Xatu: Richer Neural Network Based Prediction for Video Streaming

Yun Seong Nam
Purdue University/ Google
USA

Jianfei Gao
Purdue University
USA

Chandan Bothra
Purdue University
USA

Ehab Ghabashneh
Purdue University
USA

Sanjay Rao
Purdue University
USA

Bruno Ribeiro
Purdue University
USA

Jibin Zhan
Conviva
USA

Hui Zhang
Conviva
USA

ABSTRACT

The performance of Adaptive Bitrate (ABR) algorithms for video streaming depends on accurately predicting the download time of video chunks. Existing prediction approaches (i) assume chunk download times are dominated by network throughput; and (ii) apriori cluster sessions (e.g., based on ISP and CDN) and only learn from sessions in the same cluster. We make three contributions. First, through analysis of data from real-world video streaming sessions, we show (i) apriori clustering prevents learning from related clusters; and (ii) factors such as the Time to First Byte (TTFB) are key components of chunk download times but not easily incorporated into existing prediction approaches. Second, we propose Xatu, a new prediction approach that jointly learns a neural network sequence model with an interpretable automatic session clustering method. Xatu learns clustering rules across all sessions it deems relevant, and models sequences with multiple chunk-dependent features (e.g., TTFB) rather than just throughput. Third, evaluations using the above datasets and emulation experiments show that Xatu significantly improves prediction accuracies by 23.8% relative to CS2P (a state-of-the-art predictor). We show Xatu provides substantial performance benefits when integrated with multiple ABR algorithms including MPC (a well studied ABR algorithm), and FuguABR (a recent algorithm using stochastic control) relative to their default predictors (CS2P and a fully connected neural network respectively). Further, Xatu combined with MPC outperforms Pensieve, an ABR based on deep reinforcement learning.

CCS CONCEPTS

• **Networks** → **Application layer protocols**; **Network performance modeling**; **Network measurement**.

KEYWORDS

Video streaming; Adaptive bitrate algorithms; Predictive models; Neural Networks

ACM Reference Format:
Yun Seong Nam, Jianfei Gao, Chandan Bothra, Ehab Ghabashneh, Sanjay Rao, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. 2022. Xatu: Richer Neural Network Based Prediction for Video Streaming . In *Abstract Proceedings of the 2022 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/PERFORMANCE '22 Abstracts), June 6–10, 2022, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3489048.3522641

1 INTRODUCTION

Recent years have seen much growth in Internet video, and video traffic is expected to account for 82% of all Internet traffic by 2022 [2]. Despite much progress, consumer surveys indicate that users experience problems such as buffering when streaming video [4]. While access and core network capacity continues to grow, optimizing Internet video delivery will remain a challenge since we are only beginning to see the adoption of technologies such as 4K video [1] which may involve bitrates of tens of Megabits per second, and since there exists (and will likely remain), a wide gulf in the quality of both fixed and mobile broadband connections across households globally [3].

Video streaming today typically involves splitting video into chunks, each encoded at multiple bitrates. Clients pick bitrates for each chunk so as to balance between achieving high video quality, while avoiding rebuffering. Several Adaptive Bitrate (ABR) algorithms [5, 8–10, 12, 14, 17] have emerged to tackling this challenge. However, the design of these algorithms is complicated by the fact that clients must predict chunk download times, a challenging problem [13].

Most research in video streaming typically assumes that chunk download times are primarily determined by network throughput, and use local algorithms based on observations of prior chunks to estimate throughput [5, 9, 10, 17]. CS2P [13], a state-of-the-art approach, clusters sessions based on features such as the ISP and CDN involved in streaming, and predicts throughput based on a

SIGMETRICS/PERFORMANCE '22 Abstracts, June 6–10, 2022, Mumbai, India

ACM ISBN 978-1-4503-9141-2/22/06.
https://doi.org/10.1145/3489048.3522641

Hidden Markov Model (HMM) trained for each cluster. Yet, predicting throughput is challenging, often leading to overly conservative bitrate selections, or aggressive selections that result in rebuffering.

In this paper, we are motivated by two questions: (i) what information can be made available that can aid in predicting video chunk download times? (ii) how do we design frameworks that can leverage such information to improve the accuracy of prediction for streaming applications? We make the following **contributions**:

First, we present observations from an analysis of a dataset of nearly 100K video session traces from real users. The analysis indicates that (i) while features such as the ISP and CDN aid prediction, pre-clustering sessions based on a combination of these features and learning only from sessions in the same cluster can hurt prediction accuracy; and (ii) chunk download time depends not only on network throughput but also on factors such as the Time to First Byte (TTFB) and chunk size, which are not easy to incorporate into existing prediction approaches [13].

Second, motivated by these observations, we present Xatu[1], a prediction framework for video streaming applications. Xatu considers *static* features that do not vary during a session (e.g., ISP, CDN or city), and models sequences involving multiple *temporal* features that vary across chunks in a session (e.g., TTFB), and not just throughput. The main novel aspects are (i) Xatu learns from other sessions with similar static features, but does not apriori partition data; and (ii) Xatu proposes a gated mask neural network mechanism that uses static features to modify the output of the sequence model, and leverages neural sequence models (specifically, Long Short Term Memory networks (LSTMs) [7]) that can handle multiple temporal features. We show that our output-mask approach gives better accuracy in our task than the traditional approach of combining static and temporal features closer to the input stage of the sequence model used in other domains [6, 11, 15]. We also show that these masks help interpret the role played by static features, providing a naturally automated way of clustering sessions in contrast to pre-defining clusters [13].

Third, we evaluate Xatu using a combination of trace-driven experiments using the above dataset[2], and emulation experiments. Our results show that Xatu reduces the median of the prediction error across sessions relative to CS2P by 23.8%. The benefits come from both the ability of Xatu to exploit static features without pre-clustering data, and from its ability to model chunk-dependent features besides throughput. We integrate Xatu with multiple ABR algorithms including the widely studied MPC [17], and FuguABR [16], a recently proposed approach based on stochastic control which considers perceptual video quality. When combined with MPC, Xatu improves the median of a composite QoE metric [17] across sessions by 38.9% relative to CS2P. The resulting system outperforms Pensieve [10], an ABR algorithm that uses reinforcement learning to make bitrate selections. When combined with FuguABR, Xatu achieves significantly lower rebuffering (the 80^{th} rebuffering ratio is lowered from 10% to 0%), and higher QoE scores relative to FuguNN, a fully connected neural network proposed in [16].

We evaluate Xatu on an additional dataset obtained using controlled measurements, which also has available the CDN layer an object is served from. The results further confirm Xatu's benefits, and show that Xatu is extensible, and can easily exploit new features such as the CDN layer from which a chunk is served to further improve prediction accuracies.

[1]Named after a Pokemon character that can see into the future.
[2]We have made the data available at https://github.com/Purdue-ISL/XatuDataset

Acknowledgements We thank our shepherd, Zhi-Li Zhang, and the anonymous reviewers for their feedback which greatly helped improve the paper. We thank Humberto La Roche for his encouragement of this work. This work was funded in part by the National Science Foundation (NSF) Awards ICE-T:RC 1836889, IIS-1943364 and CCF-1918483, Purdue Integrative Data Science Initiative and Cisco.

REFERENCES

[1] Can I stream Netflix in ultra hd? https://help.netflix.com/en/node/13444.
[2] Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html.
[3] Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper. https://apps.fcc.gov/edocs_public/attachmatch/FCC-18-10A1.pdf.
[4] New research reveals buffer rage as tech's newest epidemic. https://www.prnewswire.com/news-releases/new-research-reveals-buffer-rage-as-techs-newest-epidemic-300237001.html.
[5] Zahaib Akhtar, Yun Seong Nam, Ramesh Govindan, Sanjay Rao, Jessica Chen, Ethan Katz-Bassett, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. Oboe: Auto-tuning video abr algorithms to network conditions. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '18, pages 44–58, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5567-4. doi: 10.1145/3230543.3230558. URL http://doi.acm.org/10.1145/3230543.3230558.
[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
[8] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *Proceedings of the 2014 ACM Conference on SIGCOMM*, SIGCOMM '14, pages 187–198, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2836-4. doi: 10.1145/2619239.2626296. URL http://doi.acm.org/10.1145/2619239.2626296.
[9] Junchen Jiang, Vyas Sekar, and Hui Zhang. Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 97–108, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1775-7. doi: 10.1145/2413176.2413189. URL http://doi.acm.org/10.1145/2413176.2413189.
[10] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. Neural adaptive video streaming with pensieve. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pages 197–210. ACM, 2017.
[11] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. *ICLR*, 2017.
[12] Kevin Spiteri, Rahul Urgaonkar, and Ramesh K Sitaraman. Bola: Near-optimal bitrate adaptation for online videos. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.
[13] Yi Sun, Xiaoqi Yin, Junchen Jiang, Vyas Sekar, Fuyuan Lin, Nanshu Wang, Tao Liu, and Bruno Sinopoli. Cs2p: Improving video bitrate selection and adaptation with data-driven throughput prediction. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 272–285, 2016.
[14] Guibin Tian and Yong Liu. Towards agile and smooth video adaptation in dynamic http streaming. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '12*, 2012.
[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *NIPS*, 2017.
[16] Francis Y. Yan, Hudson Ayers, Chenzhi Zhu, Sadjad Fouladi, Jam̃es Hong, Keyi Zhang, Philip Levis, and Keith Winstein. Learning in situ: a randomized experiment in video streaming. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pages 495–511, 2020.
[17] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. A control-theoretic approach for dynamic adaptive video streaming over http. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, SIGCOMM '15, London, United Kingdom, 2015.