

A Cache-level Quality of Experience Metric to Characterize ICNs for Adaptive Streaming

Wenjie Li, *Student Member, IEEE*, Marwan Fayed, *Senior Member, IEEE*,
Sharief M.A. Oteafy, *Member, IEEE*, and Hossam S. Hassanein, *Fellow, IEEE*,

Abstract—Adaptive streaming has motivated Information-Centric Network (ICN) designs to improve end-user quality of experience (QoE). However, their management and evaluation rely either on conventional cache-level metrics that are poor representations of QoE, or consumer-side indicators that are opaque to network services. This paper proposes a measure to bridge the gap between cache performance and consumer QoE. We introduce *Maximal Sustainable Bitrate (MSB)*, defined as the highest bitrate deliverable in time to be in time to meet a given request without buffering. Based on our observations, we posit that QoE is maximal when requested bitrates match a cache’s MSB for that content. We design a cache-level reward function as a benchmark metric that measures the difference between requested bitrates and MSB. We hypothesize that aggregated rewards are an indicator of overall system performance. Performance evaluations show high correlation between the sum of cache rewards and consumer QoE.

Index Terms—Dynamic Adaptive Streaming; In-network Caching; Performance Metrics.

I. INTRODUCTION

CACHING systems are growing in complexity and agility to cater for increasingly diverse and heterogeneous content. As HTTP-based paradigms evolve to meet the demands of UHD and live video [1], Information-Centric Network (ICN) developments promise responsive and dynamic in-network caching [2], [3]. In this context the characterization challenges of ICN caching systems extend beyond individual cache hits and misses [4], [5], yet the performance evaluation toolset is both stale and elusive [6]. The challenges are further compounded by adaptive video streaming, where cache performance is affected by dynamics between consumer adaptations and caching systems that are poorly understood. There exists no means to characterize the inherent dependencies, nor the impact, of in-network caching on overall quality of experience (QoE) [4].

The absence of cache-level QoE measures, which is the focus of this paper, hinders the engineering of video-specific caching networks. Conventional bitrate, buffering, and oscillation measures are available to content providers. Despite ongoing research their translations to models that reflect overall QoE can be inaccurate or complex to compute [7]. Those same measures and models are opaque both to other providers and, crucially, the networks that deliver their content. In the

traditional IP space, a need has emerged in response to identify flow- and system-level fairness metrics that preserve QoE [8].

We argue an equivalent need for a cache-level metric indicative of the overall QoE delivered by ICNs. The sets of indicators available to each of the network, cache system, and end system, are inadequate. Network indicators can only characterize end-to-end flows where, in contrast to ICN caches, content must already be available. Cache performance indicators such as load and hit rate have proven reasonable and resilient when caching non-adaptive objects, but ignore consumer adaptations. In other words, conventional metrics fail to reflect changing patterns and preferences in response to a cache hit. Finally, application-oriented indicators are per-session measures viewed in isolation from the caching system. They give no indication of caching performance, and no regard for subsequent consumer or cache behaviours.

This work is first motivated by preliminary evaluations. Observations suggest that cache placement strategies can work in opposition to bitrate adaptation, and adversely affect QoE. The same observations point to ideal cache policies and placements that may instead facilitate bitrate adaptation. In the search for an ideal, how might caching ‘quality’ be quantified? The inadequacy of conventional measures and tools to evaluate and compare caching policies recently motivated frameworks to assess ICN performance [6]. However, in our wider investigations, we find no means by which to compare caching policies and placements specifically for adaptive streaming.

This paper has two main contributions. 1) We introduce *Maximal Sustainable Bitrate (MSB)* as the highest bitrate deliverable in time to be consumed. Since consumer adaptation mechanisms estimate the highest achievable bitrate, we posit that a cache hit is most meaningful when the cached content is encoded at the MSB. 2) We design a corresponding reward function as a cache-level benchmark metric that measures the distance between requested bitrates and MSB. The rewards returned by the function reflect the ‘quality’ of interaction between cache hits and subsequent video requests.

Our evaluations suggest that cache rewards generated by a cache placement meaningfully reflect the QoE provided by the caching system. Simulations show high degrees of correlation according to Pearson’s coefficients between our reward measure and conventional consumer-side QoE metrics. Statistical significance is also evaluated; p values for the null hypothesis of no correlation are calculated to be 0 in all tests.

II. CONSUMER-SIDE ADAPTATIONS AND CACHE POLICES

Bitrate adaptations occur between fixed-duration segments of content. Consumers estimate network and system conditions

Wenjie Li and Hossam S. Hassanein are with the School of Computing, Queen’s University, ON, Canada. Email: {liwenjie, hossam}@cs.queensu.ca.

Marwan Fayed is with the School of Computer Science, University of St Andrews, United Kingdom. Email: marwan.fayed@st-andrews.ac.uk

Sharief M.A. Oteafy is with the School of Computing, DePaul University, IL, USA. Email: soteafy@depaul.edu.

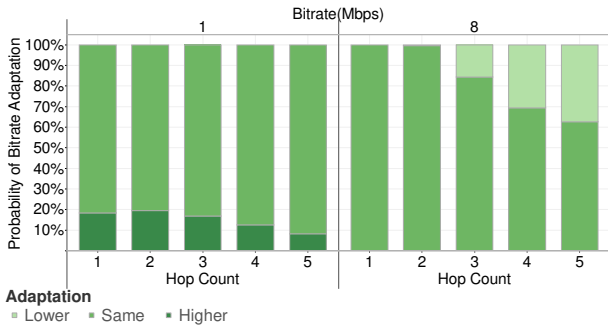


Fig. 1. Likelihood of bitrate adaptation with cache hop distance.

during downloads of the most recent segments. The bitrate of subsequent requests is selected by the consumer accordingly.

The need for ICN cache-level QoE measures is motivated by our preliminary experiments to understand the impact of cache placement on adaptive video traffic. Measurements are taken from evaluations of the benchmark *Cache Everything Everywhere* (CE2) [2] with Least Frequently Used (LFU) replacement, as described in Section IV.

The impact of cache placement on adaptive streaming is depicted by Figure 1, which gives the likelihood of bitrate adaptation as a function of hop distance from the consumer. Bitrate adaptations occur between segments. Consumers estimate network and system conditions during downloads of the most recent segments. The bitrate of subsequent requests is selected by the consumer accordingly.

The observed behaviours are best summarized by the lowest (1 Mbps) and highest (8 Mbps) video quality, as shown. Consumer adaptations are colour-coded, with the darkest shade for bitrate increases, and the lightest shade for decreases. Results indicate that adaptation-level dynamics change with cache locations, and in this example, with hop distance.

In Figure 1, the 5 left-most bars show consumer adaptations after successful requests for video content at 1 Mbps. When requests for low bitrates are satisfied within the first 4 hops, measurements indicate no significant difference in the likelihood of a bitrate increase. This suggests a degree of insensitivity to the location of low-bitrate content, with no obvious advantage to caching low-bitrate content at the edge.

A contrasting trend emerges for higher quality content. The 5 rightmost bars of Figure 1 show the likelihood of adaptation for content encoded at 8 Mbps. As hop distance increases, requests for high bitrate content increasingly trigger requests for lower bitrates. This happens because higher bitrate content consumes a disproportionately greater share of cache and network resources. As a result, service degradation for higher bitrates becomes increasingly unavoidable with hop distance. The combination of both sets of observations suggest that lower bitrate content should be moved into the network core to make room for higher bitrate content at the network edges.

A broader implication also emerges: caching policies that ignore adaptation mechanisms may hinder, rather than help, user QoE. This suggests ‘ideal’ policies for adaptive streaming that cache video content governed by relative distance or location. The same observations identify an omission in the community’s toolset: **there exists no measure to evaluate**

TABLE I
MAXIMAL SUSTAINABLE BITRATE FOR ROUTER R_i IS THE HIGHEST ENCODING DELIVERABLE TO CONSUMER C IN TIME (FOR 4S SEGMENTS).

	$(C, ..R_1)$	$(C, ..R_2)$	$(C, ..R_3)$	$(C, ..R_4)$	$(C, ..R_5)$
8 Mbps	1.6s	3.2s	5.7s	7.5s	9.1s
5 Mbps	1s	2s	3.9s	5s	6.4s
2.5 Mbps	0.5s	1s	2.9s	3.6s	4.1s
1 Mbps	0.2s	0.5s	2.1s	2.9s	3.4s

the ‘quality’ of an ICN cache placement with respect to the experience it can provide to end users.

The absence of an appropriate measure has two consequences for ICN evaluations. One consequence is that caching performance can only be assessed using consumer-side measurements. This is unrealistic in practice. Ignoring scale and sampling challenges, consumer-side measurements are unavailable to network providers. The second consequence is that cache policies cannot be compared. Given two potential policies, we have no ability to compare performance for adaptive streaming service.

In the next sections we propose a cache-level metric with a corresponding measure that mimics consumer adaptations, followed by statistical evidence of suitability and relevance.

III. A CACHE-LEVEL METRIC FOR ADAPTIVE STREAMING

Useful ICN measures for adaptive streaming must reflect cache-consumer dynamics, as well as keep pace with advancements in the understanding of that relationship. We propose a metric that follows two design principles, for dual use. First, individual measured values should reflect the impact of a cache response on subsequent requests, thus informing dynamic cache placement algorithms and optimization. Second, measurements in aggregate should characterize overall QoE being delivered by the system. A meaningful aggregate facilitates comparisons between caching schemes. The flexibility of aggregate and individual use enables the metric to evolve with consumer adaptation mechanisms, and inform caching policies.

A. Maximal Sustainable Bitrate

Figure 1 implies greater *value* for higher bitrate cache hits that occur closer to consumers. Conversely, there appears to be no less value for caching lower bitrate content upstream. This suggests a ‘best match’ between requested and cached bitrates. Intuitively, a cache hit is meaningful only if the cache can sustain the requested bitrate, i.e. when content is *delivered within the duration* of video encoded in the requested segment.

We propose the *maximal sustainable bitrate (MSB)* to evaluate cache hits, defined as the highest bitrate deliverable in time to be useful. We hypothesize that an ideal caching strategy emerges when preferred bitrates are determined based on estimates of this MSB. Note that content from upstream routers must incur additional transfer delay with greater likelihood of congestion. Consequently edge routers, where Figure 1 suggests high bitrate content should be located, will have higher maximal sustainable bitrates.

MSB estimation is realistic in practice since the duration of video segments is a known fixed value¹, and the time taken to

¹The recommended segment duration is 2-4s (Source: [Bitmovin](#), [YouTube](#)).

deliver a segment is retrievable by the respective router in the ICN. A snapshot example from our own evaluations appears in Table I. Each entry in the table is the average transfer delay for a 4s segment, of varying quality, delivered from routers R_i along the forwarding path to consumer C . In each column, the greyed entry corresponds with the bitrate that is maximally sustainable for the router. We note that MSB values monotonically decrease with hop distance from the consumer. This trend reinforces behaviours captured by Figure 1.

B. A Request-Reward Distance Metric

We propose a reward function γ that quantifies cache performance with a numerical value. The function itself represents some ideal, yet unknown, placement scheme as informed by MSB. The reward function is triggered on a per-request basis once a cache hit occurs. Thus, cache rewards are generated at a rate that mirrors consumer-side estimates of system conditions. We posit that the reward values in aggregate are an indicator of the quality of system-wide caching over all users' experience.

The reward function γ would be implemented on every ICN router and take two input parameters: (i) the consumer's requested bitrate b and, crucially, (ii) the router's Maximal Sustainable Bitrate MSB. In this study, routers maintain a MSB for each of their provider/edge router pairs, rather than per session or per content. This is because content from any single provider is similarly encoded, and that content on the forwarding path from the router to a consumer shares transmission characteristics. The reward function $\gamma(MSB, b)$ is defined as,

$$\gamma(MSB, b) = \begin{cases} \mu(b), & \text{if } b = MSB \\ \mu(b^\uparrow) * \beta(b) + \mu(b) * (1 - \beta(b)), & \text{if } b < MSB \\ \mu(b^\downarrow), & \text{if } (b > MSB) \wedge (MSB \geq b_1) \\ \mu(b_1), & \text{otherwise.} \end{cases}$$

We note that storage and transmission requirements for the encodings of any single video segment are non-uniform. In order to ensure that similar bias is reflected in the reward, μ is proportional to the base segment size. For the base bitrate at rank 1, $\mu(b_1) = 1$. Any other bitrate b is calculated as $\mu(b) = S_b/S_{b_1}$, where S_{b_1} as the size of the base bitrate segment. A bitrate b^\uparrow denotes the next higher bitrate relative to b in the set of discrete bitrates used to encode the video, while b^\downarrow denotes the next lower bitrate relative to b .

The first case triggers when the requested rate for content matches the target rate for the router, $b = MSB$. In this case there are sufficient resources to satisfy subsequent requests at the target bitrate. The reward function returns $\mu(b)$.

We shall return to the second case and proceed with the third case $b > MSB$, when a requested bitrate is greater than should be satisfied on the current cache. In this event, a lower reward than $\mu(b)$ is returned indicating that there are insufficient resources to satisfy subsequent requests at the selected rate. A reward value of $\mu(b^\downarrow)$ is returned provided that the cache can also generate sufficient throughput to satisfy the base rate b_1 . In other words, satisfying the request from this cache may cause consumers to unnecessarily reduce their

rate. The lower reward indicates that higher quality content and corresponding requests should be targeted elsewhere.

The first three cases are triggered when there are sufficient resources to satisfy the request. The final case triggers when a cache is unable to maintain even the base rate video quality. In this case the γ function returns the lowest reward of $\mu(b_1)$, since requests satisfied under such circumstances are likely to lead to buffer-induced freezing, and should be avoided.

Returning to the second case, $b < MSB$ is for requests of bitrates that are less than some ideal for the router. Referring to Table I this occurs, for example, when requests for 2.5 Mbps content is satisfied by routers R_1 or R_2 . In this case the return value is a sum of $\mu(b)$ and $\mu(b^\uparrow)$, where each value is weighted by parameter $\beta(b) \in [0, 1]$.

β relates to the placement of low bitrate content among caches. It is instructive to inspect β at its boundaries, which correspond to consumer adaptations incurred by the binary nature of cache hits and misses. $\beta(b) = 1$ represents the default notion of a cache hit for non-adaptive content that is strictly good. The consequence of a traditional cache hit encourages consumers to increase their preferred bitrate. This behaviour is reflected and encouraged by the return of the full b^\uparrow reward value. Unsustainable higher bitrates, however, lead to bitrate oscillation that adversely affect both system performance and consumer QoE. Conversely, $\beta = 0$ is no more desirable since it discourages higher bitrates despite their availability.

We next describe β in detail, before using our reward function to characterize cache-level QoE.

C. Weighted Reward Values According to β

Aggregate γ values characterize the 'quality' of a system-wide caching placement. Thus a poorly selected β may characterize performance poorly, and adversely affect subsequent caching decisions. The question then emerges: What is an appropriate weight? A fixed weight fails to capture the resource increases needed to satisfy requests for higher quality content. Linear functions are also inadequate since file sizes and delivery resources increase disproportionately with encoding rate.

We propose to define $\beta(b)$ in a manner that is inversely proportional to the rank of the bitrate, $rank(b)$, such that $\beta(b) = rank(b)^{-1}$. Encoding bitrates, as well as the resources to deliver them, increase super-linearly. As a consequence consumers are decreasingly likely to increase their preferred bitrates as quality improves. This relationship is reflected by the use of $rank(b) \in \{1, 2, \dots, k\}$ for k encoding rates.

IV. DO REWARDS REFLECT QOE?

A preliminary validation of the γ function shows high correlation with traditional consumer-side indicators of QoE.

A. Evaluation Setup

Measurements are drawn from simulations of the Named Data Networking (NDN) architecture with ndnSIM, a NS-3 based simulator. Tests are conducted over baseline caching schemes *Cache Everything Everywhere* (CE2) and Prob-Cache [9] with Least Recently Used (LRU) replacement. Each

TABLE II
SIMULATION PARAMETERS

Number of video files	200
Number of video segments per file	50
Number of NDN routers	16
Video segment duration	4 sec
Number of video consumers	32
Encoded bitrates	{1, 2.5, 5, 8}Mbps
Average time interval of video requests	400 sec
Bandwidth	20 Mbps

NDN router is allocated a Content Store (CS) from a system-wide capacity Σ CS that is tunable by parameter ω as

$$\Sigma \text{ CS} = \frac{\sum \text{Size of Video}}{\# \text{ of NDN Routers}} * \omega.$$

Consumer-side adaptation behaviour is simulated via our own implementation of FESTIVE [10], a buffer occupancy-based mechanism that captures recent advancements in bitrate adaptation. Users' interests in content vary across different videos, captured by a *Zipf* distribution (controlled via skewness parameter α). Requests for videos are triggered following a Poisson process, with an average time interval between two consecutive sessions at 400 seconds. Additional simulation parameters are listed in Table II (relevant values taken from [11]).

B. Statistical Analysis

The test design aligns with previous studies [7]. We validate the pairwise correlation between the aggregate of cache rewards taken as their sum, and the two end-user QoE metrics that show highest correlation with mean opinion score [7]. Pairwise data points are generated per session, i.e. the consumption of a complete video. Linear correlation is presented using Pearson's correlation coefficient.

The first test, *Quality-Reward (QR)*, looks for correlation between the total number of bytes that a consumer receives in a session and the sum of cache rewards generated by that session. Correlation coefficients are summarized in Table III with 95% confidence intervals, alongside a sample visual representation in Figure 2a. All but one combinations of cache size and popularity distribution reveal coefficients above 0.9.

A separate *Oscillation-Reward (OR)* test looks for bitrate oscillation, which is known to negatively impact QoE [12]. It compares the number of bitrate switches during a session against the number of times that reward values change between successive segment requests in the system. *OR* correlation coefficients are summarized in Table IV with 95% confidence intervals, alongside a representative visualization in Figure 2b. Coefficients in all cases is near or exceeds 0.8.

Strong positive correlation exists for both CE2 and ProbCache across tested parameters across cache capacity ω and interest skew α parameters. The significance p value was calculated for each experiment with a null hypothesis as "there exists no correlation between evaluated two attributes." We find that p remains 0 in all tested scenarios, reinforcing the strength of correlation that appears in *QR* and *OR* tests.

V. CONCLUSION

This paper introduced the idea of request-generated rewards to characterize adaptive streaming QoE at the cache-level. The

TABLE III
PEARSON'S CORRELATION IN QUALITY-REWARD

		$\omega = 0.05$	$\omega = 0.1$	$\omega = 0.2$
<i>CE2</i>	$\alpha = 0.8$	0.863 ± 0.003	0.909 ± 0.003	0.950 ± 0.002
	$\alpha = 1.2$	0.940 ± 0.002	0.969 ± 0.001	0.980 ± 0.001
<i>ProbCache (LRU)</i>	$\alpha = 0.8$	0.928 ± 0.002	0.961 ± 0.001	0.971 ± 0.001
	$\alpha = 1.2$	0.973 ± 0.001	0.980 ± 0.001	0.982 ± 0.001

TABLE IV
PEARSON'S CORRELATION IN OSCILLATION-REWARD

		$\omega = 0.05$	$\omega = 0.1$	$\omega = 0.2$
<i>CE2</i>	$\alpha = 0.8$	0.887 ± 0.004	0.902 ± 0.003	0.895 ± 0.003
	$\alpha = 1.2$	0.901 ± 0.002	0.890 ± 0.002	0.876 ± 0.003
<i>ProbCache (LRU)</i>	$\alpha = 0.8$	0.810 ± 0.004	0.805 ± 0.003	0.793 ± 0.002
	$\alpha = 1.2$	0.805 ± 0.002	0.801 ± 0.003	0.789 ± 0.004

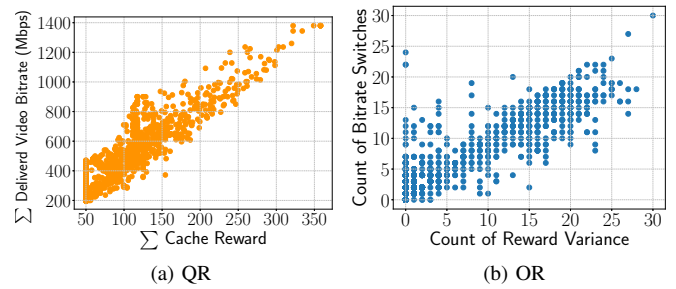


Fig. 2. Correlation between reward and QoE (CE2, $\omega = 0.1$, $\alpha = 0.8$).

reward function bridges the gap between metrics of cache performance available in the system and consumer-side QoE. The reward function is tunable, so that the function may evolve as the interactions between caching and consumer-side adaptive streaming become better understood. Simulations shows high correlation as indicated by Pearson's coefficients.

REFERENCES

- [1] BBC R&D, "Dynamic adaptive streaming over ip multicast: Distributing live television at scale over the internet." Source: last accessed 10/2018.
- [2] A. Ioannou and S. Weber, "A survey of caching policies and forwarding mechanisms in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2847–2886, 2016.
- [3] C. Xu, M. Wang, X. Chen, L. Zhong, and A. L. Grieco, "Optimal information centric caching in 5g device-to-device communications," *IEEE Transactions on Mobile Computing*, 2018.
- [4] C. Westphal, S. Lederer, D. Posch, C. Timmerer *et al.*, "Adaptive video streaming over information-centric networking (icn)," RFC 7933, IRTF, 2016, [Online]. Available: <http://www.rfc-editor.org/rfc/rfc7933.txt>.
- [5] W. Li, S. M. Oteafy, and H. S. Hassanein, "Rate-selective caching for adaptive streaming over information-centric networks," *IEEE Transactions on Computers*, 2017.
- [6] E. Ramadan, P. Babaie, and Z. Zhang, "A framework for evaluating policies in a hierarchical network of caches," in *IFIP Networking*, 2018.
- [7] Z. Duanmu, A. Rehman, and Z. Wang, "A quality-of-experience database for adaptive video streaming," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 474–487, June 2018.
- [8] A. Mansy, M. Fayed, and M. H. Ammar, "Network-layer fairness for adaptive video streams," in *Proc. IFIP Networking*, 2015.
- [9] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *ACM ICN*, 2012, pp. 55–60.
- [10] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive," in *ACM CoNEXT*. ACM, 2012, pp. 97–108.
- [11] YouTube Help, <https://support.google.com/youtube/answer/1722171?hl=en>, [Online; accessed 24-July-2018].
- [12] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," in *Proc. ACM SIGCOMM*, 2011.