Optimal Set of Video Representations in Adaptive Streaming

Laura Toni
EPFL, Switzerland

Ramon Aparicio-Pardo
Telecom Bretagne, France

Gwendal Simon
Telecom Bretagne, France

Alberto Blanc
Telecom Bretagne, France

Pascal Frossard
EPFL, Switzerland

ABSTRACT

To address the increasing and heterogenous demand of multimedia content over the Internet, adaptive streaming has been developed as a versatile solution to encode a video stream into different versions, each one catering to a specific set of users. While most of the efforts so far has been focused on optimal playout-control strategies at the client side, in this paper we concentrate instead on the providers' side. We study the set of parameters at which representations should be encoded, showing need of an optimal selection of this set. In particular, we formulate an integer linear program that maximizes users' average satisfaction, taking into account the network characteristics, the type of video content, and the user population. The solution of the optimization is a set of encoding parameters that outperforms the commonly used vendor recommendations, in terms of user satisfaction and total delivery cost. Results show that video content information as well as network constraints and users' statistics are fundamental knowledge to select proper encoding parameters to provide fairness among users and reduce network consumption. By combining patterns common to several representative cases, we propose a few practical guidelines that can be used to choose the encoding parameters based on the user base characteristics, the network capacity and the type of video content.

1. INTRODUCTION

The population of users who consume video on the Internet has become more heterogeneous in terms of requested content, of network connections, and of devices. Adaptive streaming solutions aim to address this growing heterogeneity by offering users several versions of the video contents. Each version is encoded at different bitrates and resolutions so that any user can select the most suitable data based on her streaming client and her network conditions.

Figure 1 illustrates an instance of an adaptive streaming system. The ingest server receives video data from cameras and prepares several different video *representations*, which are mainly characterized by a spatio-temporal resolution and a bit-rate. The ingest server then sends the representation package to the origin server of the content delivery network (CDN), which delivers the video representations to the edge-servers, directly connected to the end-users. On the other end, media clients send requests for video data that are available at the edge-servers. Several models have been recently proposed to formalize the adaptive streaming communication framework, like DASH [2, 3] or WebRTC [4]. Implementations of such systems mostly differ on the client adaptation strategy, or on the definition of the different video representations. While many research efforts have been invested on the first problem, the second one has been overlooked in the literature, creating a gap between the advances in clients strategies and in the ingest servers policies.

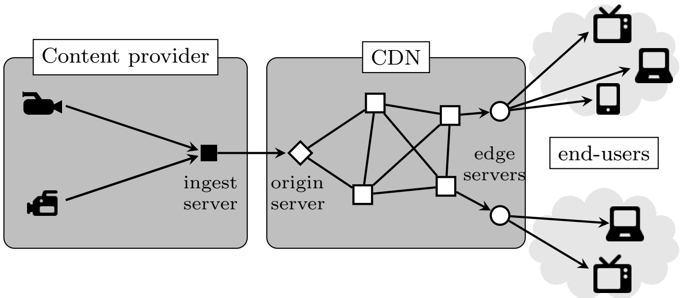


Figure 1: Live streaming: the delivery chain.

We aim at filling this gap, focusing on the *set of representations* that should be generated by the ingest server. Today, the only existing guidelines for selecting the parameters of representation set are the *recommendations* from system manufacturers, including Apple [5] and Microsoft [6]. Some content providers have also defined their own representation sets, for example Netflix [7]. However, to the best of our knowledge neither the recommendations from system manufacturers nor the choices made by content providers have been supported by any scientific study. In this paper, we show that the existing recommended sets have critical weaknesses, which calls for a better selection of the encoding parameters. However, deciding encoding parameters for representation sets is an open problem, which deals with multiple correlated constraints, including the cost of delivering video streams in a CDN, the characteristics of end-users, and the type of video in input. For example smaller sets might satisfy only a fraction of the users, while larger ones could satisfy more users. However, larger sets come at the price of increased storage costs, or larger encoding delays in the case of live-streaming. It is therefore important to study how representation set should be designed, in order to strike the appropriate balance between user satisfaction and system costs.

In this paper, we provide a theoretical study on the selection of the key encoding parameters for representation sets in an adaptive streaming system. Our contributions are as follows:

- We formulate an *optimization problem*, specifically an integer linear program (ILP), in order to maximize the average user satisfaction, given network and system constraints. The satisfaction function of clients is formalized as a function of the encoding rate, the resolution, and the content characteristics of the requested video. By using a generic solver, it is possible to solve the ILP on representative cases, gaining insights about the optimal representation sets.

- We use the ILP to study how far from the optimal are recommended sets. We compute the solution of the ILP for different user populations and compare it to the representations selected by existing recommendations. Our results show that recommended sets well perform when both the population of users and the catalogue of videos correspond to the target of each system. However, these recommended sets require too many representations and do not easily adapt to other contexts.

- We analyze the optimal representation sets in different scenarios, which leads us to provide insights on how a system provider should decide encoding parameters for its representation sets. We consider a generic system with respect to important characteristics like user population (number of devices of each type: smart phones, tablets, etc.), considered network (connection of each client and overall CDN capacity), and type of video (sport, documentary, movie, cartoon). By analyzing the solutions of the ILP in different scenarios, in which both network constraints and users statistics vary, we notice recurrent patterns. We extrapolate these insights providing *guidelines*, which can be useful for content providers in the selection of the best representation set.

Overall, we propose a theoretical framework to optimize representation sets in any possible scenario. Any system configuration can be considered as input, leading to a generic framework that can be used by any content provider. We further demonstrate the need of making the selection of the representation set based on the video content, network, and clients characteristics. The remainder of this paper is organized as follows. Related works on adaptive streaming are described in Sec. 2. Formalization of the optimization problem as ILP is provided in Sec. 3. In Sec. 4, we detail the simulation settings. In Sec. 5, results are provided to study the system performance of optimal representation sets w.r.t the recommended one. In Sec. 6 we provide analysis results of the behavior of the optimal set across different configuration to derive useful guidelines. Finally, conclusions and future works are discussed in Sec. 7.

2. RELATED WORKS

In the past decade, adaptive streaming has received a lot of attentions from the research community for systems that were mostly governed by the streaming server intelligence. Recently, a new paradigm based HTTP-adaptive streaming [2, 3], where the clients drives the streaming decision, has been developed.

The development of this new framework has attracted quite some research efforts, which have been mostly carried out to optimize the client-driven resource allocations strategies. Works have focused on the optimization of the best the representation request for each user [8,9] based on a proper estimation of the network dynamics [10] and on the control of the client buffer status; the general objective is to maximize the quality of the stream while avoiding unnecessary quality fluctuations. In [11] for example, the selection of the representation is optimized in such a way that each large variation of rates in successive segments are avoided since large rate variations, in front of varying network conditions, may lead to low Quality of Experience (QoE) levels. Beyond variations of QoE levels, timing aspects in real-time applications have been also investigated in order to minimize the re-buffering phases [8]. Researchers have also investigated the performance of HTTP-based adaptive streaming in systems with a multitude of users. As proved in experimental tests provided in [12–14], the current HTTP-adaptive streaming systems have limitations when a multitude of clients share the same network. For example, they cannot reach simultaneously fairness and efficiency in a scenario where a multitude of clients share a bottleneck channel.

In most these recent research works, however, the design of optimal representation sets is usually overlooked and representations rates have been considered as a priori recommendations, as in the case of Apple [5], Microsoft [6], or Netflix [7]. To the best of our knowledge neither the recommendations from system manufacturers nor the choices made by content providers have been supported by any scientific study. The encoding rate optimization has been investigated very recently in [15], where the encoding rates of archived videos in a storage-limited server scenario are optimized in such a way that the best possible QoE is provided to a pool of users and a total storage capacity constraint is met. Homogenous users are considered in the investigated scenario and the optimization solving method is intrinsically linked to the users uniform distribution. In our work we rather look at the optimization of the representation sets in adaptive streaming applications. Beyond content information, we also include in the optimization problem network state information and users' popularity characteristics, assuming also heterogeneous-users scenarios.

Notation	
$f_{vrs} \in \mathcal{R}^+$	Satisfaction level for the representation encoded at rate r and resolution s of the video v
$b_r \in \mathcal{R}^+$	Value in *kbps* of the encoding rate r
$b_{vs}^{\min} \in \mathcal{R}^+$	Value in *kbps* of the minimum encoding rate that the video v at resolution s can admit.
$b_{vs}^{\max} \in \mathcal{R}^+$	Value in *kbps* of the maximun encoding rate that the video v at resolution s can admit.
$c_u \in \mathcal{R}^+$	Maximum internet connection capacity in *kbps* of user u
$v_u \in \mathcal{V}$	Video channel requested by user u
$s_u \in \mathcal{S}$	Spatial resolution requested by user u
$C \in \mathcal{R}^+$	Overall network capacity hired by the CDN in *kbps*
$K \in \mathcal{R}^+$	Overall number of used representations, i.e. triples (v, r, s), used by the CDN
$P \in [0, 1]$	Minimum ratio of users u required to be served

Table 1: Notation adopted in the ILP formulation.

3. PROBLEM FORMULATION

We now provide the problem formulation for selecting the best representation set by taking into account network, users, and video content information. The behavior of both the network and the clients are modeled based on a statistical model of the system. These considered statistics (i.e., network capacities, content, and clients statistics) are constant over the complete asset and known a priori. Although adaptive streaming systems have been deployed for dynamic systems, we argue here that an a priori optimization framework allows a better understanding of the key elements that are at stake when implementing adaptive systems. It is thus able to reveal the efficiency of existing strategies in a fair manner, showing average behaviors. We also believe that encoding parameters should be set for relatively long periods of time (e.g., a few hours) in order not to disrupt viewers watching videos for prolonged periods.

In the following, first we introduce notations used in our problem and constraints imposed in the optimization. Then, we provide the ILP model to characterize and solve this problem.

3.1 Definitions

Let $\mathcal{V}$ be the set of possible video. Each video channel $v \in \mathcal{V}$ can be encoded into different representations, each of them characterized by the encoding rate $r \in \mathcal{R}$ and the spatial resolution $s \in \mathcal{S}$, being $\mathcal{R}$ and $\mathcal{S}$ respectively the sets of bit rates and spatial resolutions used to generate the representations. In our model then the triple (v, r, s) corresponds to the representation of a video $v \in \mathcal{V}$ encoded at a resolution $s \in \mathcal{S}$ and at a bit rate $r \in \mathcal{R}$. Each resolution s admits encoding rates within the range [$b_{vs}^{\min}$, $b_{vs}^{\max}$] for video v. We also denote by b_r the value (in *kbps*) of the encoding rate r.

Let $\mathcal{U}$ be the set of users that the CDN network should serve, where each user $u \in \mathcal{U}$ requests a video channel $v_u \in \mathcal{V}$ at a given resolution $s_u \in \mathcal{S}$ by means of an Internet connection with a capacity of c_u *kbps*. We assume that each user is associated with one single video resolution.

An arbitrary user watching video v at resolution s experiences a satisfaction level of $f_{vs}(r)$, which is an increasing function of the bit rate r, ranging from 0 to 1. Note that the satisfaction function depends on the resolution. For example, for a user watching a video v at resolution s, $f_{vs}(r) = 1$ if $b_r = b_{vs}^{\max}$, while the same rate might lead to a different satisfaction for the same video content but displayed at different resolutions. For sake of clarity in the notation, in the following we denote the satisfaction level by f_{vrs} rather than $f_{vs}(r)$.

Equipped with the above definitions, we define the optimal encoding parameters set as the one which maximizes the overall user satisfaction, subject to several constraints imposed by both the delivery system and the service provider. The constraints that we formulate for this problem directly derive from the real challenges that have been identified by service providers. We especially emphasize three constraints:

- **The overall CDN capacity** available to successfully deliver the video representations. We denote this overall capacity by C and it is expressed in terms of *Mbps*. In general, video service providers reserve an overall budget (in \$) for video delivery, this budget corresponding to an overall bandwidth that is negotiated with the CDN [16]. Thus, the manager of a video service provider is interested in maintaining the overall delivery bandwidth below a given value, here C.

- **The overall number of representations**, i.e. the overall number of triples (v, r, s) provided to ingest servers. Let K be the maximum number of representations that are produced by the system. In short, a larger representation set means more complexity and higher costs for the video service provider. Complexity comes from more data to handle, log, store and deliver while cost directly derives from the number of machines that have been provisioned to encode raw video in inputs. To justify this constraint, let us recall that some video service providers face some challenging issues related to scalability. Typically, a website like justin.tv has about 4,000 video channels simultaneously [17].

- **The ratio of users that must be served**. Ideally, the service provider should guarantee that a large fraction of the users is served. To take into account this need in our formulation, we introduce another constraint, taking into account the minimum ratio of users that a service provider would like to serve. We denote this ratio value by P. This constraint allows to set the level of fairness among users that service providers want to impose. The fairness definition is simplified but it conforms to the reality of operational team in video service providers. Furthermore, our problem formulation is general enough and can include any other fairness constraint.

Notations are summarized in Table 1.

3.2 ILP Model

We now describe the proposed ILP. The decision variables in the model are:

$$\alpha_{uvrs} = \begin{cases} 1, & \text{if user } u \text{ is served by a representation} \\ & \text{of video } v \text{ at resolution } s \text{ and rate } r, \\ 0, & \text{otherwise.} \end{cases}$$

Integer Linear Programming formulation

$$\max_{\{\boldsymbol{\alpha},\boldsymbol{\beta}\}} \sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} f_{vrs}\cdot\alpha_{uvrs} \tag{1a}$$

$$\text{s.t. } \alpha_{uvrs} \leq \beta_{vrs}, \qquad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1b}$$

$$\beta_{vrs} \leq \sum_{u\in\mathcal{U}} \alpha_{uvrs}, \qquad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1c}$$

$$(b_{vs}^{\min} - b_r)\cdot\beta_{vrs} \leq 0, \qquad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1d}$$

$$(b_{vs}^{\max} - b_r)\cdot\beta_{vrs} \leq 0, \qquad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1e}$$

$$\sum_{r\in\mathcal{R}} \alpha_{uvrs} \leq \begin{cases} 1, \text{ if } v = v_u \\ \quad \& \; s = s_u \\ 0, \text{ otherwise} \end{cases} \qquad u\in\mathcal{U}, v\in\mathcal{V}, s\in\mathcal{S} \tag{1f}$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} b_r\cdot\alpha_{uvrs} \leq c_u, \qquad u\in\mathcal{U} \tag{1g}$$

$$\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} b_r\cdot\alpha_{uvrs} \leq C, \tag{1h}$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \beta_{vrs} \leq K, \tag{1i}$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \alpha_{uvrs} \geq P\cdot|\mathcal{U}|, \qquad u\in\mathcal{U} \tag{1j}$$

$$\alpha_{uvrs} \in [0,1], \qquad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1k}$$

$$\beta_{vrs} \in [0,1], \qquad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1l}$$

$$\beta_{vrs} = \begin{cases} 1, & \text{if any user in the system is being served} \\ & \text{by a representation of video } v \text{ encoded} \\ & \text{at resolution } s \text{ and at rate } r, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the optimization problem can be formulated as shown in (1).

The objective function (1a) maximizes the overall user satisfaction. The constraints (1b) and (1c) set up a consistent relation between the decision variables α and β. The constraints (1d) and (1e) force to zero some β variables. They ensure that each video v at resolution s only is encoded at the bit rates in the range between the minimal and maximal admissible rates for the video v at resolution s. The constraints (1g) and (1h) resptectively limit the user link capacity and the overall network capacity. The constraint (1i) fixes the maximal number of representations. The constraint (1f) associates to each user u the demanded video channel v_u at the requested resolution s_u. And, finally, the constraint (1j) force us to server at least a certain ratio P of users.

3.3 Generalization of the model

The ILP formulation introduced above could be easily extended. In particular, one may argue that both rates and resolutions are not the only parameters that characterize a representation. Indeed, it is possible to consider also the required decoding and encoding CPU and GPU, the size of buffer at client side, or even more specific parameters like the library codec that should be installed at the client side.

Nevertheless, what we observe is that rate and resolutions are not only critical parameters (which are undebatable when identifying the key video encoding parameters), but also parameters that impose constraints that can be generalized to any other constraint. Typically, the required decoding CPU is exactly the same constraint as the constraint (1g) for network connectivity at the client side. Similarly, a limitation of the encoding CPU at the ingest server can be expressed with constraint (1h).

Therefore, we have preferred not to increase the complexity of the proposed formulation since current constraints and encoding rate and resolutions are enough to capture the main features of the optimization problem. However, it is always possible to formulate a more general problem by adding further client-side (respectively delivery system-side) constraints.

4. NUMERICAL ANALYSIS SETTINGS

In the present work, we use the ILP model introduced in the previous section as tool to perform a comprehensive numerical analysis on the optimal selection of the encoding parameters for representation sets. The ILP model is solved by the generic solver IBM ILOG CPLEX [18]. With this, we obtain optimal representation sets that can then be compared to the recommended ones and can be analyzed to provide (hopefully useful) guidelines. To this end, we define different *configurations*, described in this section, which are used in our tests. First, we present how the *user satisfaction metric* is evaluated. Second, we explain how the *user population* is synthetically generated. Finally, we describe the default settings of our tests.

4.1 User satisfaction

We characterize each video content at a given resolution by one *satisfaction function* that reflects the QoE as a function of both the rate and the resolution. How to model this behavior has been investigated in several works and a uniformly accepted model still has to be accepted [19]. In our case, we provide the satisfaction curve as an Video Quality Metric (VQM) score [20], which a full-reference metric that has higher correlation with human perception than other MSE-based metrics.

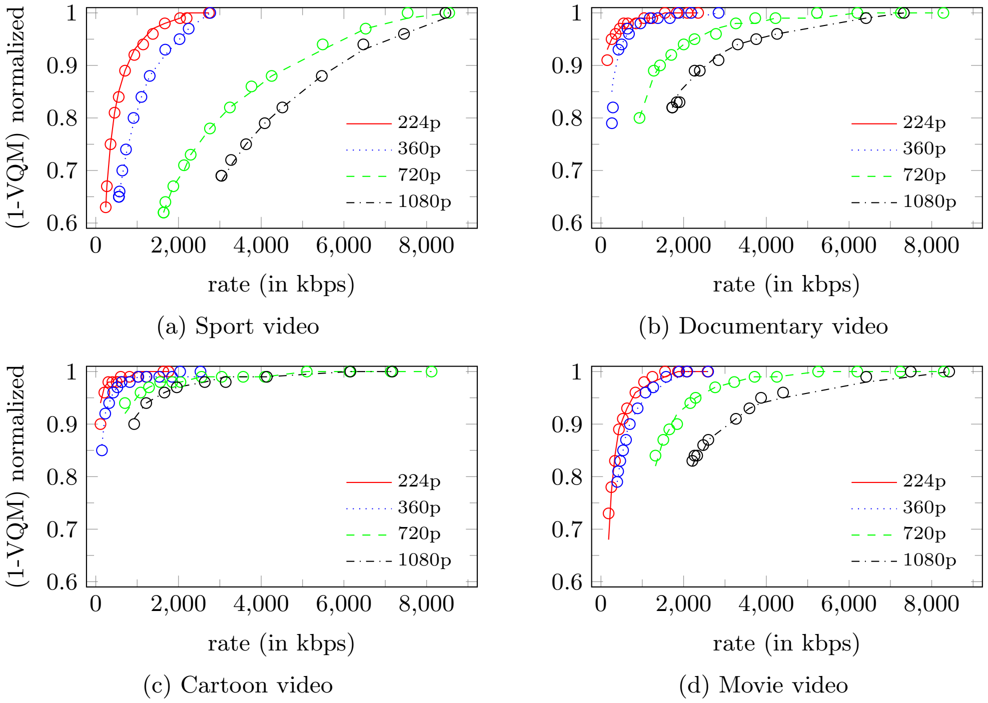
We evaluated the VQM score for four different test sequences from [21] at four different resolutions. Each of these four test sequences corresponds to a representative video type. The tested sequences and resolutions are provided in Table 4. Since the VQM score ranges from 0 to 1, representing the best and the worst QoE, respectively, we associate user satisfaction level with (1 − VQM) score. The empirical measures obtained from evaluating the aforementioned sequences are depicted as circles in Fig. 2. From these measures, we derived a satisfaction function by curve fitting. In this extrapolated function, the satisfaction level of each user receiving a video at rate R is modeled as follows

$$f(R) = a * R^b + c. \tag{2}$$

In Table 2, we provide the parameters a, b, and c used in the fitting for each video and resolution. Satisfaction curves evaluated from Eq. 2 are plotted as continuous lines in Fig. 2.

4.2 User population

We now describe the generation of a synthetic *user populations (sets $\mathcal{U}$)*. User population of different content providers can be very different in terms of video popularity, type of used devices and network connectivity. A synthetic generation of it offer us an opportunity of creating populations under a common framework of parameters.



(a) Sport video

(b) Documentary video

(c) Cartoon video

(d) Movie video

Figure 2: Curve fitting for all the considered videos. The circles are real measures taken from the video while the lines reflect the model.

Video: Big Buck Bunny			
Resolution	a	b	c
224	-1.897125	-0.703675	1.01
360	-48.287172	-1.169053	1.00
720	-1425.351349	-1.501161	1.00
1080	-244.124234	-1.144599	1.01
Video: Snow Mountain			
Resolution	a	b	c
224	-1.056339	-0.471450	1.03
360	-576.987743	-1.477734	1.00
720	-4307.239812	-1.452866	1.01
1080	-1407.140911	-1.177391	1.04
Video: Rush Field Cuts			
Resolution	a	b	c
224	-40.246497	-0.824477	1.07
360	-26.016439	-0.606764	1.21
720	-17.593112	-0.421462	1.40
1080	-57.332200	-0.546566	1.40
Video: Old Town Cross			
224	-88.612999	-1.057453	1.03
360	-56.653398	-0.893399	1.06
720	-775052.600233	-2.118902	1.01
1080	-44331.026196	-1.599378	1.02

Table 2: Parameters for the QoE model.

This feature allows us to run our performance analysis by changing the values of these parameters in a systematic way. Considering other approaches, e.g., using a variety of real traces, would have been possible, but it would have made our targeted analysis of the optimal set more difficult.

A user $u \in \mathcal{U}$ is characterized by three parameters: her requested video channel v_u, her requested resolution s_u and her local network capacity c_u. These three parameters are assigned as follows.

- For what concern the choice of the four video channel types indicated in Table 4, users are randomly assigned to one of these videos types with the same probability, *i.e.* 1 out of 4. This defines v_u.

- We set that users make use of four categories of devices, each of them associated to one displayable resolution. In particular, for smartphone, tablet, laptop and HDTV, the allowed resolution is 224p, 360p, 720p and 1080p, respectively. Again, users are randomly assigned to one of these devices with the uniform probability, *i.e.* 1 out of 4. This defines s_u, since devices and resolutions are associated by a bijective application.

- The last user property, the connection capacity c_u, is created according to the information shown in Table 3. First, users are randomly assigned to one of these Internet connections by following the discrete probably distribution corresponding to the “Ratio of users” column in Table 3. Once known the user connection, the connection capacity of user c_u is uniformly distributed between the minimum and the maximum bandwidth of the connection. These delimiting values depend on the connection type and they are in the two first columns of Table 3.

4.3 Default settings

We conclude this section by detailing the default settings, which will be used hereafter in the numerical analysis. In the following, these settings remain unchanged unless other

Technologies	Minimum bandwidth (in Mbps)	Maximum bandwidth (in Mbps)	Ratio of users
Wifi-Hotspot	0.15	0.8	30%
3G	0.4	4	20%
ADSL-slow	0.3	3	10%
ADSL-fast	0.7	10	30%
FTTH	1.5	25	10%

Table 3: Technologies of users connections.

mention. The video catalog $\mathcal{V}$ and spatial resolution set $\mathcal{S}$ correspond to the video sequences and resolutions indicated in Table 4. The set of bit rates $r \in \mathcal{R}$ ranges from 150 *kbps* up to 8,650 *kbps* with steps of 50 *kbps*, which implies 171 possible values. The minimum and maximum encoding rate for each video v and each resolution s $b_{vs}^{\min}$ and $b_{vs}^{\max}$ are shown in Table 5.

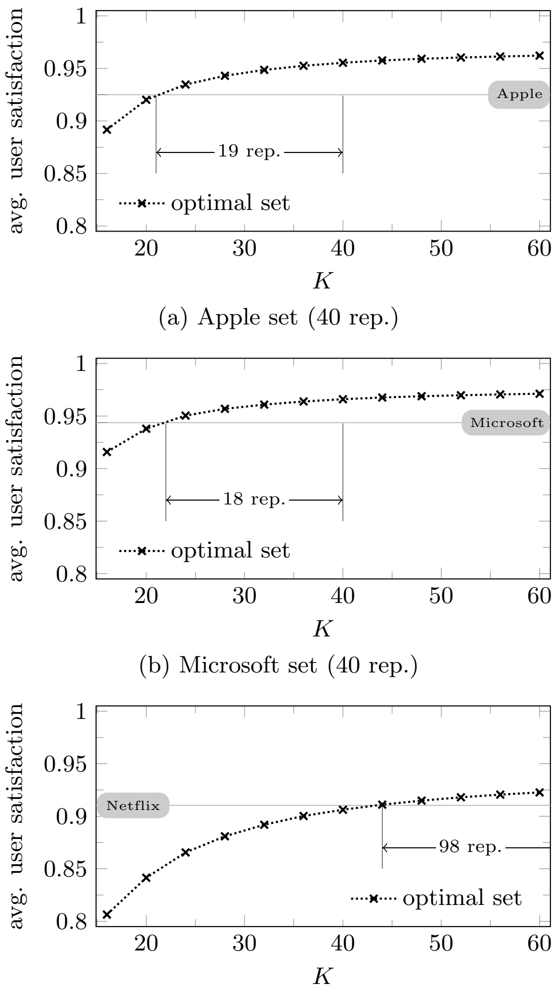
The satisfaction coefficients f_{vrs} are fixed for each triple (v, r, s) according to the extrapolated satisfaction curves plotted in Fig. 2. In our tests, we use five instances of user population sets $\mathcal{U}$, synthetically generated following in such a way that $|\mathcal{U}| = 500$, with $|\mathcal{U}|$ being the cardinality of the set. We also consider that for all generated configurations, $C = 5000$ *kbps*, $K = 60$, and $P = 0.95$.

Finally, we would like to the mention that, for instances created according to these settings, CPLEX was able to solve the ILP model in the order of a few minutes in an Intel(R) Xeon(R) CPU E5640 @ 2.67GHz with 24 GB of RAM.

5. HOW FAR FROM THE OPTIMAL ARE RECOMMENDED SETS?

As already mentioned in previous sections, today's system engineers commonly select encoding parameters for their representations following given recommendations, which are not optimized based on content or context information and which should be versatile enough to apply to any possible scenario. In this section we provide results of a comprehensive numerical analysis that we conducted to answer a critical question: *how far from the optimal are recommended sets?* With our ILP, we are able to determine the optimal representation set for any *configuration* (video catalog, user population, delivery system characteristics), evaluating the performances of any existing solution vis-à-vis the optimal one. In the following, we focus on three recommended representation sets: Apple [5, 22] for HTTP Live Streaming (HLS) (see Table 6), Microsoft [23] for Smooth Streaming (see Table 7), and Netflix [7, 24] (see Table 8).

In Fig. 3, we show the average user satisfaction as a function of the number of representations K in the optimal solution. The setting of other parameters conform the description given in Sec. 4. The gray horizontal line indicates the average user satisfaction obtained when the representation set follows the recommendations. Note that three different figures are provided, one for each recommended set. This is due to the fact that each recommendation has its own range over which video resolutions and rates are defined. Apple recommendations typically accommodate smartphones and tablets while Microsoft target more specifically laptops and home computers. Thus, we characterize one user population per recommended set. For each population, we compute the optimal representation set, and compare the resulting performance with the corresponding recommendations.



(a) Apple set (40 rep.)

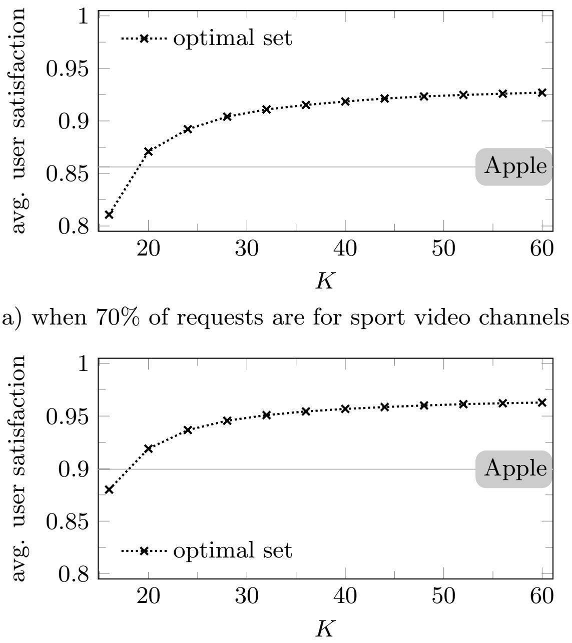
(b) Microsoft set (40 rep.)

(c) Netflix set (132 rep.)

Figure 3: Average user satisfaction: recommended sets vs. optimal sets with different number of representations

In Fig. 3, we observe that the recommended sets are able to achieve an average satisfaction level not necessarily lower than the one obtained with the optimal set. However, with respect to the optimal set, **the recommended sets need a much larger number of representations to reach a globally good user satisfaction.** We highlight by an arrow the difference in terms of number of representations between the recommended sets and the optimal sets. The average user satisfaction of 0.92 (respectively 0.945) obtained by Apple's (respectively Microsoft's) 40 representations can be obtained with 21 (respectively 22) representations in the optimal set, so roughly half the number of representations. It is worthy to recall that the more representations in the set, the more complex and costly is the encoding and delivery system.

For the case of Netflix, the result is even more critical. Netflix's representation set contains 132 representations although the same average user satisfaction (about 0.91)



(a) when 70% of requests are for sport video channels

(b) when 70% of users watch video on smartphone

Figure 4: Average satisfaction of users for the representation sets recommended by Apple in different contexts.

can be obtained with 34 representations in the optimal set. This corresponds to a reduction of about 70% in terms of representation set dimension.

We now study how far recommended sets are from optimal ones from a different perspective. In particular, we are interested in investigating how versatile the recommended sets are for different populations and different video catalogs. To measure the ability of performing well in different configurations, in Fig. 4 we depict the average user satisfaction as a function of K when two parameters differ from the parameters given in Sec. 4: both users population and video requests are not necessarily uniformly distributed. In Fig. 4(a), the popularity of videos is not the same across video types, in particular the sport video channel gets 70% of user requests. Note however that users population is uniform in terms on devices. On the other hand, in Fig. 4(b), users population is not uniform in terms of devices (70% of users watch their videos from a smartphone) while video channels requests are uniformly distributed. For sake of brevity, we compare the optimal set only with Apple's recommended sets but similar results were obtained with other recommended sets.

We can observe that, while in homogenous scenarios (Fig. 3(a)) recommended sets perform closely to optimal ones (for some large K), **the performance of recommended sets degrades when the configuration is less homogeneous.** In Fig. 4(a), Apple's recommendations experience a satisfaction level of about 0.85 while the optimal ones achieve a floor satisfaction level at about 0.92. In the analogous scenario in Fig. 4(b), an optimal set is able to reach a 0.97 of satisfaction level, while Apple's recommendations result in a relatively poor 0.9 score. Note that in our model each representation (v, r, s) is always defined such that $b_r \in [b_{vs}^{\min},\, b_{vs}^{\max}]$. From Fig. 2, it can be observed that in the range $[b_{vs}^{\min},\, b_{vs}^{\max}]$ most of the satisfaction values are between 0.7 and 1. This means that a 0.1 gain in terms of satisfaction level is already a very good improvement in our system.

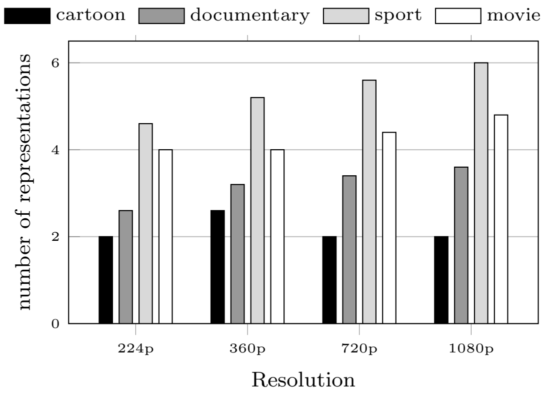


Figure 5: Average number of representations per resolution, for each type of videos.

6. GUIDELINES

From our numerical analysis of optimal representation sets evaluated across different configurations, we now derive four *guidelines*. All results provided in the following have been carried out with the default configuration model described in Sec. 4.

Guideline 1: How many representations per video?

The repartition of representation among videos needs to be content-aware. Put emphasis on the videos that are the more complex to encode.

A supposed weakness of recommended representation sets is that the number of representations is the same for any video. In Fig. 5, we show the average number of representations dedicated to any video type as a function of the video resolution for the optimal representation sets.

We observer that some videos clearly require more representations than others: about 21 in average for sport videos while only about $8 - 9$ representations in average for cartoon sequences. This is justified by the fact that the sport video has more complexity in the scene, leading to a wider range of QoE values than for the cartoon. Such analysis is straightforward when one looks at the differences between user satisfaction curves in both Figure 2(c) and Figure 2(a): for any given pair of bit-rates, the QoE gains is larger for sport video than for cartoon.

To confirm that these results are not biased by our default configuration, we changed the popularity of the videos in the catalog. Four video types are still considered, i.e., documentary, movie, sport and cartoon, but only 0.1 of users watch documentary, other 0.1 watch movie, and the remaining is shared between cartoon and sport videos. More precisely, x is the ratio of users watching sport videos, and $0.8 - x$ is the ratio of users watching cartoon. In Fig. 6, the parameter x ranges from 0 (no sport videos) to 0.8 (no cartoon videos). We measure the distribution of

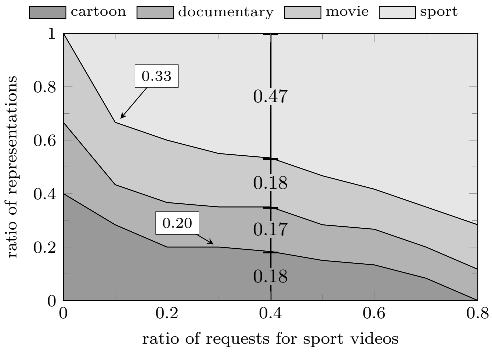


Figure 6: Distribution of representations per video.

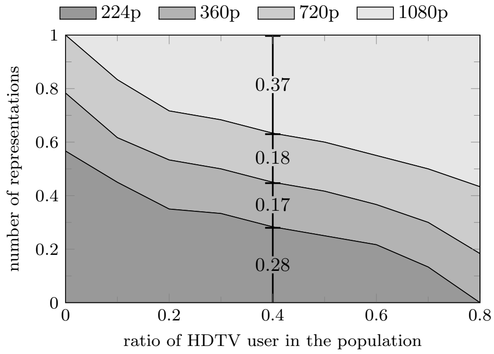


Figure 7: Distribution of representations per resolution.

the number of representations over the different videos when $K = 48$. In other words, Figure 6 shows, out of the 48 representations, how many are for dedicated to each type of videos.

Figure 6 confirms our previous observation. Cartoon videos (respectively sport videos) are under-(respectively over) represented indifferently from the popularity. For example, when sport videos are only watched by 0.1 of users, one third of representations are used by sport videos. On the opposite, cartoon videos are less that one fifth of the total of representations even when cartoon are requested by half of the population. This reveals that the QoE user satisfaction function of videos is a critical input for the setting of representation sets.

Guideline 2: For a given video, how many representations per resolution?

It mainly follows the distribution of devices in user population. Put a slight emphasis on highest resolutions.

For a first study of the representations distribution per resolution, we can refer again to Fig. 5. For a given video, the number of representations increases according with the resolution, but the increasing rate is not substantial. Although the number of representations for sport videos is 2.5 times higher than for cartoon, we find here that there is in average 13.2 representations at 224p and 16.4 representations at 1080p. This makes a difference, but it is not a major trend.

We were curious to observe whether similar observations as for video types can be done when the population of users change. So we carried out results in a way similar to Fig. 6, but rather than changing users requests we vary users devices. We denote by x the portion of HDTV users and $0.8 - x$ portion of smartphone in the user population. We measure the distribution of requests for every resolution in Fig. 7.

We observe that the impact of the heterogeneity of users on the distribution of resolutions is less significant than for the popularity of videos. The evolution of the ratio of representations per resolutions follows the evolution of the distribution of devices in user population. We also observe a slight over-representation of higher resolutions indifferently from the ratio of HDTV users.

Guideline 3: How to decide bit-rates for representations in a given resolution?

The higher is the resolution, the wider should be the range of rates. Put emphasis on lower rates.

With our ILP, we obtain an optimal set that maximizes the average user satisfaction. However, system engineers are also interested in maintaining consistency in their systems, especially they commonly try to avoid that one representation is accessed by a lot of users although another representation serves only a few users. In Fig. 8, not only we get some precious insights about the range of bit-rates in the optimal representation sets, but also we analyze the "popularity" of each representation.

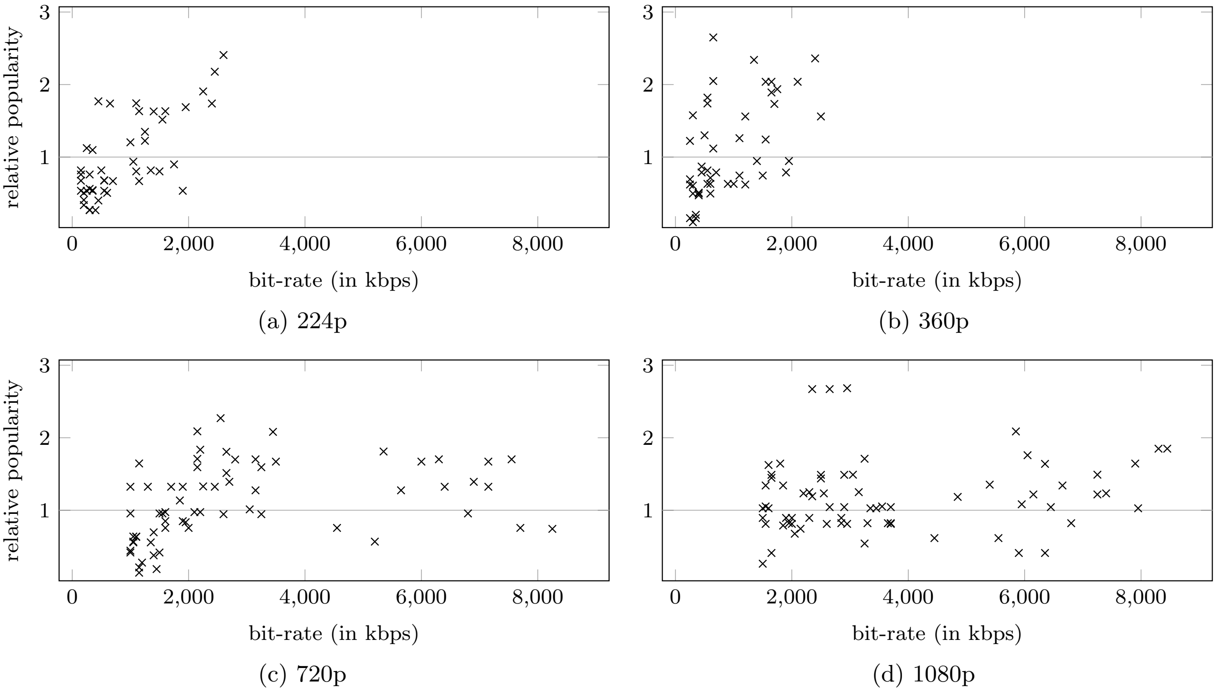
We define *relative popularity* as a value which indicates whether a representation is "over-assigned" (relative popularity greater than one) or "under-assigned" (relative popularity lesser than one). In particular, let L be a set of representations for a given video and a given resolution. Let l be one representation in L. Let n_L be the number of users who watch the said video at the said resolution. The average number of users per representation, which is hereafter noted n_L^{avg}, is given by $\frac{n_L}{|L|}$. Let n_l be the number of users assigned to representation $l \in L$. The relative popularity of the representation $l \in L$ is simply:

$$\frac{n_l}{n_L^{avg}}$$

In Fig. 8, we gather the results of five runs for the default settings. One mark shows that one representation has been created in one of the five runs for one of the videos. For each mark, we show the bit-rate and the relative popularity of the representation.

Our first observations is that the higher is the resolution, the broader is the range of bit-rates for the representations. Typically for 1080p resolution, the bit-rates ranges from 1,600kbps to more than 8,000 *kbps*. Such range cannot compare with 224p resolution where the range is from 200 *kbps* to 2,300 *kbps*.

Our second observation is that there exists a dense area of representation in the "south west" of every figure. It means both that there exist representations with the lowest possible rates in the optimal representation set, and that these representations are overall not accessed much. There are two reasons for such density in the low rates. First, the system has to ensure service for users with low network capacity. It is thus necessary to have a representation at



(a) 224p

(b) 360p

(c) 720p

(d) 1080p

Figure 8: relative popularity of representations (number of users requesting a given representation with respect to the average number of users requesting any representation in the resolution of the said representation) vs. bit-rate.

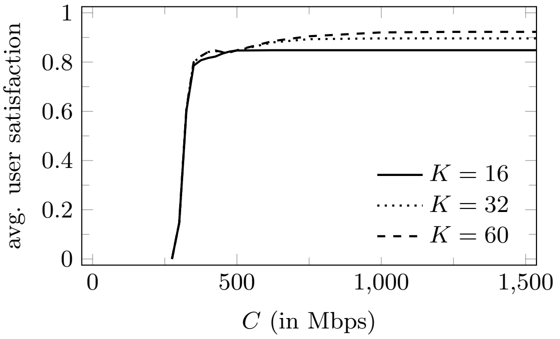


Figure 9: Average user satisfaction vs. CDN capacity C

one of the lowest possible rates. Second, the gains in terms of QoE are large in the low rates, so the encoding of a large number of representations at low rates is valuable because a small increase of network capacity at the client side can result in a significant QoE gain. In other words, the interval between two consecutive representations should be small at low rates and high for high rates.

Our third observation is that we do not see any major trouble with the distribution of user assignment on the representation set, although such metrics was not in our ILP. Note that it is trivial to add a constraint on the maximum number of users assigned to a representation if required. From our numerical analysis, such constraint is not necessary since no representation is assigned a population that is more than three times larger than the average expected population.

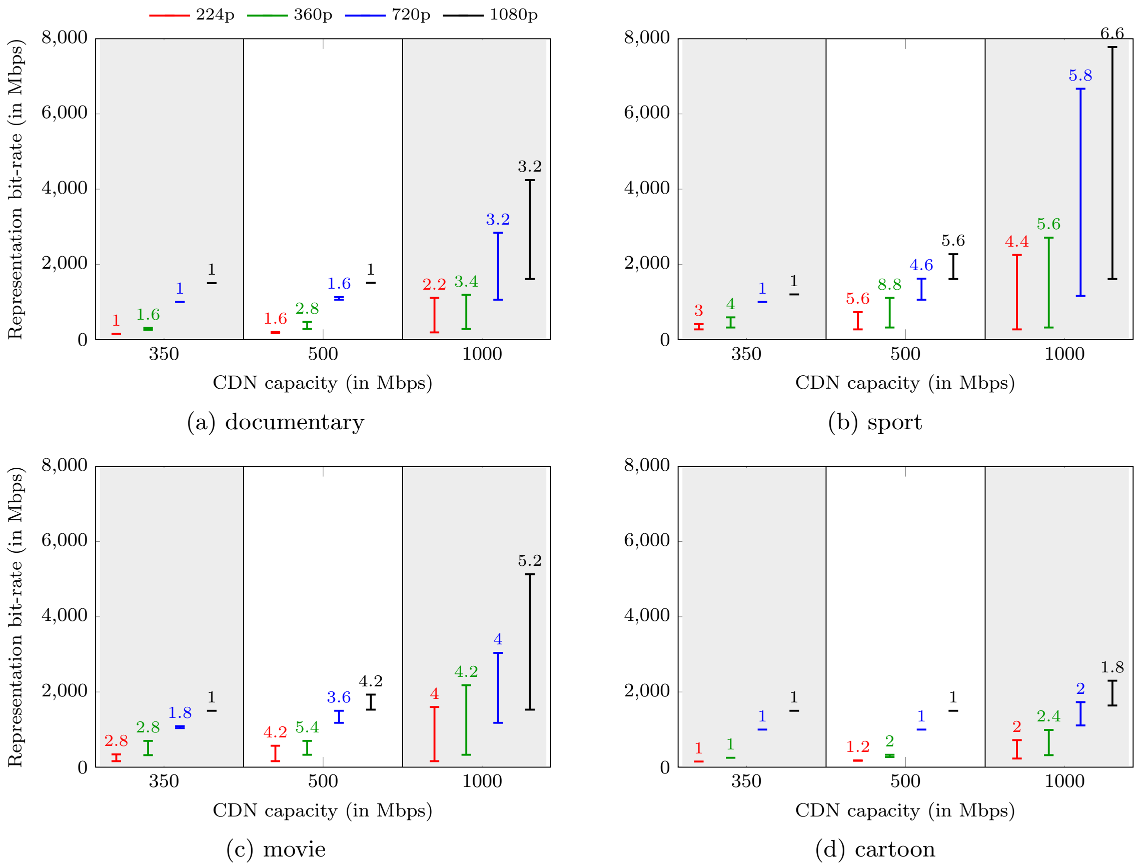
Guideline 4: How to save CDN bandwidth?

Reduce the range of rates for representations in a resolution. Reduce the number of representations at high resolutions.

One of the major concerns of content providers is to reduce the costs of delivering video streams. In the following, we study scenarios where the overall capacity C is arbitrarily restricted. The analysis of the optimal representation sets aims at identifying some ways to keep a reasonable average user satisfaction in under-provisioned configurations.

At first, we would like to observe how the average user satisfaction behaves when the CDN capacity gets low. In Fig. 9, we depict the average satisfaction of users as a function of various CDN capacities C. We can observe that i) there is a cliff effect, which means that there is a threshold value of C, around 375Mbps in our configuration, below which the QoE drops very quickly, and above which the QoE quickly reach the floor level; ii) the number of representations provides some gains in terms of user satisfaction only when the CDN capacity grows. When the delivery network is under-provisioned, there is no need to have a large number of representations.

We go more into the details of the guideline in Fig. 10, where we focus on three critical CDN capacities: $C = 350$ $Mbps$ (which is a capacity below the aforementioned threshold), $C = 500$ $Mbps$ (which is enough to deliver to users a service at good quality), and $C = 1{,}000$ $Mbps$ (which should enable the best possible user satisfaction). For each of these capacities, we represent the range of bit-rates in the optimal sets per resolution, with the minimum and the maxium bit-rates on average. The number above the bar is the average number of representations in a resolution. The maximum number of



(a) documentary

(b) sport

(c) movie

(d) cartoon

Figure 10: range of representations when CDN capacity is limited. Three different CDN capacities are given.. Bars are bounded by, at the bottom (respectively top), the average minimum (respectively maximum) value over 5 runs. The number over the bars indicate the average number of representations for the resolution.

representations K is 60.

For a low capacity ($C = 350$Mbps), not only there are very few representations (only 26 representations in average although it is possible to get 60), but also the ranges of bit-rates are very small. Simply put, an efficient set of representations in such underprovisioned context contains one representation per resolution, with the minimum possible bit-rate. A similar trend is visible for $C = 500$Mbps. The number of representations increases, but the ranges of bit-rates are still small. For the most impacting videos (here sport videos) the optimal set contains multiple representations such that their bit-rates that are very close to each other.

Please note that the scenario where $C = 1{,}000$Mbps confirms our three first guidelines. The ranges of bit-rates is larger for high resolutions, the number of representations depends on the videos and the number of representations is slightly higher for high representation.

7. CONCLUSION AND DISCUSSIONS

To the best of our knowledge, this paper is the first study on optimal encoding parameters for representation sets in adaptive streaming. More in details, first we defined an optimization problem for the selection of the representation set that maximizes the average satisfaction of users. We modeled this problem as an ILP. Using a generic solver, we were able to conduct a comprehensive numerical analysis, which allows us to measure the performances of representation sets based on recommendations, but also to identify some common patterns in the optimal sets. We derived guidelines for system engineers in charge of the encoding process in adaptive streaming delivery systems.

This paper opens a large number of perspectives.

- It reveals the gap between the practical importance of encoding video representations and the lack of theoretical foundations. Although the representation sets can severely impact the average QoE of users in adaptive streaming, this topic is still highly overlooked in the literature.
- Our optimization model depicts the complexity of today's video delivery systems. We gather information from various engineers and stakeholders to build a model that makes sense in both theoretical and practical contexts. The large number of parameters to take into account when addressing optimization problems in this area now challenges the scientific community. This paper is a first step toward a better understanding of the interaction and correlation between these parameters.

As part of our future works, automatic process for the setting of encoding parameters should be implemented at the ingest server. The combination of our guidelines and massive data retrieval from the delivery system should enable the implementation of an efficient ingest server. Also, dynamism of the system will be included into our study, developing strategies that select the best representations by leveraging forecasting algorithms.

8. REFERENCES

[1] T. Stockhammer, "Dynamic adaptive streaming over HTTP: standards and design principles," in *Proc. of ACM MMSys*, 2011.

[2] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the internet," *MultiMedia, IEEE*, vol. 18, no. 4, pp. 62–67, 2011.

[3] "Webrtc: Web browser with real-time communications." [Online]. Available: http://www.webrtc.org

[4] Apple, "Using HTTP live streaming," http://goo.gl/fJIwC.

[5] "Iis smooth streaming technical overview." [Online]. Available: http://www.microsoft.com/en-us/download/details.aspx?id=17678

[6] Netflix, "Encoding for streaming," http://is.gd/Ibo0LI.

[7] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation algorithm for adaptive streaming over HTTP," in *Proc. IEEE Packet Video Workshop*, 2012.

[8] V. Joseph and G. de Veciana, "NOVA: QoE-driven optimization of DASH-based video delivery in networks," *ArXiV*, vol. 1307.7210, 2013.

[9] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and adapt: Rate adaptation for HTTP video streaming at scale," *ArXiV*, vol. 1305.0510, 2013.

[10] R. K. P. Mok, X. Luo, E. W. W. Chan, and R. K. C. Chang, "QDASH: a QoE-aware DASH system," in *Proc. ACM MMSys*, 2012.

[11] S. Akhshabi, A. C. Begen, and C. Dovrolis, "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over HTTP," in *Proc. ACM MMSys*, 2011.

[12] S. Akhshabi, S. Narayanaswamy, A. C. Begen, and C. Dovrolis, "An experimental evaluation of rate-adaptive video players over {HTTP}," *Signal Processing: Image Communication*, vol. 27, no. 4, pp. 271 – 287, 2012.

[13] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with FESTIVE," in *Proc. ACM CoNEXT*, 2012.

[14] W. Zhang, Y. Wen, Z. Chen, and A. Khisti, "QoE-driven cache management for HTTP adaptive bit rate streaming over wireless networks," *IEEE Trans. on Multimedia*, 2013.

[15] E. Nygren, R. K. Sitaraman, and J. Sun, "The Akamai network: a platform for high-performance internet applications," *Op. Sys. Rev.*, vol. 44, no. 3, pp. 2–19, 2010.

[16] T. Hoff, "Gone fishin': Justin.tv's live video broadcasting architecture," High Scalability blog, Nov. 2012, http://is.gd/5ocNz2.

[17] IBM, "Ilog cplex optimization studio," http://is.gd/3GGOFp.

[18] Z. Ma, H. Hu, M. Xu, and Y. Wang, "Rate model for compressed video," *CoRR*, vol. abs/1206.2625, 2012.

[19] "VQM software." [Online]. Available: http://www.its.bldrdoc.gov/n3/video/vqmsoftware.htm

[20] "Xiph.org video test media." [Online]. Available: http://media.xiph.org/video/derf/

[21] Apple, "Best practices for creating and deploying HTTP live streaming media for the iphone and ipad," http://is.gd/LBOdpz.

[22] M. Grafl, C. Timmerer, H. Hellwagner, W. Cherif, D. Negru, and S. Battista, "Combined bitrate suggestions for multi-rate streaming of industry solutions," http://alicante.itec.aau.at/am1.html.

[23] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang, "Unreeling netflix: Understanding and improving multi-cdn movie delivery," in *IEEE INFOCOM*, 2012, pp. 1620–1628.

Resolutions				
	224p: 400x224	360p: 640x360	720p: 1280x720	1080p: 1920x1080
Videos				
Video Type	Documentary	Sport	Cartoon	Video
Video Name	Aspen, Snow Montain	TouchdownPass, RushFieldCuts	big buck bunny sintel trailer	old town cross

Table 4: Test sequences and resolutions considered for the quality evaluation.

	224p		**360p**		**720p**		**1080p**	
	$b_{vs}^{\min}$ (Kbps)	$b_{vs}^{\max}$ (Kbps)	$b_{vs}^{\min}$ (Kbps)	$b_{vs}^{\max}$ (Kbps)	$b_{vs}^{\min}$ (Kbps)	$b_{vs}^{\max}$ (Kbps)	$b_{vs}^{\min}$ (Kbps)	$b_{vs}^{\max}$ (Kbps)
Video	150	1757	200	2531	1000	8420	1500	7171
Sport	150	2350	200	2844	1000	8281	1500	7326
Documentary	150	2738	200	2764	1000	8545	1500	8455
Cartoon	150	2578	200	2592	1000	8291	1500	8421

Table 5: Minimum and maximum encoding rates.

Representation	1	2	3	4	5	6	7	8	9	10
Bit-rate (in kbps)	150	200	400	600	1,200	1,800	2,500	4,500	4,500	6,500
Resolutions	224p	224p	224p	360p	360p	720p	720p	720p	1080p	1080p

Table 6: Representation bit-rates recommended for Apple.

Representation	1	2	3	4	5	6	7	8	9	10
Bit-rate (in kbps)	350	400	900	1,250	1,400	2,100	3,000	3,450	5,000	6,000
Resolutions	224p	224p	224p	360p	720p	720p	720p	720p	1080p	1080p

Table 7: Representation bit-rates recommended for Microsoft.

Representation	1	2	3	4	5	6	7	8	9	10	11
Bit-rate (in kbps)	150	250	350	500	650	750	1,000	1,400	1,500	1,600	1,750
Resolutions	224p	224p	224p	224p	224p	224p	224p	224p	224p	224p	224p
Representation	12	13	14	15	16	17	18	19	20	21	22
Bit-rate (in kbps)	250	350	500	650	750	1,000	1,400	1,500	1,600	1,750	1,000
Resolutions	360p	360p	360p	360p	360p	360p	360p	360p	360p	720p	720p
Representation	23	24	25	26	27	28	29	30	31	32	33
Bit-rate (in kbps)	1,400	1,500	1,600	1,750	2,350	3,600	1,500	1,600	1,750	2,350	3,600
Resolutions	720p	720p	720p	720p	720p	720p	1080p	1080p	1080p	1080p	1080p

Table 8: Representation bit-rates recommended for Netflix.