A Real-Time Bandwidth Prediction Model for H.265 Video (Postprint)

Authors: Liu Jiangtao, Li Zeping, Lin Chuan

Date: 2018-12-13T00:00:00+00:00

Abstract

To address the issue of degraded user experience caused by video frame loss due to bandwidth jitter during live video streaming, a H.265 video bandwidth real-time prediction (VBRP) model is proposed. Based on Markov chain theory and focusing on H.265-encoded live video streams, this model investigates the statistical characteristics of B-frame occurrence under GOP (Group of Pictures) encoding mode. It is discovered that B-frame size significantly impacts video stream transmission rate. By leveraging this characteristic, the size of B-frames in live streams can be predicted, and B-frames can be selectively discarded according to network bandwidth availability. Furthermore, a step adjustment factor AF and an error threshold FT are introduced to balance the model' s training frequency and the number of predicted frames during B-frame prediction. Finally, a VBRP prediction algorithm is implemented based on the model, and its effectiveness is validated within a live streaming system.

Full Text

Preamble

Vol. 37 No. 2
Application Research of Computers
ChinaXiv Cooperative Journal

A Real-Time Prediction Model for H.265 Video Bandwidth

Liu Jiangtao, Li Zeping, Yang Bingzhao
(College of Computer Science & Technology, Guizhou University, Guiyang 550025, China)

Abstract: To address the degradation of user experience caused by video frame loss due to bandwidth jitter during live video streaming, this paper proposes a real-time prediction model (VBRP) for H.265 video bandwidth. Based on

Markov chains and focusing on H.265-encoded live video streams, the model investigates the statistical characteristics of B-frames in GOP (Group of Pictures) encoding mode. The study reveals that B-frame size significantly influences video stream transmission rates. Leveraging this characteristic, the model can predict the size of B-frames in live streams and selectively discard them based on network bandwidth conditions. To balance training frequency and the number of predicted frames, step adjustment factor AF and error threshold FT are introduced. Finally, the VBRP prediction algorithm is implemented and validated within a live streaming system.

Keywords: bandwidth prediction; H.265; Markov; media streaming

1 Related Work

According to Netflix' s 2017 annual report [?], there are over 117 million streaming members worldwide, watching more than 140 million hours of television programs and movies daily. Future projections indicate that streaming traffic will grow by 67% in the mobile sector and 29% in fixed networks [?]. This explosive growth of multimedia content has driven industry development and research, making it a priority to seek more efficient streaming distribution architectures that provide better video services for all users. In current network services, video streaming already occupies the vast majority of wired and wireless bandwidth. H.265/HEVC (High Efficiency Video Coding) [?] offers double the compression rate of H.264 at equivalent image quality, making its application in streaming transmission highly beneficial for improving network bandwidth utilization. However, network bandwidth is not always stable; it is typically a time-varying system that changes dynamically over time, particularly evident in high-jitter, high-delay mobile network environments. When network jitter occurs during video live streaming, it causes loss of critical video data and degrades video quality. Traditional solutions reduce video quality according to available bandwidth [?], resulting in poor user experience and decreased service quality. If statistical methods can predict video bandwidth and selectively discard B-frames based on network conditions, this can balance Quality of Service (QoS) with network bandwidth capacity, thereby improving Quality of Experience (QoE).

Sarkar et al. [?] proposed a Markov-modulated gamma-based framework, while Lanfranchi et al. [?] introduced an MPEG-4 prediction model to enhance user QoE and provider QoS. These models must adapt to dynamically changing network conditions such as throughput, packet loss rate, and delay jitter [?]. Kalampogia et al. [?] proposed a simulated annealing algorithm for predicting video B-frames. However, simulated annealing and MPEG-4 prediction models are insufficiently accurate for H.265-encoded videos because H.265 contains many B-frames much smaller than the mean, randomly distributed throughout the video and not easily predicted by formulas. According to analysis in [?], video models can be categorized into two types: data rate models and frame size models. Data rate models cannot identify the impact degree of individual frames on the entire video; even small data losses involving I-frames can affect received video quality. In [?], the authors calculated the probability of timely video stream packet delivery in LTE networks and allocated video packets to users based on probability without distinguishing the importance of different frame types. However, [?] follows the DASH standard [?], which is unsuitable for very low-latency video streaming. Reference [?] modeled B-frames using known distributions to find their statistical characteristics, but for live video streams, upcoming frames are unknown and cannot be statistically analyzed, significantly affecting B-frame prediction accuracy.

To address these issues, this paper proposes a Video Bandwidth Real-Time Prediction (VBRP) model for H.265. The VBRP model is based on Markov chains and aims to improve live video quality in dynamic network environments. The video bandwidth real-time prediction VBRP algorithm is implemented on an RTMP (Real Time Messaging Protocol) [?] streaming media publishing prediction system. This paper focuses on H.265 [?], the next-generation video coding standard succeeding H.264. Like H.264, video coding sequences in H.265 consist primarily of three frame types: I-frames, P-frames, and B-frames forming a GOP. I-frames are intra-coded picture frames that encode using only information from the current frame without referencing others. P-frames are predictive-coded picture frames that use previous I-frames or P-frames for inter-frame predictive coding through motion prediction. B-frames are bidirectionally predicted picture frames that provide the highest compression ratio, requiring both previous (I or P) and subsequent (P) frames for bidirectional inter-frame predictive coding through motion prediction. This paper investigates the statistical characteristics of live video streams, specifically the impact of B-frames on transmission rates under GOP encoding mode, demonstrating the significance of selectively discarding predicted B-frames based on current bandwidth capacity to improve user experience.

---

2 Characteristics of H.265 Encoded Live Video Streams

In H.265-encoded video streams, B-frames not only constitute the vast majority of each GOP but are also more easily compressed and can have information discarded more readily than other frames. This means B-frames have longer duration of bandwidth demand and are more unstable than other frames, leaving room for further optimization of bandwidth allocation to improve utilization. Additionally, B-frames are less important than other frames, so appropriately discarding them has limited impact on overall video quality. To verify the influence of B-frame size on live video stream variation, this paper adopts the coefficient of variation (CoV) method from [?] to describe video stream rate changes. Since live-transmitted video segments last only a few seconds or even less, this paper analyzes 60-minute H.265 video trace files. For a video sequence composed of N frames encoded at a given quantization parameter (QP) level, if the frame sizes are represented, the CoV for the video is defined as shown in Equation (1).

This paper obtained six different H.265 video description files from Arizona State University' s video trace library [?] for analysis, including "Finding Neverland," "Lake House," and three surveillance video description files of varying quality. The video files use G24B7 compression format, with each video file divided into three quality sub-files based on different QP values: QP=10, QP=25, and QP=40. To observe H.265 video stream characteristics, the standard deviation, mean, and CoV of B, P, and I frames were calculated for different QP values. The results in Table 1 show that the CoV of B-frames is significantly larger than that of P-frames and I-frames, indicating that B-frames are more numerous. According to Equation (1), low-quality videos (QP=40) have smaller overall frames, resulting in smaller means and standard deviations, thus larger CoV values. As video quality increases, the standard deviations and means of all three frame types increase. Although surveillance videos show larger B-frame CoV than the first two videos, their overall characteristics are similar. The mean and variance indicate that surveillance videos have higher overall quality than the first two videos, with obvious quality differences facilitating observation.

As shown in Table 2 , when all B-frames are removed from videos, the video CoV decreases significantly, indicating less size variation in I-frames and P-frames. Therefore, reducing frame size changes makes H.265-encoded video streams smoother during network congestion and instability. In Table 2, videos without B-frames show small differences between standard deviation and mean, while videos with B-frames show standard deviations approximately three times the mean. Overall, for videos with large scene changes (e.g., action movies, war movies), standard deviation is a crucial indicator of frame size variation. Such videos require not only high average bitrates but also exhibit large bitrate fluctuations. Conversely, low-quality videos typically require lower average bitrates. CoV is an excellent metric for describing bitrate variation, so selectively discarding B-frames can minimize bandwidth requirements and smooth encoded video streams.

---

3 Model Establishment and Algorithm Implementation

According to the H.265 encoding standard, B-frames in a GOP are derived by referencing I-frames and P-frames, suggesting strong correlations between these frame sizes. References [?, ?] used Equation (2) to calculate correlation coefficients between each B-frame size and each I-frame or P-frame in GOPs for MPEG-4, H.264, and H.265 videos, while Equation (3) calculated B-frame autocorrelation. Results showed strong autocorrelation among B-frames and strong correlation between B-frames and P-frames, enabling more accurate B-frame prediction.

Since H.265 video streams contain numerous randomly generated small B-frames

[?, ?], to improve prediction accuracy, this paper adopts the B-frame classification method proposed in [?, ?]. B-frames in each GOP of H.265 videos are divided into two subsets using the median of all B-frame sizes as the boundary: SBF (small B-frames) and BBF (big B-frames). Using EasyFit software to find optimal distribution fits for numerous H.265-encoded videos, analysis reveals that small frames (SBF) in H.265 videos best fit log-logistic or Pearson Type V distributions, while large frames (BBF) primarily fit Weibull distributions and less frequently uniform distributions. Like [?], maximum likelihood estimation determines these distribution parameters. During model training, relative percentage error (RPE) [?] evaluates prediction error between predicted and actual B-frame values to find optimal distributions for SBF and BBF. The experimental H.265 videos in this paper use G24B7 GOP pattern as shown in Table 3 , seeking optimal distributions for B1~B21 frames to predict B-frames in subsequent GOPs.

In addition to RPE, mean absolute percentage error (MAPE) as shown in Equation (6) and QQ-plot are used for more accurate distribution fitting. QQ-plot is a powerful graphical goodness-of-fit test [?] that visually verifies whether a dataset originates from the same distribution.

In summary, Algorithm 1 presents the training model algorithm for finding optimal distributions for B-frame prediction.

Algorithm 1: B-frame Optimal Distribution Training Algorithm
// m represents the number of B-frames in each GOP
// ArithmeticalMean is the arithmetic mean of all B-frame sizes in the GOP
// XP b and XP b represent parameters of distribution X for BBF
// XP s and XP s represent parameters of distribution X for SBF

```
Procedure FindingTheBestDistribution
  for CountBFrame = 0 to m by increment do
    Calculate ArithmeticalMean from Bframes;
    if (Bframe < ArithmeticalMean) then
      SBF ← Bframe;
    else
      BBF ← Bframe;
    end if

    // Generate random numbers from various distributions
    LogRB = LogisticRegression(LP b, LP b);
    LogRS = LogisticRegression(LP s, LP s);
    PearsonVB = PearsonV(PP b, PP b);
    PearsonVS = PearsonV(PP s, PP s);
    WeibB = Weibull(WP b, WP b);
    WeibS = Weibull(PP s, PP s);
    UniformB = Uniform(UP b, UP b);
    UniformS = Uniform(UP s, UP s);
    ··· // Other distributions
```

```
    BestDistBBF(CountBFrame) ← Best(CriterionListB);
    BestDistSBF(CountBFrame) ← Best(CriterionListS);
  end for
end Procedure
```

After training the model to find optimal distributions for BBF and SBF, predicting the size of upcoming B-frames requires first determining whether the incoming B-frame is BBF or SBF. The training results are used for real-time prediction of H.265 video B-frame types (BBF or SBF), followed by obtaining predicted B-frame sizes from the corresponding distributions. This paper proposes a Markov model-based prediction algorithm that establishes a 2-state Markov chain for each B-frame in a GOP. For example, using G24B7 GOP pattern, a GOP contains 21 B-frames, requiring 21 distinct 2-state Markov chains. Each state chain has four transition probabilities: BBF→BBF, BBF→SBF, SBF→BBF, and SBF→SBF, described by Equation (6). Four transition probabilities are defined as Pbb, Pbs, Psb, and Pss, obtained from Algorithm 1 during training. Figure 1 [Figure 1: see original paper] illustrates how the m-th B-frame in GOP t predicts the m-th B-frame type in GOP t+1 through transition probabilities. Predicting B1 frame involves two steps: first, calculating the predicted B1 frame type; second, generating B1 frame size using the optimal distribution, as shown in Figure 2 [Figure 2: see original paper].

Algorithm 2: Real-Time B1 Frame Prediction Algorithm Based on Markov Chains
// GOPn represents the number of GOPs
// (Pbb, Pbs, Psb, Pss) is the transition matrix obtained from training

```
Procedure B-FramesPrediction
  for t = InitialNum to GOPn by increment do
    // Calculate probabilities of B-frame being BBF or SBF
    (Pt+1,b, Pt+1,s) ← ((Ptb, Pts)(Pbb, Pbs, Psb, Pss))
    BestDistribution of BBF or SBF ← Max(Pt+k,b, Pt+k,s)
    // Predict B-frame size through optimal distribution
    Bframe ← BestDistribution()
  end for
end Procedure
```

4 Model Optimization

Algorithm 2 obtains the transition matrix for B-frame types (BBF and SBF) through initial model training and uses this consistent matrix to predict Bt,m, Bt+1,m, Bt+2,m, ···, Bt+n,m frames in GOP sequences (where t+n represents GOP sequence number and m represents B-frame index within a GOP). However, during live video streaming, the time-homogeneity of Markov chain transition probabilities may not hold over time. To mitigate this, a threshold determines whether to retrain. The FT value thus controls training frequency: larger FT values increase training frequency and resource consumption, while smaller FT values reduce training frequency.

Introducing step adjustment factor AF to limit the number of forward-predicted Bm frames effectively solves this problem. Let E be the accumulated error and A be a coefficient. When AF is small, the algorithm executes more frequently, increasing accumulated error, so retraining tolerance should increase (FT becomes larger). Conversely, when AF is large, FT becomes smaller. Therefore, a feasible approach is needed.

To selectively discard B-frames based on network bandwidth, this paper references the TCP congestion control algorithm Westwood [**?**] for bandwidth estimation. Bandwidth is estimated at the sender side during each RTT (Round-Trip Time) interval by counting successfully ACKed bytes. The bandwidth estimate then determines B-frame discard decisions, as shown in Equation (8), where dk represents the discard decision result.

---

5 Experiments and Simulation

5.1 Algorithm Simulation

To evaluate VBRP algorithm performance, a B-frame prediction program was developed to predict 1000 B-frames from "Finding Neverland," "Lake House," and surveillance videos. Video files with QP=10 quality served as input to the prediction program, outputting predicted B-frame sequences to simulate live video streams. This experiment verified the actual prediction error percentage. The physical platform used an Intel Core i5-3470 CPU with 16GB DDR3 and CentOS 6.8 operating system. As shown in Figure 3 [Figure 3: see original paper], the x-axis represents B-frame sequence numbers and the y-axis represents absolute percentage error between predicted and actual values. When accumulated error is large, the proposed VBRP algorithm demonstrates greater stability than simulated annealing prediction and One-Set-Fits-All algorithms. When predicting B-frames in "Finding Neverland," "Lake House," and surveillance videos, VBRP algorithm (black curve in Figure 3 [Figure 3: see original paper]) achieves error rates between 3% and 25%, significantly outperforming the other two algorithms.

5.2 System Implementation

Current live streaming solutions include Adobe' s RTMP protocol, Apple' s HLS (HTTP Live Streaming) protocol, and Google' s WebRTC. RTMP is mature with low latency and easy deployment but lacks universality. HLS suits various browsers, especially mobile terminals, with strong compatibility. WebRTC has low latency and moderate compatibility but complex implementation and low concurrency. For experimental convenience, this paper implemented an RTMP-based live video streaming system referencing [?, ?, ?]. The system consists of three components: CentOS as the base server operating system, Nginx + Nginx-rtmp-module as the streaming media publishing server, and FFmpeg as the streaming server.

Analysis of prediction algorithm topology reveals two integration methods. The first integrates the prediction algorithm into the Nginx streaming server, reorganizing RTMP chunks for prediction and performing: reorganize → discard B-frames → split on subsequent chunks. The second integrates the prediction algorithm into FFmpeg, creating a separate training thread and performing B-frame discard before chunk generation: discard B-frames → split. The first method increases latency, causing prediction lag and affecting the entire system. The second method performs prediction after H.265 encoding but before RTMP chunk generation, then discards B-frames based on network bandwidth. Clearly, the second method has lower latency and consumes fewer resources than the first.

Therefore, this paper implemented the second method, as shown in Figure 4 [Figure 4: see original paper].

---

6 Conclusion

This paper' s VBRP model studies H.265-encoded video stream characteristics, classifying B-frames into BBF and SBF types based on these features. During training, optimal fitting distributions are found for BBF or SBF frame types. A novel Markov chain predicts upcoming B-frame types, and the optimal fitting distribution generates the predicted frame size. The model introduces FT as a training threshold and AF as a step adjustment factor for B-frame prediction, enabling dynamic variation of training frequency based on FT and accumulated error E, further improving prediction accuracy. After simulation validation, the VBRP algorithm was applied to a streaming media publishing system composed of Nginx and FFmpeg. Experiments demonstrate that the proposed VBRP algorithm outperforms simulated annealing prediction and static video file prediction algorithms [?]. However, the FT value is empirical and requires further improvement. The network bandwidth prediction for selective B-frame discard relies solely on RTT, which only partially reflects network congestion. B-frame discard cannot depend entirely on RTT-derived bandwidth values, otherwise discarding loses rationality. Future work will investigate more accurate network bandwidth prediction models and analyze the impact of network bandwidth on AF values, making AF vary with network bandwidth to further improve prediction accuracy.

References

[1] Annual Meeting of Stockholders of Netflix, Inc. [EB/OL].(2017) [2018-07-10]. https://s22.q4cdn.com/959853165/files/doc_financials/annual_reports/0001065280-18-000069.pdf.

[2] Cisco visual networking index: global mobile data traffic forecast update, 2015–2020 [EB/OL]. (2016)[2018-07-10]. https://www.cisco.com/c/dam/m/en_in/innovation/enterprise/asset white-paper-c11-520862.pdf

[3] J Sullivan G, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard [J]. IEEE Trans on Circuits and Systems for Video Technology, 2012, 22 (12): 1649-68.

[4] Qi Xin, Yang Qing, Nguyen D T, et al. A context-aware framework for reducing bandwidth usage of mobile video chats [J]. IEEE Trans on Multimedia, 2016, 18 (8): 1640-1649.

[5] Sarkar U K, Ramakrishnan S, Sarkar D. Modeling full-length video using Markov-modulated gamma-based framework [J]. IEEE/ACM Trans on Networking, 2003, 11(4): 638-649

[6] Lanfranchi L I, Bing B K. MPEG-4 bandwidth prediction for broadband cable networks [J]. IEEE Trans on Broadcasting, 2008, 54(4): 741-751.

[7] Parmar H, Thornburgh M. Real-Time Messaging Protocol (RTMP) specification (version 1.0) [EB/OL]. (2012)[2018-07-10]. http://wwwimages.adobe.com/www.adobe.com/content/dam/

[8] Miller K, Bethanabhotla D, Caire G, et al. A Control-theoretic approach to adaptive video streaming in dense wireless networks [J]. IEEE Trans on Multimedia, 2015, 17 (8): 1309-1322.

[9] Kalampogia A, Koutsakis P. Using simulated annealing for improved video bandwidth prediction [C]//Proc of IEEE Conference on Computer Communications Workshops.2017: 701-705.

[10] Colonnese S, Russo S, Cuomo F, et al. Timely delivery versus bandwidth allocation for DASH-based video streaming over LTE [J]. IEEE Communications Letters, 2016, 20(3): 586-589.

[11] Sodagar I. The MPEG-DASH standard for multimedia streaming over the Internet [J]. IEEE Multimedia, 2011, 18(4): 62-67.

[12] Kalampogia A, Koutsakis P. H.264 and H.265 video bandwidth prediction [J]. IEEE Trans on Multimedia, 2017, PP (99): 1-1.

[13] Arizona State University. Trace files and statistics [EB/OL]. (2012) [2018-07-10]. http://trace.eas.asu.edu/tracemain.html.

[14] Ito M, Yoshida K, Hachiya H, et al. Quantification of the scatter distributions for liver fibrosis using modified Q-Q probability plot [C]//Proc of Ultrasonics Symposium. 2014: 2394-2397.

[15] Mascolo S, Casetti C, Gerla M, et al. TCP westwood: end-to-end bandwidth estimation for efficient transport over wired and wireless networks [C]//Proc of ACM MOBICOM.2001.

[16] Zhao Pengyu, Li Jianwei, Xi Jianxiao, et al. A mobile real-time video system using RTMP [C]//Proc of the 4th International Conference on Computational Intelligence and Communication Networks. 2012: 164-168.

[17] Huang Jian, Wu Dongmei, Liu Xiaopei. Implementation of the Rtmp Server Based on Embedded System [C]//Proc of International Conference on Computer Science and Information Processing. 2012: 193-196.

[18] Lei Xiaohua, Jiang Xiuhua, Wang Caihong. Design and implementation of streaming media processing software based on RTMP [C]//Proc of International Congress on Image and Signal Processing. 2013: 192-196.

Note: Figure translations are in progress. See original paper for figures.