * (\text{IW}\{1,1\}) + \text{jitter} * (\text{IW}\{1,2\}) + \text{bit rate} * (\text{IW}\{1,3\}) + \text{B1}\{1\}$$

$$\beta = \text{loss} * (\text{IW}\{2,1\}) + \text{jitter} * (\text{IW}\{2,2\}) + \text{bit rate} * (\text{IW}\{2,3\}) + \text{B1}\{2\}$$

$$\gamma = \tanh(\alpha)$$

$$\zeta = \tanh(\beta)$$

$$\boxed{\text{MOS} = \gamma * (\text{LW}\{1,1\}) + \zeta * (\text{LW}\{1,2\}) + \text{B2}\{1\}}$$

FIGURE 10: MOS calculation using neural network model parameters.

TABLE 6: TFIFO model parameters showing weights and biases for QCIF and QVGA resolutions.

Params.	QCIF MPEG-2	QCIF MPEG-4	QCIF H.264	QVGA MPEG-2	QVGA MPEG-4	QVGA H.264
IW{1, 1}	1.5925	73.9938	0.54	-1.805	0.8271	-53.9923
IW{1, 2}	-6.6656	99.6796	0.2806	-1.1208	0.3625	0.9155
IW{1, 3}	-0.4831	-30.1671	0.0507	-17.6184	-0.0008	8.7659
IW{2, 1}	0.5170	0.7311	289.0135	-0.7044	-2.9727	-1.249
IW{2, 2}	0.5438	0.4927	88.5316	-0.3372	-16.2994	-0.5441
IW{2, 3}	0.0400	0.0803	-25.5375	-0.009	-19.3425	-0.0562
B1{1}	1.7810	65.876	0.6902	-2.9083	1.0195	-31.2057
B1{2}	1.1284	-0.2967	53.7415	-2.3335	48.4183	0.3567
LW{1, 1}	-0.7164	-0.3313	-11.176	-133.596	-19.0374	0.2548
LW{1, 2}	-31.2439	-3.2402	-0.7112	271.4444	0.2697	3.5383
B2{1}	31.5663	4.6002	12.7213	138.3723	19.9051	4.51

for each queuing discipline. Tables 6 and 7 show the model parameters for the 12 TFIFO queuing discipline models. Tables 8 and 9 show the model parameters for the 12 PFIFO queuing discipline models. Figure 10 shows how the MOS calculation is performed using the neural network model parameters. We remark that we used the Matlab Neural Network Toolbox's "training state" plots to terminate the model training and derive the weight and bias parameters that are shown in Tables 6–9; the plots help us to identify the weight and bias parameters that do not overtrain the model with a given data set.

6. Performance Evaluation

In this section we present our performance evaluation methodology used and results obtained after critical examination of our models. The results correspond to our models prediction performance in terms of overall characteristics, accuracy, consistency, and speed.

6.1. Model Characteristics. We verified the characteristics of all our multi-resolution models by observing the MOS predictions for test cases with systematically increasing jitter and loss inputs, at different bit rate settings in the resolution-specific ranges. The observations were made by plotting 3-dimensional graphs of jitter, loss, and model predicted MOS.

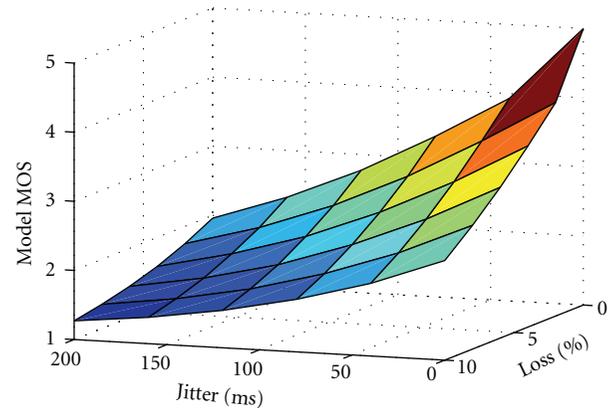


FIGURE 11: TFIFO QCIF model characteristics at 512 Kbps bit rate.

Figures 11 and 12 show example graphs, which correspond to the MOS predictions of the TFIFO models for QCIF and HD resolutions, respectively, at 512 Kbps bit rate. In each model case, we were able to confirm that the MOS predicted by the model decreases with the increase in the jitter and loss levels. Also, the predicted MOS was 5 at the best network condition with 0 jitter and 0 loss; it decreased with increasing jitter and loss and reached 1 at and after the highest value of jitter and loss used. Thus, owing to our above systematic evaluation of

TABLE 7: TFIFO model parameters showing weights and biases for SD and HD resolutions.

Params.	SD MPEG-2	SD MPEG-4	SD H.264	HD MPEG-2	HD MPEG-4	HD H.264
IW{1, 1}	-235.618	69.198	1.9839	-2.705	18.0678	13.2719
IW{1, 2}	-59.4654	66.7891	2.24103	-0.7994	-4.0983	-4.5941
IW{1, 3}	-0.0141	16.6377	0.2509	-0.0137	2.0437	-0.0316
IW{2, 1}	1.3085	-1.5325	-272.955	0.9113	-2.7116	2.9783
IW{2, 2}	1.4055	-0.9056	-69.1575	-6.7765	-0.7866	0.276
IW{2, 3}	0.1048	-0.0147	-0.0183	-4.5513	-0.0187	0.0075
B1{1}	30.533	-127.695	-2.9503	-1.4931	-14.7136	5.0295
B1{2}	-1.4295	-1.6538	35.7638	1.5997	-0.6561	1.6912
LW{1, 1}	0.8211	-0.2376	-0.7411	51.6476	0.129	65.2392
LW{1, 2}	-1.2851	62.1292	1.194	-26.1423	12.0581	-70.5904
B2{1}	3.0135	63.3574	2.9553	26.7493	13.1543	6.2374

TABLE 8: PFIFO model parameters showing weights and biases for QCIF and QVGA resolutions.

Params.	QCIF MPEG-2	QCIF MPEG-4	QCIF H.264	QVGA MPEG-2	QVGA MPEG-4	QVGA H.264
IW{1, 1}	29.5296	5.1324	5.4489	0.0934	-23.7788	-4.4705
IW{1, 2}	-53.6772	-1.3168	27.4114	0.1402	4.3746	2.3579
IW{1, 3}	0.7985	-5.0446	-1.5200	0.0025	8.4277	0.0110
IW{2, 1}	-0.0350	0.0167	-0.0479	5.6176	-0.0377	-0.0672
IW{2, 2}	-0.1508	0.0736	-0.2183	-0.6570	-0.1304	-0.0397
IW{2, 3}	-0.0210	0.0131	-0.0360	-26.1136	-0.0205	-0.0015
B1{1}	-15.6984	10.9947	-2.2131	-0.2708	19.0489	10.5100
B1{2}	0.8192	-0.9046	1.4721	16.4723	0.3639	1.0951
LW{1, 1}	0.1918	-0.0691	-0.3595	-3.5616	0.2382	-1.6414
LW{1, 2}	3.1491	-6.7489	2.2291	0.5117	3.3048	17.6250
B2{1}	2.5604	-0.3827	2.4283	4.4878	3.6572	-7.5768

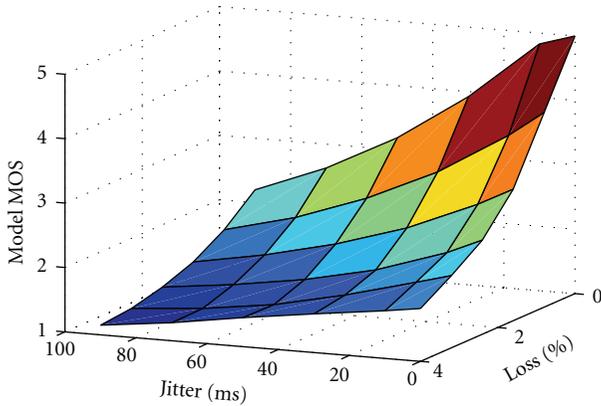


FIGURE 12: TFIFO HD model characteristics at 512 Kbps bit rate.

the model prediction characteristics for a comprehensive set of control inputs, we can conclude that the model prediction behavior follows expected patterns.

6.2. Model Accuracy and Consistency. In order to validate the accuracy and consistency of our models, we choose 60 validation test cases that encompass video sequences and network conditions that are different from the ones used in the model training. Given that our model construction:

(a) considered additional factors (e.g., PFIFO/TFIFO router queuing disciplines, online network emulation, frame-freeze impairment events) and (b) was targeted specifically at multi-resolution video sequence-based IPTV applications, it is not reasonable to compare our models with existing real-time QoE models such as [13, 17, 18] that have different considerations and application contexts. Hence, we took a poll of 10 human subjects by administering the 60 validation test cases to collect Validation MOS (V-MOS) rankings using the same testbed setup explained in Section 3. We compared the V-MOS rankings with the corresponding test case model MOS predictions (M-MOS). The metrics: (i) correlation coefficient (r) and (ii) root mean square error (rmse) are used to evaluate model prediction accuracy, and the outlier ratio (or) metric is used to evaluate model consistency.

The correlation coefficient r of the pairs (x, y) , where x is the V-MOS and y is the M-MOS, is calculated as follows:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}. \quad (1)$$

Having r as close to 1 indicates that the model MOS ranking predictions closely match the human subject MOS rankings.

TABLE 9: PFIFO model parameters showing weights and biases for SD and HD resolutions.

Params.	SD MPEG-2	SD MPEG-4	SD H.264	HD MPEG-2	HD MPEG-4	HD H.264
IW{1, 1}	0.0774	8.2689	6.0167	0.1001	0.0470	18.6600
IW{1, 2}	0.3627	-47.4086	10.5060	0.9631	0.2612	1.4276
IW{1, 3}	0.0256	2.1463	0.0344	0.0179	0.1268	3.4165
IW{2, 1}	0.6096	163.2712	14.4661	-0.1703	-0.5565	-0.4194
IW{2, 2}	-3.0682	46.1962	14.4351	-0.8077	-0.0374	-0.1042
IW{2, 3}	-0.1490	-35.2372	0.1807	-0.0177	0.0179	-0.0079
B1{1}	-0.6770	-18.4515	0.7189	-1.2166	-1.0055	-48.3245
B1{2}	3.4885	-129.5034	-41.3358	1.0665	-1.1830	-0.9670
LW{1, 1}	-3.5498	0.0098	-43.2699	15.4020	-0.9678	-0.2964
LW{1, 2}	-0.7699	-0.9262	-0.9502	18.0452	12.9946	13.3309
B2{1}	3.7081	3.3711	46.3764	3.8736	15.0390	14.9439

The difference between the V-MOS and M-MOS for test case i is defined as the absolute prediction error P_{error} given by

$$P_{\text{error}}(i) = \text{MOS}(i) - \text{MOS}_p(i). \quad (2)$$

The rmse of P_{error} is calculated as follows:

$$\text{rmse} = \sqrt{\frac{1}{N-d} \sum_N P_{\text{error}}[i]^2}, \quad (3)$$

where N denotes the number of samples and d denotes the mapping function's number of degrees of freedom.

The outlier ratio (or) is defined as the ratio of "outlier points" to total points N and is calculated as follows:

$$\text{or} = \frac{\text{Total Number of Outliers}}{N}, \quad (4)$$

where an outlier is defined as a point for which -

$$|P_{\text{error}}(i)| > 2 * \sigma(\text{MOS}(i)), \quad (5)$$

where $\sigma(\text{MOS}(i))$ represents the standard deviation of the individual scores associated with sample i .

The performance of the model predicted MOS (M-MOS) to the human subject MOS rankings for the validation test cases (V-MOS) can be seen from Figure 13 for the TFIFO queuing discipline and Figure 14 for the PFIFO queuing discipline. The accuracy of the model is evaluated by determining the correlation coefficient between the M-MOS and V-MOS rankings for corresponding network conditions. The results of the average r and rmse for the TFIFO and PFIFO models are presented in Table 10. It is evident from the results that even in the worst-case scenarios, the M-MOS is very close to V-MOS because r is >0.9 and rmse is <0.7 across all the 4 resolutions for both TFIFO and PFIFO models. Upon checking the consistency of the model using the or metric, we observed that all of the models had zero outliers.

There is further evidence that the model predictions are consistent with what actual users experience while using multi-resolution IPTV applications on the Internet.

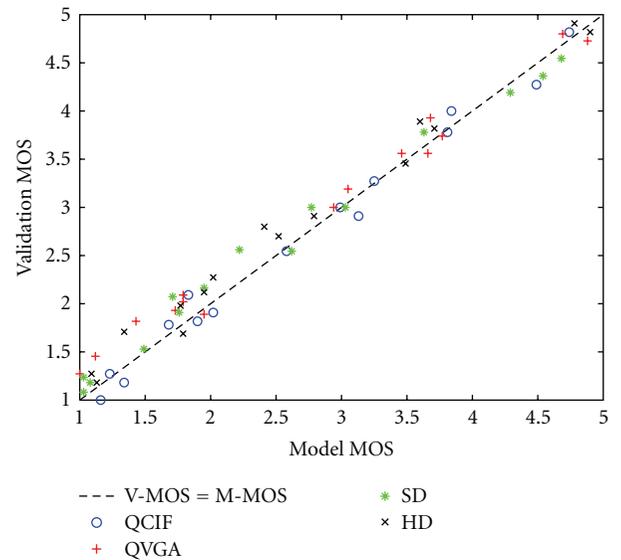


FIGURE 13: Correlation between TFIFO Model MOS and TFIFO Validation MOS.

TABLE 10: Accuracy results for TFIFO and PFIFO models.

Resolution	r (TFIFO)	rmse (TFIFO)	r (PFIFO)	rmse (PFIFO)
QCIF	0.98	0.24	0.98	0.37
QVGA	0.98	0.61	0.94	0.51
SD	0.94	0.60	0.92	0.55
HD	0.96	0.56	0.91	0.55

6.3. Model Speed. We evaluate the prediction speed of the models to determine if they can be successfully deployed on devices such as routers, gateways, set-top boxes, or measurement servers in a scalable manner. Specifically, even in large-scale IPTV content delivery networks, where QoE monitoring points process several tens to thousands of IPTV flows, the model speed should be small enough that it does not overburden a device's processor. The flows correspond to the measurements of numerous network paths on which critical IPTV services are provided to consumer

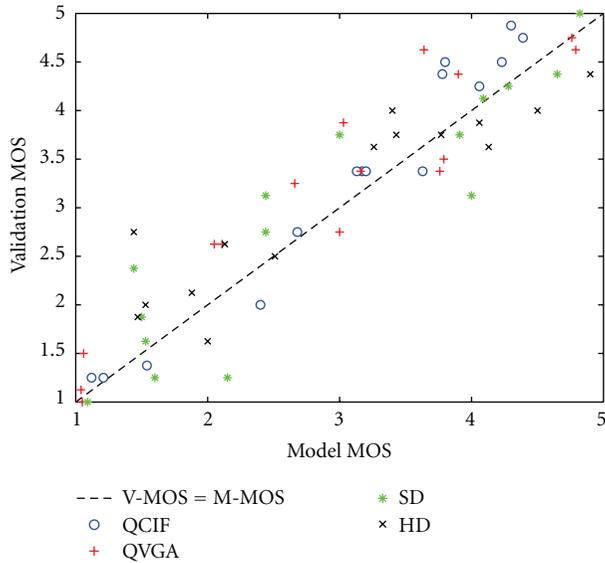


FIGURE 14: Correlation between PFIFO model MOS and PFIFO validation MOS.

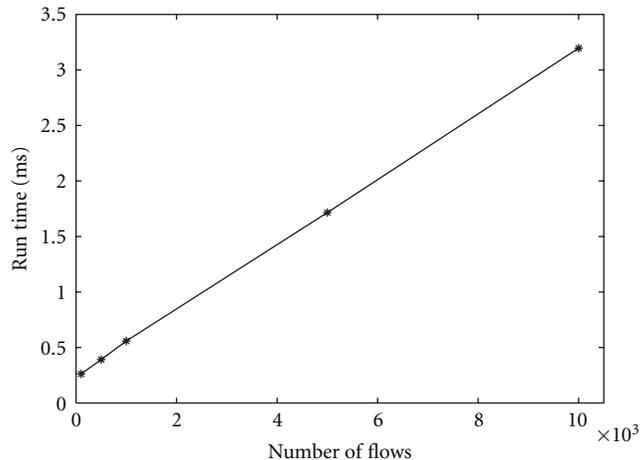


FIGURE 15: Model speed for increasing number of flows.

groups. To evaluate the model speed, we conduct a run time analysis simulation in which we calculate the amount of time taken by any of the neural network models to predict the MOS rankings for increasing number of flows with random network conditions. Four separate test runs of this simulation are conducted, and the average run-time is calculated. Representative model run time results are shown in Figure 15. From the results, we can see that even for the case of online prediction of MOS rankings for 10,000 flows, the model speed is <4 ms. Thus, the prediction speed of any of the models is negligible, making them suitable for integration even in embedded-monitoring protocols such as RTCP-extended reports (RTCP-XR) [28] in considerably large-scale IPTV content delivery networks.

7. Conclusion

In this paper, we presented a novel methodology to develop *real-time* and *zero-reference* multimedia QoE models for IPTV applications. The models were developed using neural network principles for multiple resolution (QCIF, QVGA, SD, HD) video sequences streamed with popular codec combinations (MPEG-2 video with MPEG-2 audio, MPEG-4 video with AAC audio, and H.264 video with AAC audio) and bit rates for different network health conditions. Also, we showed the impact of PFIFO and TFIFO router queuing disciplines on multimedia QoE in IPTV content delivery networks and developed separate models for each of them. The models developed can be used for online QoE estimation given measurable network factors such as jitter and loss. To develop the models, we proposed novel test case reduction schemes, namely, network sample space identification and cross-codec redundancy elimination that provided a manageable sample space to collect MOS rankings from human subject experiments in compliance with the ITU-T P.911 recommendation. Using our test case reduction schemes and two novel objective QoE metrics at the user-level (PIR) and network-level (FPL), we were able to reduce 1080 test scenarios for the modeling to 602 test cases without comprising model coverage. We evaluated our models prediction performance in terms of overall characteristics, accuracy, consistency, and speed. In worst-case scenarios, we observed our model predicted MOS rankings to have >0.9 correlation coefficient with <0.7 root mean square error and zero outlier ratio when compared with actual human subject validation MOS rankings. Also, our model speed results suggest that over 10,000 flows can be handled in <4 ms. Thus, our impressive performance of models demonstrate their suitability to be used for: (a) continuous multimedia QoE monitoring in devices such as routers, gateways, set-top boxes, or measurement servers and (b) real-time adaptation of system and network resources, in small-to-large scale IPTV content delivery networks.

References

- [1] E. Jopling and A. Sabia, "Forecast: IPTV subscribers and service revenue, worldwide, 2004–2010," *Gartner IPTV Market Report*, 6 pages, 2006.
- [2] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid—A framework for video transmission and quality evaluation," in *Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 255–272, 2003.
- [3] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," 2001.
- [4] "Subjective audiovisual quality assessment methods for multimedia applications," ITU-T Rec. P.911, 1998.
- [5] P. Calyam, M. Haffner, E. Ekici, and C. G. Lee, "Measuring interaction QoE in Internet videoconferencing," *IEEE MMNS*, vol. 4787, pp. 14–25, 2007.
- [6] G. M. Muntean, P. Perry, and L. Murphy, "Subjective assessment of the quality-oriented adaptive scheme," *IEEE Transactions on Broadcasting*, vol. 51, no. 3, pp. 276–286, 2005.

- [7] X. Lu, S. Tao, M. Zarki, and R. Guerin, "Quality-based adaptive video over the Internet," in *Proceedings of the CNDS*, 2003.
- [8] M. Ghanbari, D. Crawford, M. Fleury et al., "Future performance of video codecs," in Ofcom Research Report (SES2006-7-13), 2006.
- [9] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross, "A measurement study of a large-scale P2P IPTV system," *IEEE Transactions on Multimedia*, vol. 9, no. 8, pp. 1672–1687, 2007.
- [10] M. Claypool and J. Tanner, "The effects of jitter on the perceptual quality of video," in *Proceedings of the ACM Multimedia*, pp. 115–118, New York, NY, USA, 1999.
- [11] Y. Won and M. Choi, "End-User IPTV traffic measurement of residential broadband access networks," in *Proceedings of the IEEE Workshop on End-to-End Monitoring Techniques and Services*, pp. 95–100, Salvador, Bahia, Brazil, 2008.
- [12] S. Winkler and R. Campos, "Video quality evaluation for Internet streaming applications," *Human Vision and Electronic Imaging*, vol. 3, pp. 104–115, 2003.
- [13] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1071–1083, 2002.
- [14] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
- [15] "The E-Model: a computational model for use in transmission planning," ITU-T Rec. G.107, 1998.
- [16] A. P. Markopoulou, F. A. Tobagi, and M. J. Karam, "Assessment of VoIP quality over Internet backbones," in *Proceedings of the IEEE Infocom*, pp. 150–159, June 2002.
- [17] S. Tao, J. Apostolopoulos, and R. Guérin, "Real-time monitoring of video quality in IP networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1052–1065, 2008.
- [18] F. Massidda, D. D. Giusto, and C. Perra, "No reference video quality estimation based on human visual system for 2.5/3G devices," in *Proceedings of the SPIE and T Electronic Imaging*, pp. 168–179, January 2005.
- [19] H. J. Kim and S. G. Choi, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," in *Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT '10)*, pp. 1377–1382, February 2010.
- [20] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra, "Predicting quality of experience in multimedia streaming," in *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia, (MoMM '09)*, pp. 52–59, December 2009.
- [21] S. Balasubramaniam, J. Mineraud, P. McDonagh et al., "An evaluation of parameterized gradient based routing with QoE monitoring for multiple IPTV providers," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 183–194, 2011.
- [22] H. Demuth and M. Beale, *Matlab Neural Network Toolbox User's Guide*, The MathWorks, 2000.
- [23] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 78–84, 2008.
- [24] "VLC media player," <http://www.videolan.org/vlc>.
- [25] S. Hemminger, "Netem—Emulating real networks in the lab," in *Proceedings of the Linux Conference*, Australia, 2005.
- [26] J. Bellardo and S. Savage, "Measuring packet reordering," in *Proceedings of the 2nd ACM SIGCOMM Internet Measurement Workshop (IMW '02)*, pp. 97–105, November 2002.
- [27] P. Calyam, M. Sridharan, W. Mandrawa, and P. Schopis, "Performance measurement and analysis of h.323 traffic," in *Proceedings of the Passive and Active Measurement Workshop*, vol. 3015, pp. 137–146, 2004.
- [28] T. Friedman, R. Caceres, and A. Clark, "RTP control protocol extended reports (RTCP XR)," IETF RFC 3611, 2003.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

