Nonlinear Auto-Regressive Neural Network Model for Forecasting Hi-Def H.265 Video Traffic Over Ethernet Passive Optical Networks

Collin Daly, David L. Moore, Rami J. Haddad
Department of Electrical Engineering
Georgia Southern University
Statesboro, Georgia 30458
Email: {cd03292, dm05190, rhaddad}@georgiasouthern.edu

***Abstract*—Video bandwidth forecasting can help optimize the transmission of video traffic over optical access networks. In this paper, we propose the use of a nonlinear auto-regressive (NAR) neural network model for forecasting H.265 video bandwidth requirements to optimize video transmission within Ethernet Passive Optical Networks (EPONs). The video's constituent I, P, and B frames are forecast separately to improve model forecasting accuracy. The proposed forecasting model is able to forecast H.265 encoded High-Definition videos with an accuracy exceeding 90%. In addition, using the video bandwidth requirement predictions as grant requests within EPONs improved the efficiency of dynamic bandwidth allocation (DBA). The use of nonlinear autoregressive neural network grant sizing predictions within EPONs reduced the video packet queueing delay significantly when the network was saturated near capacity.**

***Keywords*—Video prediction, H.265, video forecasting, EPON, Nonlinear Auto-regressive, NAR, neural network.**

I. Introduction

Cisco Virtual Networking Index (VNI) predicts that IP video traffic in all its forms will constitute 82% (1.89 zettabytes) of the total global IP traffic by the year 2020 [1]. As this trend is expected to continue, the networking demands of multimedia video will increase due to more widespread consumer demand. As a result of the sheer bulk of videos sent through the network, the amount of bandwidth required to transmit video continues to grow with advances in recording and playback quality. Newly proposed standards for ultra-high definition (UHD) video, such as the High Efficiency Video Coding (H.265) standard for 4K and 8K video, are gaining popularity and being adopted for use due to higher video resolution compared with previous standards such as H.264. These high resolution standards have high transmission bandwidth requirements, which necessitate implementing either more efficient compression techniques or higher bandwidth communication infrastructure. Multiple studies have determined there is a 40-52% reduction in the size of a compressed H.265 encoded file as compared to the corresponding H.264 encoded file [2], [3], [4]. However, these compression algorithms generate a large variance in the video frame sizes which, in turn, reduces the transmission bandwidth utilization due to the inconsistent grant sizing requirements [5].

The current global deployment rate of optical fiber is constantly increasing in an effort on the part of internet service providers to stay ahead of the growing bandwidth demand caused by the increased commonality of transferring large files and streaming videos to and from mobile communications equipment. If there is not a drastic decrease in networking demand nor a fundamental shift in transmission paradigms, the demand on the network will far exceed the ability of the network to deliver data, and the deployment of optical fiber will not be able to keep pace with the increasing demand. Actively amplified optical fiber is used to connect large cities and datacenters across long distances to maintain Ethernet network continuity. These systems are expensive to install and maintain due to the cost of active optical devices.

To create a link between service providers and consumers that is capable of delivering the speed and performance characteristics associated with fiber optic networks, industry professionals have created Ethernet passive optical networks (EPONs) that are capable of spanning tens or hundreds of kilometers without the need for active optical devices. These EPONs can be used in place of the existing copper cable infrastructure, as fiber is not susceptible to the same level of interference and attenuation which degrades data integrity in a copper network over a much shorter distance than a typical EPON spans. In addition, the available transmission bandwidth is orders of magnitude higher than copper cable, minimizing the bottleneck effect of the delivery network. EPONs are often deployed directly to the house or to the curb, limiting copper cable to inside the house and to the short distance between the curb and the house. These short runs of copper cable can be used for higher bandwidth and faster speeds than is possible over a longer run, so the inherent service degradation of copper is not noticed by the end user.

EPONs consist of three main components: the optical line terminal (OLT), the optical network units (ONU), and the passive optical fiber splitter. Optical fiber forms the communications channel between the OLT and ONUs. A laser at each end of the fiber medium is aimed so that all light is reflected internally and confined to a path along the length of the fiber. Figure 1 depicts the basic components and layout

978-1-5386-1539-3/17/$31.00 ©2017 IEEE

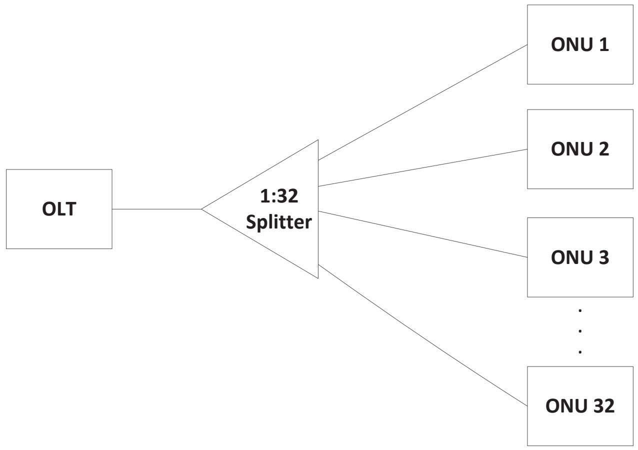


Fig. 1. EPON with 32 ONUs

of the EPON simulated for this paper. The fiber splitter is used to send an incoming pulse of light from the OLT to all attached output fibers and the associated ONUs. Additionally, it is used to route any packets coming from the ONUs directly to the OLT. This allows one OLT to broadcast instructions to all of the connected ONUs. Packets are addressed to the ONUs and the packets not addressed to the ONU are ignored. Packets received by the OLT are relayed to the next section of the network. While a simple setup with similar lasers and hardware is sufficient to service several clients, the very property that makes the splitter useful also has a deleterious effect on the efficiency of the network. If more than one ONU transmits packets at the same time to the OLT, the packets will collide at the splitter/combiner, and the message is rendered unintelligible.

Limiting upstream and downstream communications to one wavelength each way remedies the hardware issues presented by multiple clients but requires timing instructions to be sent to the downstream units by a central controller. This is accomplished by setting a timing structure which is regulated by the OLT serving as a medium access controller (MAC). The MAC is designed to allow only one transmission from any ONU at any time, eliminating the possibility of conflicts or destruction of data. Available bandwidth is allocated to every ONU in the system. The simplest method of allocation is to award a fixed grant to each ONU, regardless of the length of the transmission queue at each or the utilization of the ONU in question at the time [6]. This centralized protocol operates utilizing a cyclical polling process to service all the optical network units. However, this polling process can increase the upstream queueing delay, especially if packets arrive just after polling is done, which is very critical when streaming delay sensitive traffic such as video traffic [7].

Many different forecasting algorithms have been proposed to decrease this delay. Auto-Regressive models were used in [8], [9], [10], and so were Moving Average models [11], and Auto-Regressive Moving Average models [12]. These models were able to forecast video bandwidth, but with relatively low accuracy. Some other techniques that were utilized include time lagged feed-forward neural networks [13], [14], [15], [16] and Recurrent neural networks [17], [14], [18], [19]. Feed-Forward Bandwidth Indication (FFBI), which is a technique that feeds forward video bandwidth requirements into the header packets to be used as an accurate predictive model, has shown to optimally reduce video queueing delay [20]. However, FFBI has limitation when used with live interactive video which the technique proposed here can help resolve. In this paper, an exhaustive grant sizing scheme using Nonlinear Auto-Regressive (NAR) neural network based queue size prediction is proposed as a solution to eliminate this additional delay. The proposed technique predicts the constituent video frame types separately with optimized neural networks to increase the prediction accuracy.

The rest of the paper is organized as follows. Section II discusses the proposed approach in details. Section III discusses the method and procedure. Section IV discusses the simulation model and the results. Finally, Section V concludes the paper with a summary of the findings.

II. Proposed Approach

In an EPON, the OLT issues grants to the ONUs based on polled requests for the next grant cycle which are appended to the end of each ONU transmission. These grants allow each ONU to transmit upstream to the OLT during a specific grant window each cycle. Since each ONU transmits at the same wavelength, it is imperative to regulate the flow of data in order to avoid destructive interference within the fiber medium. The OLT provides a transmission grant based upon the queue length of the each ONU which is provided to the OLT at the end of the previous data transmission grant. An OLT working under a limited grant sizing rule will then issue each ONU the requested grant size up to a maximum limit. While the full grant is not necessarily always issued, any requests over the limit will be allocated during subsequent grant cycles. This limited grant sizing technique creates a delay of one polling cycle between request and grant. If packets arrives at the ONU after the transmission of the grant request, the packets will have to wait up to two grant cycles to be serviced. To minimize this delay, the OLT may utilize previous data transmission grants to forecast requirements and more accurately allocate transmission grants. Multimedia traffic, such as the new H.265 ultra-high definition video standard, is often affected by the inherent cycle delay, slowing the overall progression of video traffic through the network. Accurate predictive algorithms employed in the grant allocation process provide a method to mitigate polling delay and limit the queueing delay to only one grant cycle. This will also eliminate any used portion of a grant common to techniques such as fixed grant sizing, which results in under-utilization of the available bandwidth. A NAR neural network model is used to provide video frame size forecasts for high-definition multimedia traffic because the

NAR network structure is particularly well suited to predicting time-variant series with multiple step-ahead delays.

III. Experimental Procedure

A set of experiments were conducted to test the effect of using the NAR forecast model on the EPON queueing delay performance using exhaustive service at various video quantization levels. Four H.265 encoded video trace files from the Arizona State University video trace library were utilized as inputs to the video transmission queues [21]. The four video traces used are: Finding Neverland, Harry Potter, Lake House, and Speed. Each video trace is tested with quantization levels of 20, 25, 30, 35, 40 and 45, which correspond to the amount of video compression. The video files are formated using the Full High Definition (FHD) with 1920x1080 resolution and a frame rate of 24fps. The group of pictures (GoP) structure for all the videos is G24B7, consisting of identity (I), predictive (P) and bi-directionally predictive (B) frames, represented as: IBBBBBBBPBBBBBBBPBBBBBBB.

Each of these video trace files are broken up into input files based on frame type. This process generated an I, P, and B video dataset. In general, neural networks do not model trends within the time series data very well, so separating the video frames as sets of I, B, and P frames detrends the video data, stabilizing the mean and variance for each set and improving the modeling efficiency of the neural network. Separate single-frame-ahead prediction neural networks are used to forecast each of the resulting video frame sets. The first 50% of each video is used for training while the rest is used for testing. The neural network chosen for the process is a NAR model due to the time-varying nature of the video frame sizes. The network uses a fixed delay of 10, 15, and 20 frames for I, P, and B frames, respectively, with a general network structure of one input layer, two hidden layers, and one output layer. The two hidden layers consist of 11 and 6 neurons respectively, as illustrated in Figure 2. The number of hidden layers is maintained at 2 to avoid reducing the generalization performance by over-fitting.

Once the predicted values are generated for each frame type set, the results are recombined back into the trace files as next-frame predictions. The traces are then encapsulated as Ethernet packets to allow network simulation. These packets are inputs for an EPON simulator developed using the CSIM discrete event simulation library. The simulated EPON has a transmission rate of 1 Gbps and services a total of 32 ONUs. Each ONU has two distinct queues: a data queue and a video queue. A shared limited grant sizing mechanism is implemented at the OLT based upon the predicted grant requirements.

The data queues are supplied with a Poisson traffic generator to allow for rapid convergence, and the video queues are supplied with both predicted and unpredicted video trace data. The simulator was set to run at different network utilization levels to determine the effectiveness of prediction at varying levels of network utilization. The following quad modal packet size distribution was used for the Poisson traffic generator: 60% 64 bytes, 4% 300 bytes, 11% 580 bytes, and 25% 1518 bytes.

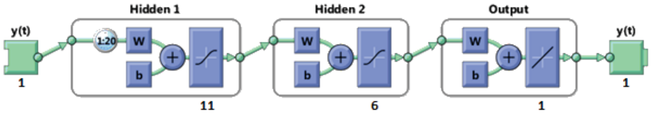


Fig. 2. NAR Artificial Neural Network

IV. Experimental Analysis

Using the Ethernet encapsulated video trace files as inputs, the EPON simulator outputs consist of data packet delay and video packet delay. To establish a benchmark, the EPON network was simulated without using prediction of any kind to provide a baseline point. Once the video and data queueing delay baselines are established for the case when no prediction is used, baselines for 10% under-prediction, 10% over-prediction, and 100% accurate prediction are established for each video trace. Finally, the video frames are forecast using the NAR neural network and compared with the baseline data to determine viability. For each of the following figures (3-9), subfigures (a) and (b) depict the video delays with video prediction compared to the video delays without video prediction. Subfigures (c) and (d) depict the data delay with video prediction compared to the data delay without video prediction. Subfigues (a) and (c) are for quantization levels 20, 25, and 30; while (b) and (d) are for quantization levels 35, 40, and 45. Simulated network saturation is depicted on the x-axis, ranging from 0.1 Gbps data traffic (10% utilization) to 0.9 Gbps data traffic (90% utilization).

A. Under-prediction Baseline

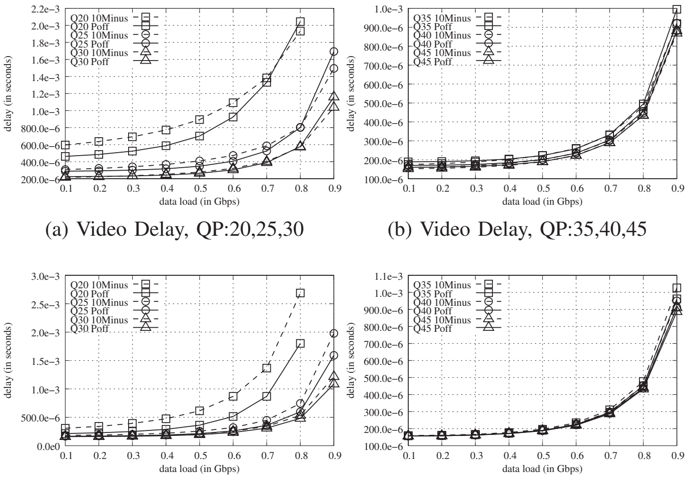
For video bandwidth forecast with 10% under-prediction of the necessary grant size, the data for Finding Neverland is indicative of the overall trends, as illustrated in Figure 3. The data delay and video delay are higher than the case without video prediction at each quantization level. It is concluded that under-prediction of video bandwidth will not reduce the video queueing delay. However, it will ensure that allocated grant is fully utilized, which improve bandwidth utilization.

B. Accurate Prediction Baseline

Video bandwidth forecast with accurate prediction data provides the idea solution to ensure minimal queueing delay with maximum bandwidth utilization. The results for the Harry Potter video are depicted in Figure 4. There is once again a large data queueing delays in illustrated in Figures 4(c) and 4(d). However, there is a higher reduction in video queueing delays observed for videos with quantization level exceeding Q30 as illustrated in Figure 4(a), Q35, Q40, and Q45 as illustrated in Figure 4(b). It is evident that the overall queueing delay (video and data) will be less than the baseline.

TABLE I
STATISTICS OF VIDEO TRACES

Video Trace Name	Quantization Parameter	Mean Bit Rate (kbit/s)	Coefficient of Variation	Data Delay Decrease (µs)	Video Delay Decrease (µs)	Overall Delay Decrease (µs)	Prediction Accuracy (%)
Finding Neverland	30	559.6184	2.317	-138	250	112	92.89
Finding Neverland	35	294.2985	2.509	-64.6	245	181	93.24
Finding Neverland	40	164.891	2.57	-36.2	228	192	93.21
Finding Neverland	45	89.8925	2.577	-20.5	220	199	92.10
Harry Potter	30	527.8378	2.247	-180	250	69.5	91.75
Harry Potter	35	273.7205	2.416	-67.4	242	175	92.18
Harry Potter	40	150.9942	2.476	-40.9	247	206	92.60
Harry Potter	45	81.0442	2.499	-20.8	243	222	90.84
Lake House	30	506.8175	3.083	-119	223	104	94.88
Lake House	35	268.3858	3.341	-61.6	220	159	93.30
Lake House	40	148.3874	3.379	-32.1	215	183	92.24
Lake House	45	78.9218	3.293	-17.5	236	218	92.54
Speed	30	603.0512	1.634	-179	300	121	91.34
Speed	35	316.6073	1.696	-84.3	254	170	90.40
Speed	40	180.9275	1.734	-47.9	225	177	90.42
Speed	45	101.2876	1.768	-28.2	223	195	91.41



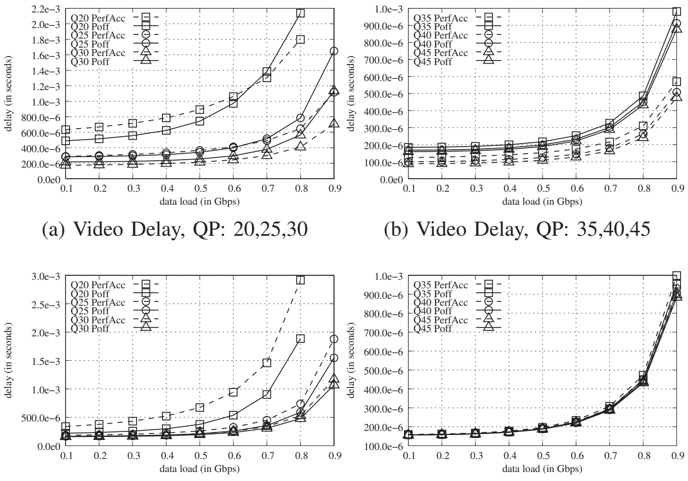
(a) Video Delay, QP:20,25,30

(b) Video Delay, QP:35,40,45

(c) Data Delay, QP:20,25,30

(d) Data Delay, QP:35,40,45

Fig. 3. Under-prediction delay for video (Finding Neverland) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.



(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

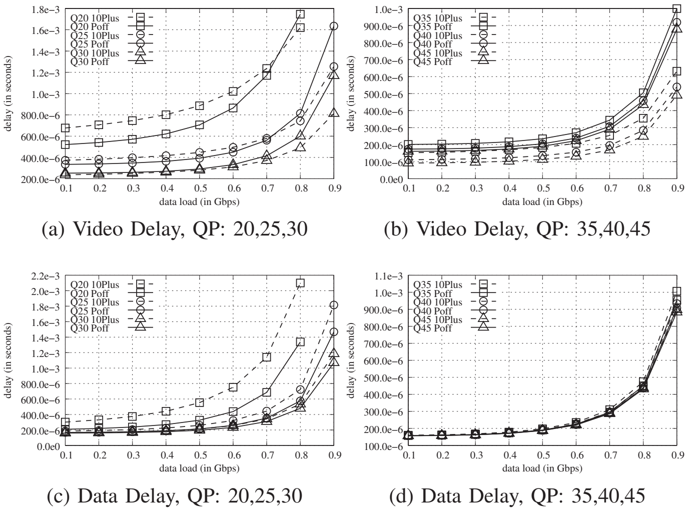
Fig. 4. Accurate prediction delay for video (Harry Potter) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.

C. Over-prediction Baseline

A trend was noted based upon the under-prediction and accurate prediction baselines, so a series of 10% over-predicted video grant sizing is generated and tested. Lake House, which is indicative of the results, is depicted in Figure 5. For lower quantization values of 20 and 25, there is a greater delay for both data and video delay (Figure 5(a) and 5(c)) compared to the case without prediction. As with the accurate prediction baseline shown previously, the values for Q30 illustrated in Figure 5(a) and Q35, Q40, Q45 as illustrated in Figure 5(b) videos has lower video queueing delays compared to those without prediction. Improvement in the system performance is inferred from this data, but this method could instead introduce unnecessary extra queueing delay by requesting a larger grant than is required and should therefore be avoided. This results in lower bandwidth utilization compared to the other cases. A more accurate grant size prediction is preferable to over-prediction.

D. NAR Neural Network Predictions

Following the trends set by the baseline results, the data queueing delay is higher for the case with video prediction compared to without video prediction, which is expected due to serving more video packets instead of data. Therefore, video queueing delay is used to benchmark the effect of the NAR predictions on the performance of the EPON. If the video queueing delay is decreased by a margin large enough to also account for the data delay, that is, if the video delay difference subtracted from the data delay difference is less than or equal to zero net delay added, the NAR prediction is found to be worthwhile. For the four videos used, the remaining data is



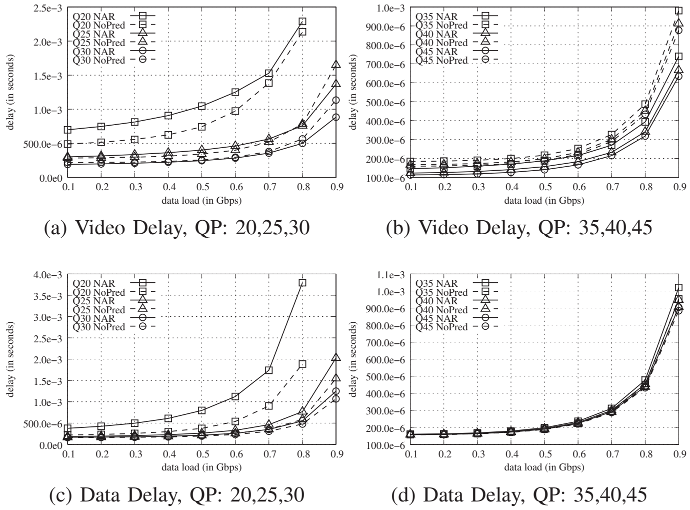
(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

Fig. 5. Over-prediction delay for video (Lake House) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.



(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

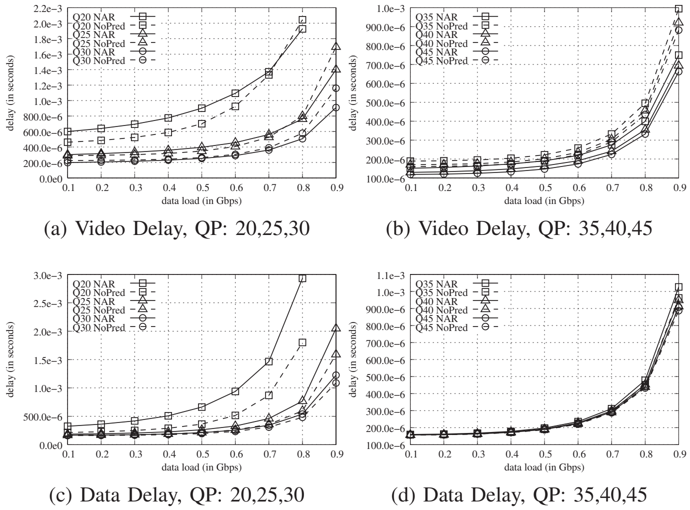
(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

Fig. 6. Delay for video (Harry Potter) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.

plotted in a manner similar to the baseline data.

The video queueing delay for Harry Potter is depicted in Figures 6 (a) and (b), showing NAR prediction results in data queueing delay which is consistently higher than the case without prediction for quantization levels 20 and 25. The data queueing delay values for quantization levels 30 through 45 are all smaller than the case without prediction. Figures 6 (c) and (d) are the plots of the data queueing delay, showing an increase in delay for all quantization levels. As the utilization of the network increases from 10% to 90%, the benefit of prediction is shown to become more pronounced, and the video delay for quantizaton level 25 intersects with the curve of without prediction and decrease the overall delay (Figure 6), with similar results for the other videos. While this marginal increase is not enough to justify the delay increase when a network is well below capacity, this result shows an increased benefit using video prediction when the network is highly utilized or nearing capacity. It is also noted that the data queueing delay values are much lower than the video queueing delay values. As shown in Table I, the data and video queueing delays can be combined to determine reduction in the network overall queueing delay. By adding the minor data queueing delay increase and the larger video queueing delay decrease an overall delay decrease is obtained. Because the delays are consistently large for parameters 20 and 25, it would provide no benefit to predict such videos. The table shows the data for all quantization parameters from 30 through 45. Similar trends are present in the data for the other videos, depicted in Table I and Figures 7, 8, and 9. In all cases discussed, the video prediction improve the performance of the Ethernet networks when the network utilization is near capacity, providing a larger benefit in terms of decreased queueing delay when the network demand is higher.



(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

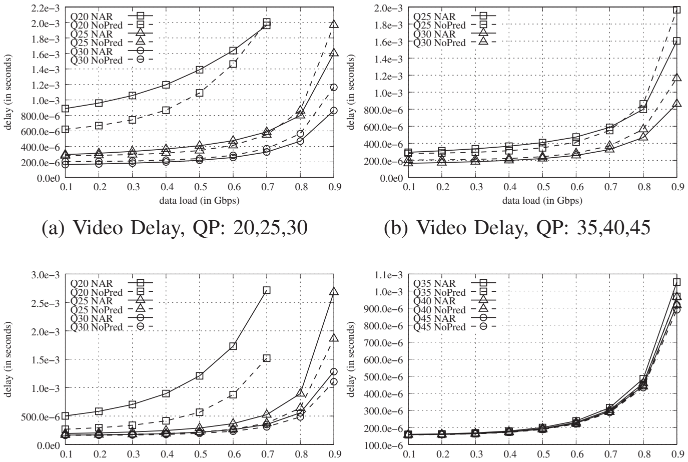
(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

Fig. 7. Delay for video (Finding Neverland) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.

V. Conclusion

Using a NAR neural network to forecast bandwidth requirements for H.265 encoded 4K video traffic appears to be a promising method. Test results indicate NAR network predictions increase the data and video traffic delay at low quantization levels of 20 and 25, which increases the overall delay unnecessarily, to the detriment of system efficiency. High quantization levels of 35, 40, and 45 provide increased performance in terms of improved video delay, but with slightly increased data delay values as compared to the baseline, unpredicted values. The overall delay is decreased for the higher quantization levels due to the relatively large decrease in video delay compared to the slight increase in data delay. Quantization level 30 showed only marginal improvement



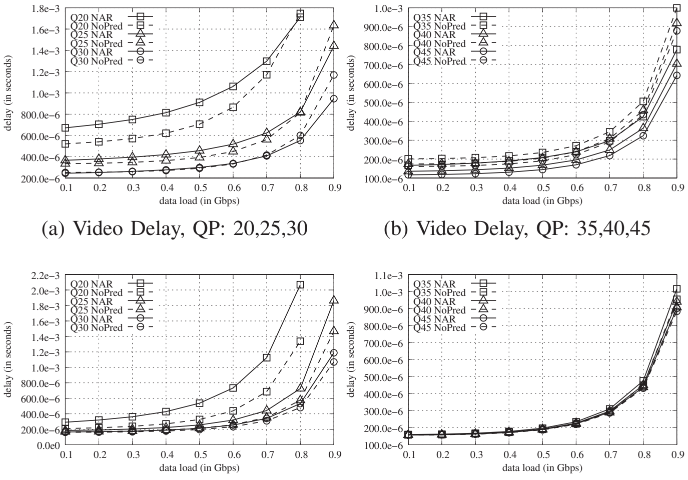
(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

Fig. 8. Delay for video (Speed) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.



(a) Video Delay, QP: 20,25,30

(b) Video Delay, QP: 35,40,45

(c) Data Delay, QP: 20,25,30

(d) Data Delay, QP: 35,40,45

Fig. 9. Delay for video (Lake House) and data packets, load values are for the background data traffic. Total load is video load plus the background data load.

in most cases but never increased the overall delay of the system, unlike quantization levels 20 and 25. In the cases of quantization levels 30, 35, 40, and 45, the system delay is decreased most significantly when the network nears saturation, typically around 0.8 Gbps throughput on the simulated EPON system. Below 0.8 Gbps utilization, the benefit of NAR video prediction still exists, but to a lesser extent.

Overall, there is a good case for predicting frame sizes using the NAR neural network model for videos compressed with a quantization levels 30 through 45. Higher quantization values were not tested in this experiment. Quantization level 30 provides the least benefit under light to normal use, but provides a benefit at heavy EPON saturation without detrimental effect when the EPON is not saturated. In contrast, prediction values for quantization levels 20 and 25 provide no benefit and actually appear to increase overall delay in all cases. Therefore no NAR neural network predictive model should be employed for quantization levels below 30.

References

[1] Cisco, "The zettabyte eratrends and analysis," *Cisco Visual Networking Index*, 2016.

[2] H. Koumaras, M. Kourtis, and D. Martakos, "Benchmarking the encoding efficiency of h.265/hevc and h.264/avc," *Future Network & Mobile Summit (FutureNetw)*, pp. 1 – 7, July 2012.

[3] D. Zegarra Rodriguez and G. Bressan, "Performance assessment of high efficiency video coding - hevc," *IEEE Global High Tech Congress on Electronics (GHTCE)*, pp. 110 – 111, Nov. 2013.

[4] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of h.265/mpeg-hevc, vp9, and h.264/mpeg-avc encoders," *Picture Coding Symposium (PCS)*, pp. 394 – 397, Dec. 2013.

[5] P. Seeling and M. Reisslein, "Video traffic characteristics of modern encoding standards: H.264/avc with svc and mvc extensions and h.265/hevc," *The Scientific World Journal*, p. 16, 2014.

[6] M. P. Mcgarry, M. Reisslein, and M. Maier, "Ethernet passive optical network architectures and dynamic bandwidth allocation algorithms," *IEEE Communications Surveys Tutorials*, vol. 10, no. 3, pp. 46–60, Third 2008.

[7] R.J. Haddad, M.P. McGarry, and P. Seeling, "Video bandwidth forecasting," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 4, pp. 1803 – 1818, Fourth 2013.

[8] A. M. Adas, "Using adaptive linear prediction to support real-time vbr video under rcbr network service model," *IEEE/ACM Transactions on Networking*, vol. 6, no. 5, pp. 635–644, Oct 1998.

[9] X. Wang, Souhwan Jung, and J. S. Meditch, "Dynamic bandwidth allocation for vbr video traffic using adaptive wavelet prediction," in *Communications, 1998. ICC 98. Conference Record. 1998 IEEE International Conference on*, Jun 1998, vol. 1, pp. 549–553.

[10] Sang-Jo Yoo, "Efficient traffic prediction scheme for real-time vbr mpeg video transmission over high-speed networks," *IEEE Transactions on Broadcasting*, vol. 48, no. 1, pp. 10–18, Mar 2002.

[11] K. Y. Lee, K. S. Cho, and B. S. Lee, "Efficient traffic prediction algorithm of mutimedia traffic for scheduling the wireless network resources," in *2007 IEEE International Symposium on Consumer Electronics*, June 2007, pp. 1–5.

[12] N. Sadek and A. Khotanzad, "Multi-scale high-speed network traffic prediction using k-factor gegenbauer arma model," in *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, June 2004, vol. 4, pp. 2148–2152.

[13] A. D. Doulamis, N. D. Doulamis, and S. D. Kollias, "An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of mpeg video sources," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 150–166, Jan 2003.

[14] A. Abdennour, "Evaluation of neural network architectures for mpeg-4 video traffic prediction," *IEEE Transactions on Broadcasting*, vol. 52, no. 2, pp. 184–192, June 2006.

[15] A. Doulamis and G. Tziritas, "Content-based video adaptation in low/variable bandwidth communication networks using adaptable neural network structures," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 4037–4044.

[16] A. D. Doulamis, N. D. Doulamis, and S. D. Kollias, "Recursive non linear models for on line traffic prediction of vbr mpeg coded video sources," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, 2000, vol. 6, pp. 114–119.

[17] A. Bhattacharya, A. G. Parlos, and A. F. Atiya, "Prediction of mpeg-coded video source traffic using recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2177–2190, Aug 2003.

[18] Po-Rong Chang and Jen-Tsung Hu, "Optimal nonlinear adaptive prediction and modeling of mpeg video in atm networks using pipelined recurrent neural networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1087–1100, Aug 1997.

[19] A. F. Atiya, M. A. Aly, and A. G. Parlos, "Sparse basis selection: new results and application to adaptive prediction of video source traffic," *IEEE Transactions on Neural Networks*, vol. 16, no. 5, pp. 1136–1146, Sept 2005.

[20] R. J. Haddad and M. P. McGarry, "Feed Forward Bandwidth Indication (FFBI): Cooperation for an accurate bandwidth forecast," *Elsevier Computer Communications*, vol. 35, no. 6, pp. 748–758, Mar 2012.

[21] "Arizona State University video trace library," Website, http://trace.eas.asu.edu.