4K Video Traffic Prediction using Seasonal Autoregressive Modeling

Dejan R. Marković, Ana M. Gavrovska, Member, IEEE, and Irini S. Reljin, Senior Member, IEEE

Abstract — From the perspective of average viewer, high definition video streams such as HD (High Definition) and UHD (Ultra HD) are increasing their internet presence year over year. This is not surprising, having in mind expansion of HD streaming services, such as YouTube, Netflix etc. Therefore, high definition video streams are starting to challenge network resource allocation with their bandwidth requirements and statistical characteristics. Need for analysis and modeling of this demanding video traffic has essential importance for better quality of service and experience support. In this paper we use an easy-to-apply statistical model for prediction of 4K video traffic. Namely, seasonal autoregressive modeling is applied in prediction of 4K video traffic, encoded with HEVC (High Efficiency Video Coding). Analysis and modeling were performed within R programming environment using over 17.000 high definition video frames. It is shown that the proposed methodology provides good accuracy in high definition video traffic modeling.

Keywords — 4K Video traffic, HEVC, autoregressive model, seasonal technique, video tracing, prediction.

I. INTRODUCTION

In the Cisco Visual Networking Index: Forecast and Methodology for period 2015-2020, it is clearly stated that global consumer internet video traffic is expected to reach 82% of all (both business and consumer) internet traffic by 2020 (compared to 70% in 2015) [1]. The expansion of video streaming services is not surprising. YouTube video-sharing website is among the two most popular [2]. It would take an individual more than 5 millions of years to watch the amount of video that will

Paper received May 15, 2015; accepted June 20, 2017. Date of publication July 31, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Grozdan Petrović.

This paper is a revised and expanded version of the paper presented at the 24th Telecommunications Forum TELFOR 2016 [21].

This research is partially supported by the Serbian Ministry of Education, Science and Technological Development through the project TR32048.

Dejan R. Markovic is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia and Telenor, Omladinskih brigada 90, 11070 Belgrade, Serbia (phone: 381-11- 3218383; e-mail: 1.dejan.markovic@gmail.com).

Ana M. Gavrovska is with the Telecommunications Department, School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia (e-mail: anaga777@gmail.com, anaga777@etf.rs).

Irini S. Reljin is with the Telecommunications Department, School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia (e-mail: irinitms@gmail.com, irini@etf.rs).

cross global IP networks each month in 2020, since every second is going to be filled with over millions of minutes of video traffic going through the network. Cisco also estimates that by 2020, more than 40% of connected flat-panel TV sets will be 4K, Fig. 1 [1].

These trends brought higher interest for researching high quality video and the network behavior. The main objective is to ensure the best possible resource allocation and better bandwidth control that are dependent on video traffic prediction and management of complex video streaming. Therefore, the need for analysis and modeling of HD video traffic is of essential importance [3].

However, modeling of video traffic is not an easy task, especially when knowing about high variability of HD and UHD video frame sizes. Such variability has increased with the introduction of H.264/AVC (*Advanced Video Coding*) high definition video codec standard. AVC provides a better compression rate (i.e. lower mean values) than its predecessors. Yet at the same time, it results in higher frame size variability [4]. Therefore, there is a bigger challenge by using HEVC (*High Efficiency Video Coding*) encoded video for purpose of this analysis, since HEVC is bringing higher frame size variability compared to H.264 [5, 6].



Fig. 1. Connected 4K TV sets in millions.

In this paper, we used video traces in order to analyze 4K video traffic. They are convenient to use since network researchers do not need to have detailed knowledge of video coding nor video coding equipment to carry out video networking research [7].

Different strategies of network resources allocation can be discussed through various simulations. Environments such as NS/2 software are allowing creation of the necessary network topology [3, 8]. On the other hand, it is necessary to work with realistic video traffic and adapt to different content and different simulation conditions. In that context, as it is shown in Fig. 2 when considering scenarios for video traffic analysis and testing, two streams can be recognized: actual video traces used by trace-driven simulations and statistical models that can be used to generate the required video sequences for simulations.



Fig. 2. Scenarios for video traffic analysis.

Statistical models are considered to be a good choice since once representative modeling is adopted, it is easy to apply the model according to different video traffic parameters. It also allows a better insight and understanding of different video traffic characteristics. On the other hand, most of the statistical models are complex and often developed for short video sequences [3]. Therefore, an easy-to-apply and easy-to-understand statistical model is desired, as long as it is giving a good presentation of real network video traffic.

In this paper, Seasonal Autoregressive integrated moving average (Seasonal ARIMA) model is used for modeling and prediction of 4K video traces, encoded with H.265 encoding standard. The analysis is based on publicly available HD video traces [9] in R programming environment.

There are different approaches available for analysis, statistical modeling and prediction of high definition video traces [3-4, 6, 10-11]. Mostly used models are Autoregressive (AR) models and their combinations, such as Autoregressive integrated moving average (ARIMA), Seasonal ARIMA (SARIMA) and Stochastic Autoregressive Mean (SAM).

This paper is organized in five sections. In Section II we discuss the basics of HD/UHD video concept, Autoregressive models and applied SARIMA model. Section III describes the proposed methodology, while Section IV sums the obtained experimental results. Finally, Section V concludes this paper.

II. 4K VIDEO TRAFFIC ANALYSIS

A. 4K Video Basics

High definition digital video became omnipresent in everyday life. Advances in technology made it possible to record and reproduce video material in ultra-high definition. However, the current internet and broadcasting networks should deal with distribution of large amounts of such content. Limits of current technology have led MPEG (International Standards Organization/International Electrotechnical Commission Moving Picture Experts Group) and VCEG (International Telecommunication Union - Telecommunication Standardization Sector Video Coding Experts Group) to establish JCT-VC (Joint Collaborative Team on Video Coding) with the aim of developing a new video coding standard for a high performance. The result was the development of a new compression standard known as HEVC or H.265 in order to significantly improve the compression efficiency in comparison with the existing H.264.

Concept of HD video is the best known as a video with better quality compared to the standard definition (SD), where the quality is firstly seen through higher resolutions. However, high definition is much more than that. Distribution of high definition video to end-user is a challenging task. Better image quality comes with higher bandwidth needs and requires advanced video compression and storage. 4K video can take ten times more memory usage compared to SD and three times more memory usage compared to full HD equivalent video sequence.

TABLE 1: VIDEO DEFINITION STANDARDS

TABLE 1. VIDEO DEFINITION STANDARDS				
Format	Pixels	Aspect Ratio		
Standard Definition (SD)				
480p / 480i	720×480 (704×480)	4:3 (approx)		
576p / 576i	720×576 (704×576)	4:3 (approx)		
High Definition (HD)				
720p	1280×720	16:9		
1080p / 1080i	1920×1080	16:9		
Ultra High Definition (UHD)				
4K (2160p)	3840×2160	16:9		
8K (4320p)	7680×4320	16:9		
8640p	15360×8640	16:9		
Digital Cinema (DCI)				
2K	2048×1080	1.90:1		
4K	4096×2160	1.90:1		

Details regarding UHD format can be found in ITU-R Recommendation 2020, known as Rec. 2020 or BT.2020. This recommendation defines different aspects of UHD TV, such as screen resolution, frame rate, chroma subsampling, bit depth, color space. As shown in Table 1, there is a difference between UHD and 4K terms. The easiest way to make a difference is to consider the 4K as a standard for professional production and cinema, and the UHD for enduser's displays and broadcast [12]. The 8K standard is slowly starting to be commercially more available, as it is already available on YouTube [13]. Moreover, mobile data video traffic is increasing despite limitations regarding device's screen dimension, screen resolution, limited network bandwidth and lower computing power. It is a challenge to ensure video quality to mobile users that is matched to quality that TV users can have.

B. HEVC encoded video

Continuous technology development on one hand and more demanding user expectations on the other, have a major impact on the development direction of standards for video compression. Increased technical capabilities of today's TV sets such as higher contrast, a larger number of pixels, lower consumption, realistic color display, etc. give a stronger basis and determine the way of different video formats development. Proper optimization software is necessary with the complete hardware support. Therefore, the core of implemented software used for reproducing video material is video encoding.

Designed for a wide range of applications, H.264 is currently the most used compression standard. It is used from video in mobile phones through web applications to TV broadcasting [14].

HEVC has the same basic structure as H.264/AVC. However, HEVC contains many incremental improvements such as better compression, at the cost of potentially increased processing power. It is designed to achieve a higher coding efficiency, ease of transport, data loss resilience etc. [14]. It enables more efficient video storage and transmission, compared to H.264. In practice, this means that for the same picture size and quality, an HEVC encoded video should occupy less storage or transmission capacity than the equivalent H.264 encoded video. On the other hand, the quality and/or resolution of an H.265/HEVC encoded video should be higher at the same storage or transmission bandwidth, compared to the same video sequence encoded with H.264 standard [14]. Some of the results in comparing H.265 and H.264 can be found in [6].

C. Criteria for Modeling of Video Traffic

Many models, which are suggested for video traffic modeling, can be found in literature. Most of them are based on the Markov chain, wavelet and fractal techniques, ARIMA, Fractional ARIMA (FARIMA), etc. [15-18]. In this paper we apply Seasonal ARIMA (SARIMA) model, according to [4] as a model that is simple to implement, without requiring complex steps. We considered Akaike's Information Criterion (AIC) as an accuracy criterion for optimization, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Additional performance evaluation was done by using Empirical Cumulative Distribution Function (ECDF) graphs.

AIC from [19] is calculated as:

$$AIC = N \log(\frac{SEE}{N}) + 2(k+2), \tag{1}$$

where N is the number of the video frames (number of observations), SEE is the sum of squared errors and k is the number of parameters (predictors). The k+2 part means that there are k+2 parameters in model: the k coefficients of predictors, the intercept and the variance of residuals. The smaller the value of the AIC, the better is the model for later forecasting. The lower AIC values lead to better model validity and simplicity. Although they should all lead to the similar model being selected, different computer packages use slightly different definitions for AIC [19].

Performance is evaluated using RMSE and MAE, which are applied to evaluate the difference between the models and the actual video traffic. These are calculated as shown in equation (2), where N is the number of frames and e_i is the modeling error at index *i*.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |e_i|,$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} e_i^2},$$
(2)

In order to examine the potential existence of a certain pattern in the residuals, correlation between the modeling residuals is calculated. This is performed using ACF – Autocorrelation Function (and PACF – Partial ACF), that measures the linear relationship between lagged values of a time series and whose coefficients can be calculated according to [19] as:

$$r_{k} = \frac{\sum_{t=k+1}^{T} (y_{t} - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{T} (y_{t} - \overline{y})^{2}},$$
(3)

where T is the length of the time y_t series with the corresponding mean value \overline{y} .

In this paper, modeling was performed in R programming environment [20]. R is a commonly used tool for statistical analysis, appropriate for all platforms (Windows, UNIX, and Mac OS). Other programming languages, such as SAS and Python, could also be used for ARIMA modeling and time series forecasting.

D. Seasonal ARIMA in R

In this paper Seasonal ARIMA model is used for modeling and prediction of 4K video traffic. One of the reasons for choosing SARIMA models is because SARIMA models consider both non-seasonal and seasonal parts of traces. SARIMA defined data is as SARIMA $(p, d, q)(P, D, Q)_m$, where parameters p, d and qrepresent the orders of non-seasonal autoregressive part, differencing and moving average part, respectively, and the capital letters, P, D and Q represent the orders of seasonal autoregressive part, differencing and moving average, respectively. Parameter *m* stands for the number of seasons (periods) which is equal to the frame rate here.

R was recognized as a very good tool because it has developed functions for ARIMA modeling, such as *auto.arima(), arima(), arima.sim(), garsim()* and *Arima()*. We decided to use *Arima()* function, from the *forecast* package, as it is suggested to be the best for time-series analysis [19]. On the other hand the function *auto.arima()* is found as useful, but during the extensive experiments it is showed that it only takes care of stationarity and AIC comparison. Function *arima()* is not quite suitable for later usage of *forecast()* function [19], while *arima.sim()* and *garsim()* seem not to be adjusted for SARIMA (seasonal ARIMA) modeling in comparison to ARIMA. To adjust *arima.sim()* and *garsim()* for adequate usage, additional statistical knowledge is needed [4].

III. SIMULATION

A. Materials and Experimental Settings

The analysis of 4K video traffic is performed using the extracted video traces from [9]. A scene/frame from the 4K video "Tears of Steel" (2012) is shown in Fig. 3. The characteristics of the sequence are given in Table 2. After the short animation films "Elephants Dream" (2006), "Big Buck Bunny" (2008) and "Sintel" (2010), this is a short created film used as a reference for the international standard of visual effects. It provides filmmakers and developers around the world with optimal reference footage.



Fig. 3. Scene from the Tears of Steel movie.

Name	Value
Resolution	4096×1744
Frames per Second	24
Encoder	H.265/HEVC
Number of Frames	17592
Minimum Frame Size (Bytes)	61
Maximum Frame Size (Bytes)	2037283
Average Frame Size (Bytes)	281065.30
Number of Groups of Pictures	733

As shown in Fig. 4, in order to obtain experimental results, four main steps are presented: having 4K video sequence as an input encoded with HEVC, extraction of corresponding 4K video trace, separation of frames for the learning and prediction phases and, at the end, comparison of original and modeled video traces.



Fig. 4. Experimental steps.

B. Methodology

As mentioned in Section II, modeling in this paper was done in R programming environment. A schematic view of the steps performed while obtaining the best Seasonal ARIMA model in R, is illustrated in Fig. 5. R programming environment is chosen as convenient for statistical modeling. The ARIMA modeling functions are available in R, and free of charge. R based functions can be used on all platforms and can interact with other programming languages, such as C.



Fig. 5. Schematic view of implemented procedure.

For the performance evaluation, we considered Akaike's Information Criterion (AIC) [19] as the accuracy criterion. Then we calculated Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), as described in Section II. In order to illustrate the obtained results, ECDF plots for the obtained models are presented.

IV. EXPERIMENTAL RESULTS

A. Modeling 4K Video Traces

For the purposes of modeling, we extracted 11728 frames (67%) from an original video trace, in order to use them in the training (learning) phase, Fig. 6. The model is generated for this dataset, as described in Section III.



Fig. 6. Training (learning) phase dataset plot in R.

As shown in Fig. 5, after plotting the observed dataset, the first indication of which SARIMA model could be a good one is to plot ACF (PACF) function that measures the relationship between lagged values, as mentioned in Section II. ACF and PACF functions of dataset used in the learning phase are shown in Fig. 7.



Fig. 7. Training phase dataset ACF and PACF plots.

Several SARIMA models are obtained to see which one fits the best our dataset, based on criteria described in Section III. Differences between models are found by changing the SARIMA parameters while following AIC, RMSE and MAE values, Fig. 5. While applying different models, processing and memory limitations are often faced since the analysis is performed on a very high number of data. Therefore, the best model search ended with Model 5. Comparison of applied models is shown in Table 3.

 TABLE 3: COMPARISON OF CONSIDERED SARIMA MODELS

Model	AIC	RMSE	MAE
SARIMA model 1	25020.94	156591.8	69477.86
SARIMA model 2	15737.52	120586.9	47547.33
SARIMA model 3	-241.9	139264.8	19965.11
SARIMA model 4	-246.2	139145.8	19986.65
SARIMA model 5	5926.6	61041 72	17529 49
(The proposed)	-3830.0	01041.72	1/556.46

The presented results indicate that the best considered model is SARIMA model 5. This model has the lowest AIC value, as well as the lowest RMSE and MAE values and has quite good following of the ECDF curve of the original trace. Fig. 8 illustrates the comparison of SARIMA model 1 and model 5 ECDFs with original ECDF. It is shown that SARIMA model 5 has better following of the original ECDF curve.

B. Forecasting 4K video traces

In order to make a prediction for 33% (5.864 frames) of frames based on 67% used for modeling, we used *forecast()* function. Our results show that it is not good to have this long prediction interval, because the video frame sizes change frequently and do not follow a certain pattern for a long period of time. That may result in severe prediction errors. Prediction error results either in inefficient use of network resources, or in an increased rate of dropped packets.



Fig. 8. Graphical ECDF comparisons between the SARIMA model 1 and the chosen SARIMA model 5.

One of the signs that the model will bring prediction errors, because of the high frame size variability, can be seen in Fig. 9 where we compare plots of original dataset with chosen SARIMA model 5. As it is shown, the model brings significant distortion in the second half of the modeled dataset and that has an impact on frame prediction.



Fig. 9. Mismatch of original dataset (gray) and SARIMA model 5 (blue) due to frame size variability.

Prediction accuracy was conducted by using RMSE and MAE and accuracy results are shown in Table 4. It can be noticed that RMSE and MAE are not the lowest for the proposed SARIMA model 5. Therefore, the results shown in Table 4 illustrate the consequence of choosing a long range prediction interval, based on the model obtained for a long range of observed 4K video frames.

This was also confirmed while comparing ECDF graphs of original and model frames, Fig. 10. The modeled traffic shows a significant deviation compared to the original 4K video traffic. However, our additional research has shown that the chosen model is good enough to predict up to 1.000 video frames (6% instead of 33%), Fig. 11.

Model	RMSE	MAE
SARIMA model 1	283830.8	186754.3
SARIMA model 2	284761.7	187217.8
SARIMA model 3	300648.9	198181.0
SARIMA model 4	300233.7	197828.6
SARIMA model 5	2868262	196942-1
(The proposed)	280830.3	180845.1

Original vs. Generated 4K video trace

TABLE 4: COMPARISON OF PREDICTION RESULTS



Fig. 10. ECDFs of original (red) and generated (blue) 4K video traffic (5.864 frames prediction).



Fig. 11. ECDFs of original (red) and generated (blue) 4K video traffic (1.000 frames prediction).

V. CONCLUSION

In this paper, we performed an analysis of the 4K video traffic encoded with H.265/HEVC coding standard. For the purpose of experimental analysis, we used a publicly available video trace library. We decided to test an easy-to-apply and easy-to-understand statistical model for video traffic modeling and prediction. In this paper it is shown that Seasonal ARIMA can be used for modeling 4K video traffic in R. However, a high frame size variance was not taken into account on purpose, and this had a negative effect on both, 4K video traffic modeling and prediction. Therefore, this paper confirms that a frame size variance

could have a significant negative impact on 4K video traffic modeling and prediction, especially if wide intervals are used in learning phase. Hence, this should be taken into consideration when defining the prediction interval in order to avoid severe prediction errors.

Future work should be oriented towards making a balance between the observed and the predicted time intervals according to the selected evaluation measures.

REFERENCES

- [1] Index, Cisco Visual Networking. "Forecast and methodology, 2015-2020 white paper." *Retrieved 1st June* (2016).
- [2] Alexa, http://www.alexa.com/topsites (last accessed 01.02.2017.)
- [3] A. Tamimi, A. Karim, R. Jain, and C. So-In, "Modeling and generation of AVC and SVC-TS mobile video traces for broadband access networks." In *Proc. of the first annual ACM SIGMM conference on Multimedia systems*, pp. 89-98, 2010.
- [4] A. Tamimi, A. Karim, R. Jain, and C. So-In, "Statistical analysis and modeling of high definition video traces." In *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 596-601, IEEE, 2010.
- [5] P. Seeling, 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 2014.
- [6] A. Gavrovska, I. Reljin, "Comparative Analysis of Different Video Compressions at Different Resolution Formats," *IEEE Broadcast Symposium*, Hartford CT, USA, 12-14 October 2016.
- [7] M. Reisslein, L. Karam, and P. Seeling. "H. 264/AVC and SVC video trace library: a quick reference guide." May 2009.
- [8] J. Wu, Jiyan, C. Yuen, M. Wang, and J. Chen, "Content-aware concurrent multipath transfer for high-definition video streaming over heterogeneous wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, 27(3), pp. 710-723, 2016.
- M. Reisslein, "Video trace library." Arizona State University, 2012, http://trace.eas.asu.edu (last accessed 01.09.2016.)
- [10] M. Macedo, "Three Notes on the SAM Video Statistical Model.", (last accessed 01.09.2016)
- [11] A. Tamimi, A. Karim, R. Jain, and C. So-In, "Dynamic resource allocation based on online traffic prediction for video streams." In *Internet Multimedia Services Architecture and Application (IMSAA)*, 4th International Conference on, pp. 1-6. IEEE, 2010.
- [12] D. Caleb, "Your 1080p TV is old already: everything you need to know about ultra HD 4K", 2014, https://www.digitaltrends.com (*last accessed* 01.09.2016.)
- [13] Pantagonia 8K,https://www.youtube.com/watch?v=ChOhcHD8fBA (last accessed 01.02.2017.)
- [14] M. Uhrina, J. Frnda, L. Sevcik, and M. Vaculik, "Impact of H. 264/AVC and H. 265/HEVC compression standards on the video quality for 4K resolution," *Advances in Electrical and Electronic Engineering* 12(4), pp. 368, 2014.
- [15] D. C. Montgomery, L. A. Johnson, J. S. Gardiner, "Forecasting and Time Series Analysis", Second Edition, McGraw-Hill Companies, 381 pages, July 1990.
- [16] X. Huang, Y. Zhou and R. Zhang, "A Multiscale Model for MPEG-4 Varied Bit Rate Video Traffic," *IEEE transactions on broadcasting*, 50(3), pp. 323- 334, 2004.
- [17] L. de la Cruz et al., "Self-Similar Traffic Generation Using a Fractional ARIMA Model Application to the VBR MPEG Video Traffic," *ITS 98 Proceedings*, pp. 102-107, IEEE, 1998.
- [18] I. Reljin, A. Samčović, B. Reljin, "H. 264/AVC video compressed traces: multifractal and fractal analysis," *EURASIP Journal on Advances in Signal Processing*, pp.1-13, 2006.
- [19] R. J. Hyndman, and G. Athanasopoulos. Forecasting: principles and practice, OTexts, 2014.
- [20] The project R: http://www.r-project.org (last accessed 01.09.2016).
- [21] D. R. Marković, A. M. Gavrovska, I. S. Reljin, "4K video traffic analysis using seasonal autoregressive model for traffic prediction," 2016 24th Telecommunications Forum (TELFOR), Belgrade, 2016, pp. 1-4.