Optimal Set of Video Representations in Adaptive Streaming

Laura Toni
EPFL, Switzerland
laura.toni@epfl.ch

Ramon Aparicio-Pardo
Telecom Bretagne, France
ramon.aparicio@telecom-bretagne.eu

Gwendal Simon
Telecom Bretagne, France
gwendal.simon@telecom-bretagne.eu

Alberto Blanc
Telecom Bretagne, France
alberto.blanc@telecom-bretagne.eu

Pascal Frossard
EPFL, Switzerland
pascal.frossard@epfl.ch

ABSTRACT

Adaptive streaming addresses the increasing and heterogenous demand of multimedia content over the Internet by offering several streams for each video. Each stream has a different resolution and bit rate, aimed at a specific set of users, e.g., TV, mobile phone. While most existing works on adaptive streaming deal with optimal playout-control strategies at the client side, in this paper we concentrate on the providers' side, showing how to improve user satisfaction by optimizing the encoding parameters. 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 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 play a crucial role in selecting proper encoding parameters to provide fairness among users and reduce network usage. 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 consuming video on the Internet has become more heterogeneous in terms of content requested, network connections, and devices. Adaptive streaming solutions aim to address this growing heterogeneity by offering users several versions of the video contents. Each version is encoded at a different bitrate and resolution so that any user can select the most suitable version depending on her streaming client and her network conditions.

MMSys '14, March 19 - 21 2014, Singapore.

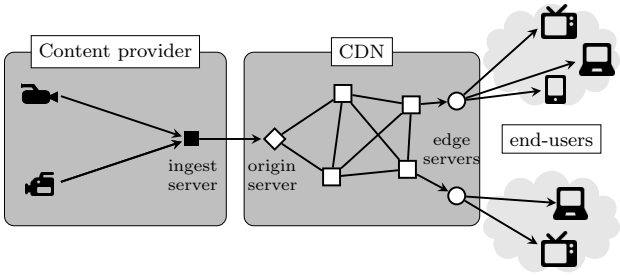


Figure 1: Live streaming: the delivery chain.

Figure 1 illustrates an instance of an adaptive streaming system for live video. The ingest server receives video data from cameras and prepares several different video *representations*, each one characterized by a different resolution and a bit rate. The ingest server sends the streams corresponding to each representation 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. At the other end, media clients send requests for video data that are available at the edge-servers. Several models have been recently proposed to standardize the adaptive streaming communication framework, like DASH [1, 2] and WebRTC [3]. Implementations of such systems differ in two ways: (i) the client adaptation strategy, and (ii) the selection of the different video representations. So far, the first problem has been at the center of the attention of the research community, while the second one has rarely been considered.

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 *recommendations* from system manufacturers, including Apple [4] and Microsoft [5]. Some content providers have also defined their own representation sets, for example Netflix [6]. 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, calling for a better selection of the encoding parameters.

Optimizing the encoding parameters for representation sets is an open problem, dealing with multiple correlated constraints, including the cost of delivering video streams using a CDN, the characteristics of end-users, and the type of video in input. For example, smaller sets (i.e., with few representations for each video) might satisfy only a fraction of the users, while larger ones could satisfy more users, but at a larger cost, in terms of increased storage costs for on-demand video, or larger encoding delays in the case of live-streaming. It is therefore important to study how the representation set should be designed, in order to strike the appropriate balance between user satisfaction and the cost of the system.

In this paper, we present and study an optimization problem to select the best encoding parameters of the representation set in an adaptive streaming system. We further demonstrate the need of making the selection of the representation set based on the video content, network, and clients characteristics. In more details, our contributions are as follows:

- We formulate an *optimization problem*, specifically an integer linear program (ILP), in order to find the best representation set, defined as the one that maximizes the average user satisfaction under network and system constraints. The satisfaction function of each client is 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. Results show that recommended sets perform well 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.

- In order to provide insights on how a system provider should select the encoding rates sets, we analyze the optimal representation sets in different scenarios. We consider several representative cases, by varying key parameters like user population (number of devices of each type: smart phones, tablets, etc.), network connection (capacity of each client connection and overall CDN capacity), and type of video (sport, documentary, movie, cartoon). By analyzing the solution of the ILP in each scenario, we notice recurrent patterns that lead us to formulate a few generic *guidelines*, which can be useful for content providers in the selection of the best representation set.

It should be stressed that even though we present detailed results only for several representative cases, the optimization problem we propose is a generic one and can address any real case.

The remainder of this paper is organized as follows. Related works on adaptive streaming are described in Section 2. Formalization of the optimization problem as an ILP is provided in Section 3. In Section 4, we detail the simulation settings. In Section 5, results are provided to study the system performance of optimal representation sets w.r.t the recommended one. In Section 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 Section 7.

2. RELATED WORKS

During the last decade, adaptive streaming has been an active research area, with most efforts aimed at developing server-controlled streaming solutions. Recently, a different approach, based on HTTP-adaptive streaming [1, 2], has gained popularity and attention. In this case the clients, and not the server, decide when and which segments to get. In other words the clients and not the server are in charge of making most of the decisions, including the selection of one of the representations available at the server.

Most papers dealing with these systems propose different ways of optimizing the representation selection for each user [7, 8] based on an estimate of the network dynamics [9] and on the control of the client buffer status. The objective is to maximize the Quality of Experience (QoE) of the users while avoiding unnecessary quality fluctuations. In [10], for example, the selection of the representation is optimized in such a way that large variations of rates in successive segments are avoided since large rate variations may lead to low QoE levels. Timing aspects in real-time applications have also been investigated in order to minimize the re-buffering phases [7]. Researchers have also investigated the performance of HTTP-based adaptive streaming in systems with a large number of users. The current HTTP-adaptive streaming systems have limitations when a large number of clients share the same network, as illustrated by the experimental results presented in [11–13]. For example, users cannot reach simultaneously fairness and efficiency in a scenario where many clients share the same bottleneck link.

Most existing works, however, do not address the problem of deciding which representations should be available at the server. Usually the available representations are an input parameter. These are often selected based on vendor recommendations, as in the case of Apple [4] and Microsoft [5]. It happens that a content provider builds its own representation set with regards to the supposed characteristics of its content and its clients, as in the case of Netflix [6]. 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.

Encoding rate optimization has been investigated very recently in [14], for on-demand videos in a storage-limited scenario. Rates 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. All the scenarios presented in [14] consider a homogeneous user population and this is a key assumption exploited in the solution of the optimization problem. In this paper, instead, we

explicitly model different types of users, in terms of access link capacity and device used (smartphone, tablet, laptop, television). We also take into account different types of video as this has a non-negligible impact on the perceived QoE. Finally, we do not restrict our study to VoD and storage constraints. Instead, our optimization problem applies indifferently to live streaming system.

3. PROBLEM FORMULATION

We now provide the problem formulation for selecting the best representation set by taking into account the network capacity, users' requests, and video content information (i.e., different types of video). For the sake of simplicity we consider for every instance of the problem the user population and the network capacity as known constants. Even though these quantities are not necessarily constant in real systems, we argue that considering them as constants is a reasonable first step in addressing this complex problem as it allows to better assess the influence of the other parameters on the optimal solution. This way we can also more easily, and more fairly, compare different solutions, including those based on existing recommendations.

In the remainder of this section, we first introduce the notation used to formalize the problem, including the constraints. Then we present the ILP model used to solve the optimization problem.

3.1 Definitions

Let $\mathcal{V}$ be the set of videos. Each video $v \in \mathcal{V}$ can be encoded in different representations, each one 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. More precisely, we use r as an index in the set of rates and we use b_r for the actual value (in bits per second) of the encoding rate.

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 bits per second. 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 is different for every resolution. For example, for a user watching a video v at resolution s, $f_{vs}(r) = 1$ if $b_r = b_{vs}^{\max}$, but the same rate might lead to a satisfaction lesser than 1 for the same video content but displayed at a higher resolution. For sake of clarity in the notation, in the following we denote the satisfaction level by f_{vrs} rather than $f_{vs}(r)$.

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 derive directly from real challenges identified by service providers. We highlight three constraints:

Name	Description
$f_{vrs} \in \mathcal{R}^+$	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
$C \in \mathcal{R}^+$	Total network capacity in *kbps*
$K \in \mathcal{R}^+$	Total number of representations used, i.e., triples $(v,\ r,\ s)$, used by the CDN
$P \in [0, 1]$	Fraction of users that must be served

Table 1: Notation adopted in the ILP formulation.

- **The global CDN capacity** available to successfully deliver all the video streams. We denote this overall capacity by C (measured in bits per second). In general, video service providers reserve an overall budget (in $) for video delivery and use it to buy a delivery service from a CDN provider. In today's CDN, the price depends on the sum of all the rates of all the video streams originating at the content provider [15]. Thus, the manager of a video service provider is interested in maintaining the total delivery bandwidth below a given value, here denoted by C.

- **The total number of representations**, denoted by K, i.e., the total number of triples $(v,\ r,\ s)$ provided to ingest servers. 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 to be provisioned to encode raw video. 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 [16].

- **The fraction of users that must be served**. Ideally, the service provider would like to serve all the users. But in certain cases, especially when the number of representations K is small, the optimal solution can exclude all the users that have a small c_u (capacity of the link connecting user u to the Internet/CDN). In this case, service provider could prefer serving a certain fraction of the users, even if this would lead to a suboptimal solution, in the interest of fairness. We introduce an additional constraint to address this problem, denoting by P the fraction of users that must be served. As there exist different definitions of fairness, this constraint can be modified according to the definition.

Table 1 summarizes the notation used in this paper.

3.2 ILP Model

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

Integer Linear Programming formulation

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

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

$$\beta_{vrs} \le \sum_{u\in\mathcal{U}} \alpha_{uvrs}, \qquad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \quad (1c)$$

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

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

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

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

$$\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} b_r\cdot\alpha_{uvrs} \le C, \quad (1h)$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \beta_{vrs} \le K, \quad (1i)$$

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

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

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

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

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

With these definitions, 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 is encoded only 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) respectively limit the user link capacity and the overall network (CDN) capacity. The constraint (1i) sets the maximal number of representations. The constraint (1f) ensures that, if user u is served, she receives the requested video stream v_u with the correct resolution s_u. And, finally, the constraint (1j) ensures that the fraction of users served is larger than P.

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 (at the client) and encoding (at the server) CPU and GPU cycles, the size of the client buffer, or even more specific parameters like the library codec that should be installed at the client side.

Nevertheless, what we observe is that the constraints imposed by rates and resolutions are somehow generic in the sense that the structure of the inequality can be reused to express other constraining parameters. Typically, a constraint on the required decoding CPU would have exactly the same structure 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.

Video Type	Video Name
Documentary	Aspen, Snow Mountain
Sport	Touchdown Pass, Rush Field Cuts
Cartoon	Big Buck Bunny, Sintel Trailer
Video	Old Town Cross

Table 3: Test videos and corresponding type.

Name	Width x Height
224p	400x224
360p	640x360
720p	1280x720
1080p	1920x1080

Table 4: Resolutions used.

4. NUMERICAL ANALYSIS SETTINGS

In the following section, we use the ILP introduced in Section 3.2 as tool to perform a comprehensive numerical analysis of the optimal selection of the encoding parameters for representation sets. We have used the generic solver IBM ILOG CPLEX [17] to solve different instances of the ILP, obtaining the 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 analysis. First, we present how the *user satisfaction* is evaluated. Second, we explain how the *user population* is synthetically generated. Finally, we describe the default settings that we have used. It should be stressed that we have chosen some representative scenarios in order to carry out our analysis. These scenarios are not meant to be an exhaustive list covering all possible cases, nor are they meant to represent the most common cases. Rather they are meant to illustrate how the optimal solution changes in several realistic cases.

4.1 User Satisfaction

We characterize each video at a given resolution by one *satisfaction function*, expressing the QoE as a function of both the rate and the resolution. Several works have investigated how to model this behavior and a uniformly accepted model still has to be accepted [18]. In our case, we model the satisfaction function as an Video Quality Metric (VQM) score [19], which is a full-reference metric

	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 2: Minimum and maximum encoding rates.

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

Table 5: Parameters of the QoE model.

that has higher correlation with human perception than other MSE-based metrics.

We evaluated the VQM score for four different test sequences from [20] 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 3 and Table 4 respectively. 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 f_{vrs} receiving a video v encoded at rate r and resolution s is modeled as follows

$$f_{vrs} = m_{vs} * b_r^{n_{vs}} + o_{vs}. \quad (2)$$

Table 5 gives the parameters m_{vs}, n_{vs}, and o_{vs} used in the fitting for each video v and resolution s. Recall that b_r is the nominal value in $Kbps$ of rate r. Satisfaction curves evaluated from Eq. 2 are plotted as continuous lines in Fig. 2. Note that, even if many parameter (delay variations, network capacity fluctuations, etc.) can potentially affect the satisfaction level, we assume that their influence is negligible compared to the encoding rate.

Network Type	Minimum Bandwidth (in Mbps)	Maximum Bandwidth (in Mbps)	Attachment Probability
Wifi	0.15	0.8	0.3
3G	0.4	4	0.2
ADSL-slow	0.3	3	0.1
ADSL-fast	0.7	10	0.3
FTTH	1.5	25	0.1

Table 6: Different network types and corresponding parameters.

4.2 User Population

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

v_u : Users are randomly assigned to one of the four video types given in Table 3. Each video type has the same probability (1 out of 4) of being selected.

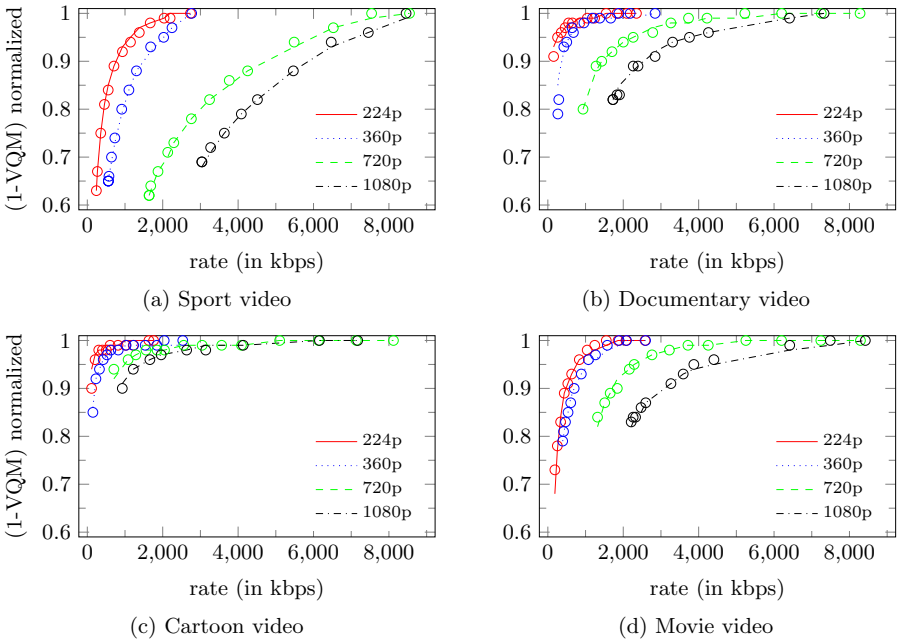
s_u : Users are randomly assigned to one of four device types: smartphone, tablet, laptop and high definition television (HDTV). Each device is associated to a resolution: 224p, 360p, 720p and 1080p for smartphone, tablet, laptop and HDTV respectively. Again, each device type has the same probability (1 out of 4).

c_u : Users are randomly assigned to one of the four network types in Table 6, using the probability given in the last column. Once a user is associated to a given type of network, c_u is selected as a uniformly distributed random value between minimum and maximum capacity (second and third column in Table 6).

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 otherwise mentioned. The video catalog $\mathcal{V}$ and spatial resolution set $\mathcal{S}$ correspond to the video sequences and resolutions indicated in Table 3 and 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 2.

The satisfaction coefficients f_{vrs} are fixed for each triple (v, r, s) according to the extrapolated satisfaction curves plotted in Fig. 2. The global network capacity (C) is 5 000 kbps, the maximum number of representations (K) is 60 and the fraction of users that must be served (P) is 0.95.



(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 represent the model.

Representation	Apple		Microsoft	
	Rate (kpbs)	Resolution	Rate (kpbs)	Resolution
1	150	224p	350	224p
2	200	224p	400	224p
3	400	224p	900	224p
4	600	360p	1,250	360p
5	1,200	360p	1,400	720p
6	1,800	720p	2,100	720p
7	2,500	720p	3,000	720p
8	4,500	720p	3,450	720p
9	4,500	1080p	5,000	1080p
10	6,500	1080p	6,000	1080p

Table 8: Representation and corresponding bit rates recommended by Apple and Microsoft.

For each problem instance we have generated five instances, each one with a population of 500 users, whose characteristics are selected as described in Section 4.2. In the following, when we provide average metrics (e.g., average user satisfaction, average number of representations), the average is taken over these five instances.

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

5. HOW FAR FROM THE OPTIMAL ARE THE RECOMMENDED SETS?

As already mentioned in previous sections, today's system engineers commonly select encoding parameters for the representations following given recommendations, which are not optimized based on content or contextual information but 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 the recommended sets?* With the 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 [4, 21] for HTTP Live Streaming (HLS), Microsoft [22] for Smooth Streaming (see Table 8), and Netflix [6, 23] (see Table 7).

In Fig. 3, we show the average user satisfaction as a function of the number of representations K in the optimal solution. Note that K is the total number of representations for all four video channels. From Tables 7-8, $K = 40$ for Apple and Microsoft recommendations while $K = 132$ for Netflix recommendations. The other parameters are selected as discussed in Section 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. To accomodate the settings of the different recommendations, we slightly modified the minimum rate for each class of users so that every user has a network capacity greater than the rate of the lowest representation. In our opinion, the differences between the recommendations come from the fact that the vendors target different populations. Apple recommendations

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

Table 7: Representation and corresponding bit rates recommended by Netflix.

typically accommodate smartphones and tablets while Microsoft target more specifically laptops and home computers.

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) with 40 representations can be obtained with 21 (respectively 22) representations in the optimal set, so roughly half the number of representations. It is worth to recall that the more representations in the set, the more complex and costly the encoding and delivery system are.

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) can be obtained with 34 representations in the optimal set. This corresponds to a reduction of 98 representations, i.e., a reduction of about 70% in terms of representation set cardinality.

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 performance in different configurations, Fig. 4 shows the average user satisfaction as a function of K when two parameters differ from the parameters given in Section 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 of 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 homogeneous 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.

6. GUIDELINES

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

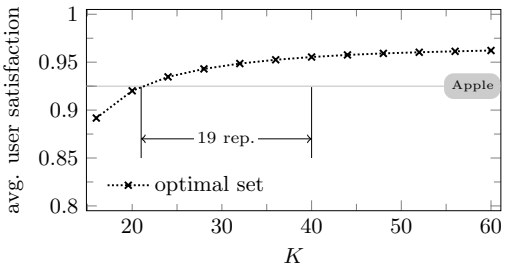
Guideline 1: How many representations per video?

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

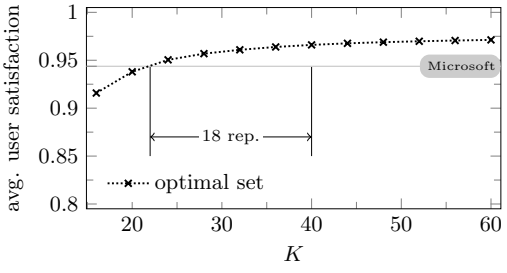
A weakness of the recommended representation sets is that the number of representations is the same for any video. In Fig. 5, we show the average number of representations dedicated to any video type as a function of the video resolution for the optimal representation sets.

We observe that some videos clearly require more representations than others: about 21 on average for sport videos while only about $8-9$ representations on 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 cartoons. 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 the sport video than for the cartoon.

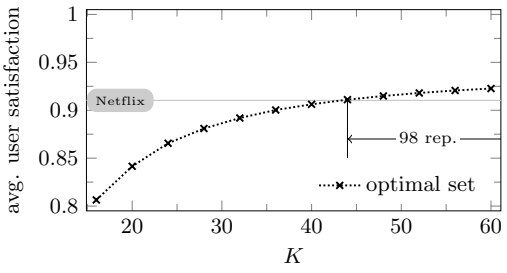
To confirm that these results are not biased by our default configuration, we changed the popularity of the videos in the catalog. Four video types are still considered, i.e., documentary, movie, sport and cartoon, but only 10% of users watch the documentary, another 10% of them watch the movie, and the remaining is shared between cartoon and sport videos. More precisely, x is the ratio of users watching the sport video, and $0.8 - x$ is the ratio of users watching the cartoon. In Fig. 6, the parameter x ranges from 0 (no sport videos) to 0.8 (no cartoon videos).



(a) Apple set (40 rep.)



(b) Microsoft set (40 rep.)



(c) Netflix set (132 rep.)

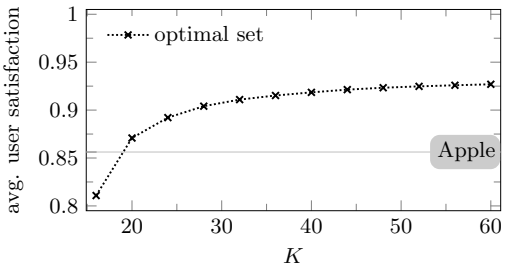
Figure 3: Average user satisfaction: recommended sets vs. optimal sets with different number of representations.

We measure the distribution of 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 video.

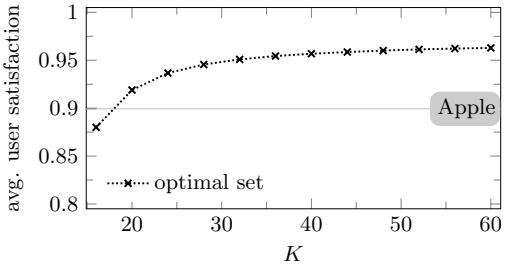
For example, when both sports and cartoon are requested by 40% of the population, representations are unequally distributed among videos (47% for sport while 18% for the cartoon). This confirms our previous observation. Cartoon videos (respectively sport videos) are under-(respectively over) represented indifferently from the popularity. Even when sport videos are watched by only 10% of users, one third of representations are used by the sport video. 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.



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



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

Figure 4: Average satisfaction of users for the representation sets recommended by Apple (40 rep.) in different contexts.

For a first analysis of the representation distribution per resolution, we can refer again to Fig. 5. For a given video, the number of representations increases with the resolution, but the increase 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 on average 13.2 representations at 224p and 16.4 representations at 1080p. Despite the difference, this is not a major trend.

We wondered whether conclusions similar to those for the type of video content would hold when the distribution of users' devices changed. To find out, similarly to what we did for Fig. 6, we changed users' devices. We denote by y the portion of HDTV users and $0.8 - y$ portion of smartphone users. Fig. 7 shows the distribution of requests for every resolution.

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 resolution 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 the ILP, we obtain an optimal set that maximizes the average user satisfaction. However, system engineers are also interested in maintaining consistency in their systems, trying to avoid for example that one representation is accessed by a lot of users although

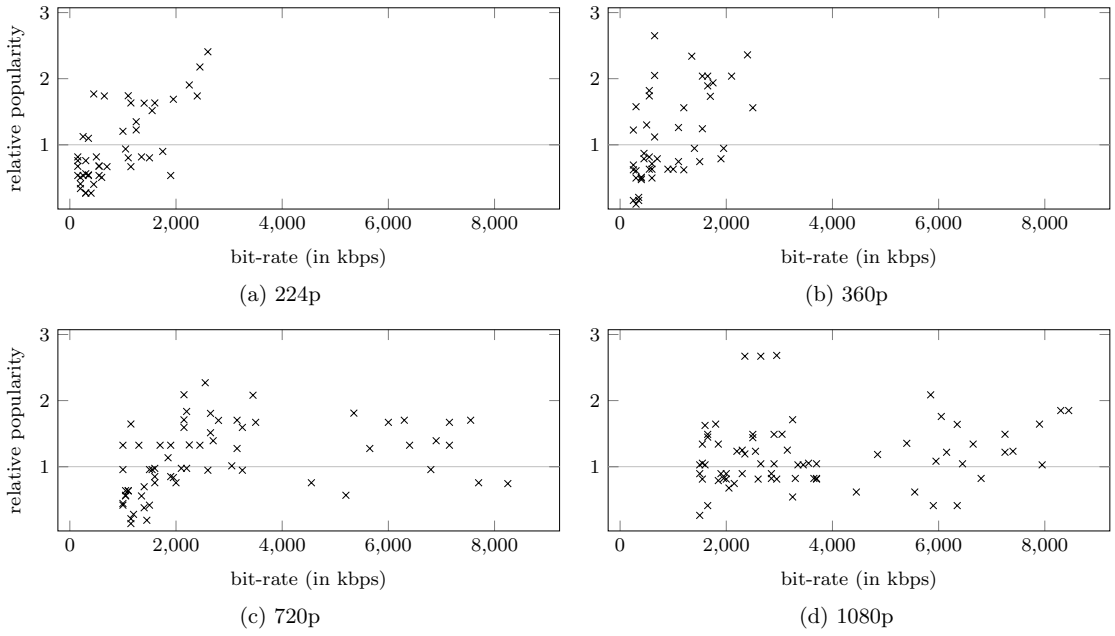


Figure 8: Relative popularity of representations (number of users requesting a given representation with respect to the average number of users requesting any representation in the resolution of said representation) vs. bit rate.

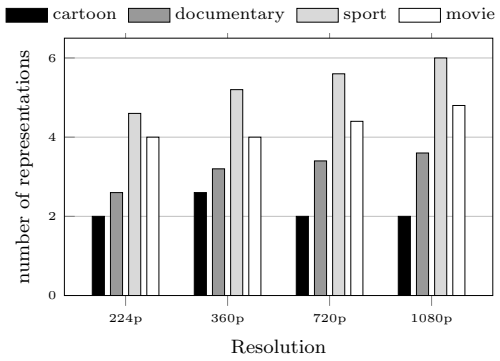


Figure 5: Average number of representations per resolution, for each type of videos.

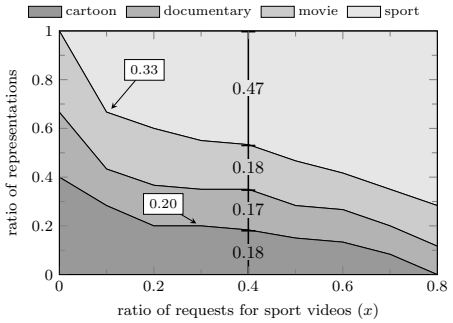


Figure 6: Distribution of representations per video.

another representation serves only a few users. In Fig. 8, not only we get some valuable insights about the range of bit rates in the optimal representation sets, but we can also analyze the "popularity" of each representation.

We define the *relative popularity* as a value that 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 said video at 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 observation is that the higher the resolution, the broader the range of bit rates for the representations. Typically for the 1080p resolution, the bit rates ranges from 1,600 *kbps* to more than 8,000 *kbps*. Such range is much larger than the one for the 224p resolution, from 200 *kbps* to 2,300 *kbps*.

Our second observation is that there exists a dense area of representations in the "south west" of every figure. It

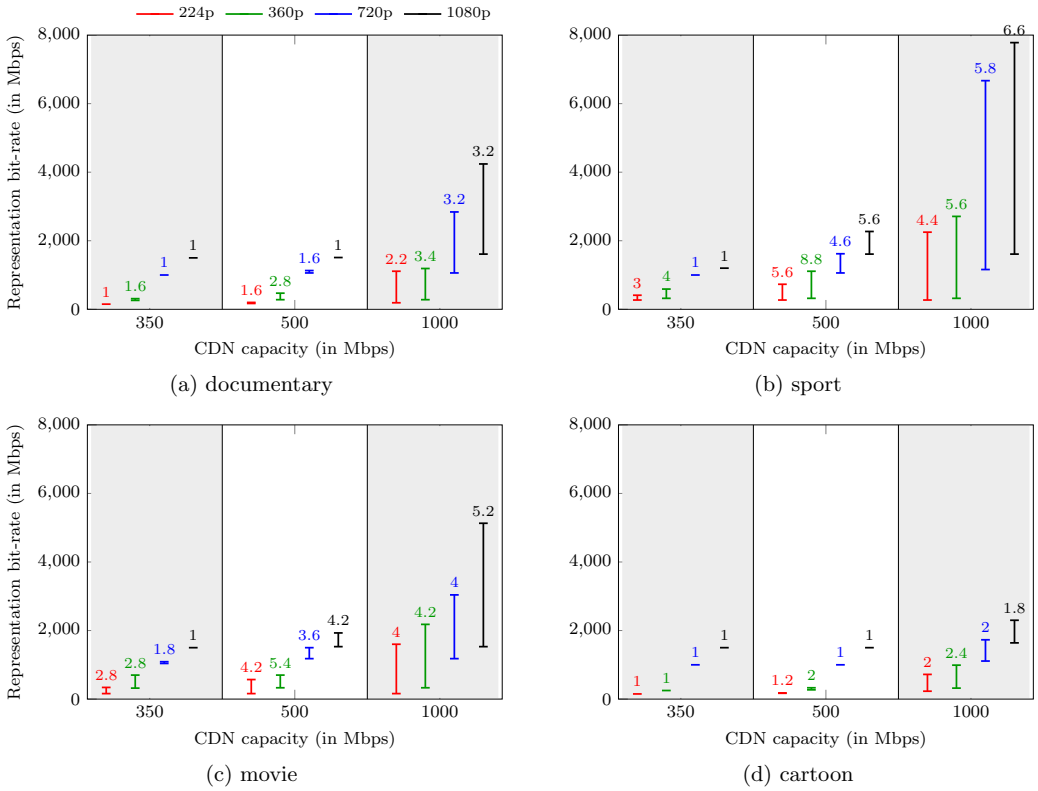


Figure 10: Range of representations when CDN capacity is limited. Three different CDN capacities are given. Bars are bounded, at the bottom (respectively top), by the average minimum (respectively maximum) value over 5 runs. The number over the bars indicates the average number of representations for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

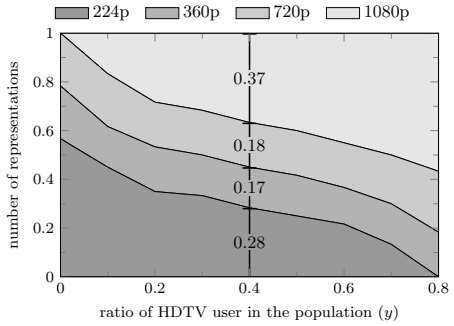


Figure 7: Distribution of representations per resolution.

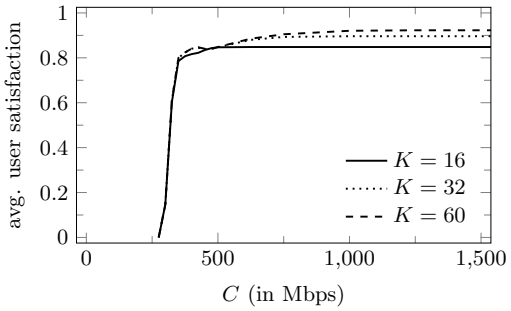


Figure 9: Average user satisfaction vs. CDN capacity C.

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 connected by low capacity links (i.e., small values of c_u). It is thus necessary to have a representation at one of the lowest possible rates. Second, the gains in terms of QoE are usually large in the low rates, so the encoding of a large number of representations at low rates is valuable because a small increase of the link 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 the number of users assigned to each representation, even though we did not constrained it in the ILP formulation. If required, it would actually be trivial to add a constraint on the maximum

number of users assigned to a representation. But our numerical analysis shows that such constraint is not necessary since no representation is assigned to 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 remainder of this section, we study scenarios where the overall capacity C is arbitrarily restricted. The analysis of the optimal representation sets aims at identifying ways to keep a reasonable average user satisfaction in under-provisioned configurations.

First, we would like to investigate how the average user satisfaction behaves when the CDN capacity decreases significantly. In Fig. 9, we depict the average satisfaction of users as a function of different CDN capacities C. We can observe that *i*) there is a cliff effect, which means that there is a threshold value of C, around 375 $Mbps$ in our configuration, below which the QoE drops very quickly, and above which the QoE quickly reaches 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 this 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 a good quality service to the users), 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 maximum bit rates on average. The number above the bar is the average number of representations per resolution and per video. The maximum number of representations K is 60 to be distributed among all videos and resolutions.

For a low capacity ($C = 350$ $Mbps$), there are very few representations - only 26 representations on average (evaluated by summing the number above the bar for all resolutions and videos in the 350 subplots) despite the maximum being 60. The ranges of bit rates are very small as well. Simply put, an efficient set of representations in such an under provisioned 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 high impact videos (here sport videos) the optimal set contains multiple representations such that their bit rates 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 higher resolutions.

7. CONCLUSION AND DISCUSSION

To the best of our knowledge, this paper is the first study on optimal encoding parameters for representation sets in adaptive streaming. We have 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, whose optimal solution can be computed by a generic solver. We were able to conduct a comprehensive numerical analysis, which allowed us to measure the performance of representation sets based on recommendations, but also to identify some common patterns in the optimal sets. We have also 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 existing recommendations and solutions that maximize the average user satisfaction. Although the representation sets can severely impact the average QoE of users in adaptive streaming, this topic is still overlooked in the literature.
- Our optimization model captures 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, we plan to improve the performances of ingest servers. Our work opens persepectives toward the design of processes that automatically set encoding parameters at the ingest server. Furthermore, the combination of our guidelines and the analysis of statistical data from the delivery system should enable the implementation of more efficient ingest server. Also, our study is based on a snapshot of the system although adaptive streaming systems have been designed to cope with dynamic environment. One possible approach to enhance the setting of representation set is to leverage forecasting algorithms so that the most probable changes in the environment (including video popularity and network changes) are anticipated.

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 considering impacts of spatial, temporal and amplitude resolutions and its applications for video coding and adaptation," *ArXiv*, 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.