BONES: Near-Optimal Neural-Enhanced Video Streaming

Lingdong Wang
lingdongwang@umass.edu
University of Massachusetts Amherst
Amherst, Massachusetts, USA

Simran Singh
s.singh.xzy@gmail.com
New Jersey Institute of Technology
Newark, New Jersey, USA

Jacob Chakareski
jacobcha@njit.edu
New Jersey Institute of Technology
Newark, New Jersey, USA

Mohammad Hajiesmaili
hajiesmaili@cs.umass.edu
University of Massachusetts Amherst
Amherst, Massachusetts, USA

Ramesh K. Sitaraman
ramesh@cs.umass.edu
University of Massachusetts Amherst
Amherst, Massachusetts, USA

ABSTRACT

Accessing high-quality video content can be challenging due to insufficient and unstable network bandwidth. Recent advances in neural enhancement have shown promising results in improving the quality of degraded videos through deep learning. Neural-Enhanced Streaming (NES) incorporates this new approach into video streaming, allowing users to download low-quality video segments and then enhance them to obtain high-quality content without violating the playback of the video stream. We introduce BONES, an NES control algorithm that jointly manages the network and computational resources to maximize the quality of experience (QoE) of the user. BONES formulates NES as a Lyapunov optimization problem and solves it in an online manner with near-optimal performance, making it the first NES algorithm to provide a theoretical performance guarantee. Comprehensive experimental results indicate that BONES increases QoE by 5% to 20% over state-of-the-art algorithms with minimal overhead. Our code is available at https://github.com/UMass-LIDS/bones.

CCS CONCEPTS

• **Networks** → **Network resources allocation**; • **Information systems** → **Multimedia streaming**.

KEYWORDS

adaptive bitrate streaming, Lyapunov optimization, neural enhancement, super-resolution

ACM Reference Format:
Lingdong Wang, Simran Singh, Jacob Chakareski, Mohammad Hajiesmaili, and Ramesh K. Sitaraman. 2024. BONES: Near-Optimal Neural-Enhanced Video Streaming. In *Abstracts of the 2024 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/PERFORMANCE Abstracts '24), June 10–14, 2024, Venice, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3652963.3655047

SIGMETRICS/PERFORMANCE Abstracts '24, June 10–14, 2024, Venice, Italy

ACM ISBN 979-8-4007-0624-0/24/06
https://doi.org/10.1145/3652963.3655047

1 PROBLEM FORMULATION

Traditional video streaming uses an adaptive bitrate (ABR) algorithm to transmit videos from the server to the client at the highest possible quality while ensuring continuous playback. Recently, neural-enhanced streaming (NES) algorithms have opened up the possibility of transmitting degraded videos and then improving the video quality via neural enhancement at the client. We propose **B**uffer-**O**ccupancy-based **N**eural-**E**nhanced **S**treaming (BONES), a simple and efficient control algorithm for on-demand NES. Please refer to the journal version [1] for complete details.

System Model. We illustrate the system model of BONES in Fig. 1. At each time slot t_k, BONES will decide the download quality level d_i and enhancement quality level e_j for a video segment. Then, the segment is downloaded and pushed into the download buffer Q^d, while its enhancement task is placed into the enhancement buffer Q^e. The enhancement task is executed when it leaves the buffer, producing an enhanced segment that is ready to play.

Optimization Objective. The first optimization target of BONES is to maximize the time-average utility $\bar{u}$, summing the base video quality u^d and the extra quality brought by enhancement $\tilde{u}^e$. We calculate the total utility by averaging across all K_N intervals, each with a duration T_k:

$$\bar{u} = \frac{\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{kij} d_i u^d(i,t_k) + d_i e_j \tilde{u}^e(i,j,t_k)\right\}}{\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{k=1}^{K_N} T_k\right\}}. \quad (1)$$

BONES also maximizes the time-average smoothness $\bar{s}$ in order to minimize the rebuffering ratio. Denote the temporal duration of each video segment as p; this can be formulated as:

$$\bar{s} = \frac{\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{kij} d_i e_j p\right\}}{\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{k=1}^{K_N} T_k\right\}}. \quad (2)$$

To prevent buffer overflow, we require both buffers to be rate stable, which is a relaxation of the strict buffer constraint. We then establish the download buffer constraint as:

$$\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{ki} d_i p\right\} \leq \lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{k} T_k\right\}, \quad (3)$$

and the enhancement buffer constraint as follows, where t^e denotes the computation time of an enhancement task:

$$\lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{k} d_i e_j t^e(i,j)\right\} \leq \lim_{K_N\to\infty} \frac{1}{K_N}\mathbb{E}\left\{\sum_{k} T_k\right\}. \quad (4)$$

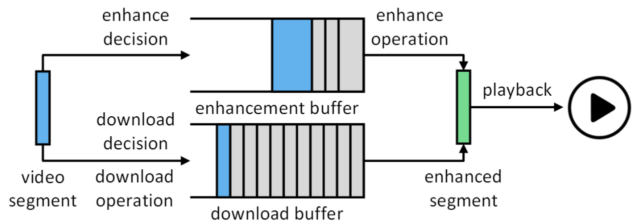

Figure 1: System model of BONES.

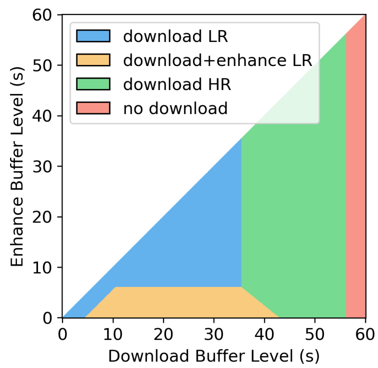

Figure 2: Decision plane of BONES.

The overall optimization objective of BONES is to maximize $\bar{u} + \lambda\bar{s}$ while meeting the buffer capacity constraints. λ is a hyper-parameter that trade-offs utility for smoothness.

2 PROPOSED ALGORITHM

Control Algorithm. BONES solves the aforementioned problem by greedily minimizing the time-average drift-plus-penalty for each time slot. Specifically, the objective function involves the Lyapunov drift, defined as:

$$O_D(t_k) = Q^d(t_k)\sum_i d_i p + Q^e(t_k)\sum_{ij} d_i e_j t^e(i,j), \tag{5}$$

and a penalty function defined as:

$$O_P(t_k) = -\sum_{ij} d_i u^d(i,t_k) + d_i e_j \tilde{u}^e(i,j,t_k) + \gamma d_i p. \tag{6}$$

We divide the objective by the time slot duration and simplify the formula. Altogether, BONES solves the following problem in each time slot, where S_i is the segment size, and V is the tradeoff factor:

$$\min_{d_i, e_j} \frac{O_D(i,t_k) + VO_P(i,j,t_k)}{\sum_i d_i S_i(t_k)}, \tag{7}$$

$$\text{s.t., } d_i \in \{0,1\}\ \forall i,\ \sum d_i \le 1,$$

$$e_j \in \{0,1\}\ \forall j,\ \sum e_j = 1.$$

This problem can be efficiently solved by traversing all combinations of download and enhancement options (d_i, e_j).

Decision Plane. BONES relies solely on the download and enhancement buffer levels $Q^d(t_k)$ and $Q^e(t_k)$ in its operation. We depict its reasonable behavior with respect to the buffer levels in Fig. 2. When the download buffer is empty, BONES quickly fills it by downloading low-resolution (LR) segments. As the buffer level increases, BONES begins requesting and enhancing LR segments to achieve higher visual quality. BONES downloads high-resolution (HR) segments for maximum quality when the buffer level is high and stops downloading if the buffer is nearly full.

Performance Bound. BONES has two performance guarantees. Firstly, it never violates the buffer capacity, as shown in Theorem 1:

THEOREM 1. *Assume $Q^d(0) = 0, Q^e(0) = 0$, and $0 < V \le \frac{(Q^d_{\max} - p)p}{u_{\max} + \gamma p}$, where $u_{\max}$ denotes the maximum utility. Then, the following holds: $Q^d(t_k) \le V\frac{u_{\max}+\gamma p}{p} + p$, and $Q^d(t_k) \le Q^d_{\max}$.*

Secondly, its performance is within an additive factor of the offline optimal solution $\bar{u}^* + \gamma\bar{s}^*$, as shown in Theorem 2:

THEOREM 2. *Assume $t^e \le t^e_{\max}$, $T_{\min} \le T_k \le T_{\max}$, $\forall k$, and $t^e_{\max}, T_{\min}, T_{\max}$ are finite. Then, we have:*

$$\bar{u}' + \gamma\bar{s}' \ge \bar{u}^* + \gamma\bar{s}^* - \frac{p^2 + (t^e_{\max})^2 + 2T_{\min}T_{\max}}{2VT_{\min}}, \tag{8}$$

where $\bar{u}' + \gamma\bar{s}'$ is the objective score of BONES.

3 EMPIRICAL EVALUATION

Simulation Results. We developed a simulation environment that examines BONES using recorded network traces. We adopt a composite QoE metric that includes visual quality, quality oscillation, and rebuffering ratio. We comprehensively compare BONES with six ABR algorithms and eight NES algorithms under six enhancement method settings and four network trace datasets. Experimental results demonstrate that BONES improves the delivered performance by 3.56% to 13.20%. BONES outperforms the default ABR algorithm of the *dash.js* video player by 7.33%, which is equivalent to increasing the average visual quality by 5.22 in the VMAF score.

Prototype System. We further developed a prototype video streaming system to verify the performance of BONES in practice. We propose heuristics that can improve its practical performance, like timing out stale download tasks and automatically searching hyper-parameters. We also provide three tradeoff options between performance and overhead, ranging from weaker performance with zero overhead to maximum benefit with 310-KB additional download size and 0.68-second startup latency. Real-world experiments show that BONES can improve the QoE by 4.66% to 20.43%.

ACKNOWLEDGMENTS

The work is supported by the National Science Foundation under awards CNS-1763617, CNS-1901137, CNS-2106463, CNS-2102963, CAREER-2045641, CNS-2106299, CCF-2031881, ECCS-2032387, CNS -2040088, CNS-2032033, and CNS-2106150; by the National Institutes of Health under award R01EY030470; and by the Panasonic Chair of Sustainability at the New Jersey Institute for Technology.

REFERENCES

[1] Lingdong Wang, Simran Singh, Jacob Chakareski, Mohammad Hajiesmaili, and Ramesh K. Sitaraman. 2024. BONES: Near-Optimal Neural-Enhanced Video Streaming. *Proc. ACM Meas. Anal. Comput. Syst.* 8, 2, Article 19 (June 2024), 28 pages. https://doi.org/10.1145/3656014