Split the cash from cache-friendly recommendations

Tsigkari, Dimitra ; Iosifidis, George; Spyropoulos, Thrasyvoulos

DOI
10.1109/GLOBECOM46510.2021.9685088

Publication date
2021

Document Version
Final published version

Published in
2021 IEEE Global Communications Conference (GLOBECOM)

Split the cash from cache-friendly recommendations

Dimitra Tsigkari*, George Iosifidis†, and Thrasyvoulos Spyropoulos*
* EURECOM, Biot, France. Email: {dimitra.tsigkari, thrasyvoulos.spyropoulos}@eurecom.fr
† Delft University of Technology, Netherlands. Email: g.iosifidis@tudelft.nl

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685088

***Abstract*—Recommender systems have been established as a key component of video streaming services, shaping up to 80% of content requests. Hence, recommendations are employed by the Content Providers (CPs) of these services to increase the viewing time and their revenues. Furthermore, it has been recently suggested that recommendations could be a means to reduce the operational costs of the Content Delivery Networks (CDNs) when they are related to already cached items, *i.e.*, when they are cache-friendly. Clearly, these conflicting objectives, *i.e.,* increasing revenue for the CP and reducing costs for the CDN, can create tensions between the two entities, and hence, prevent the full utilization of recommendations. In this work, we propose a model for capturing these tradeoffs, and an economic mechanism, based on the Nash bargaining solution, for reconciling the potentially conflicting objectives of the CP and the CDN. Our scheme enables the CP and CDN to jointly design the recommendations in a way that balances the revenue gains and cost savings, ensuring a fair and Pareto optimal split of the accrued benefits for both entities. Our numerical experiments in realistic scenarios show that the proposed scheme leads to important financial gains of up to 30%.**

I. Introduction

A. Motivation and Related Work

Recommendation systems are inherent in the Over-the-top (OTT) streaming services such as Netflix, YouTube, and Disney+, and are affecting a large portion of the issued content requests. In Netflix, for example, 80% of the requests derive from recommendations [1]. Therefore, it is not surprising that OTT Content Providers (CPs) comprehend the business value of these systems and invest research and financial resources to improve their accuracy. By proposing contents of high potential interest to users, the CPs can increase the users' engagement to the service, thereby boosting their revenues [1], [2]. Furthermore, it has been reported that recommendations can generate annual savings worth 1 billion US dollars ($) by reducing user churn [1].

Recently, it became clear that recommendations can also be used to reduce the operating expenditures of the Content Delivery Network (CDN) the CP employs for reaching its users. Clearly, recommendations that favor cached items can lead to increased cache hits and reduced delivery costs. In fact, requests that entail retrieving contents from a cache deep in the network or even the CP's origin server, can significantly push up the operational costs of the CDN [3]. This is a promising area of research with recent works proposing cache-friendly or network-friendly recommendation policies (*e.g.*, [4], [5]), or even policies that jointly optimize caching and recommendation decisions (*e.g.*, [6], [7]). Such approaches are also beneficial from the point of view of users who can experience better Quality of Service (QoS) [7], *e.g.*, higher bitrate or lower start-up delays.

An aspect that has yet to be explored is the tension between CPs and CDNs when it comes to the use of recommendations. Clearly, the CDN would prefer cache-friendly recommendations, however, these may be oblivious to users' interests, and hence, affect the CP's revenues. On the other hand, when jointly decided by the CP and the CDN, as suggested in related work, cache-friendly recommendations create a potential revenue surplus in terms of increased viewing income (for the CP) or reduced content delivery costs (for the CDN). This surplus could create tension among the potential recipients on the way it will be split between them. Alleviating this issue will enable more efficient recommendations that are possible through collaboration. In this work, we show how these entities can collaborate to design a cooperative recommendation policy that is mutually beneficial.

B. Our Approach and Contributions

We propose a novel collaboration scheme between an OTT streaming service CP and a CDN, on the grounds of recommendations, that leads to a win-win situation. Given the cache allocation made by the CDN in its network, the proposed scheme allows the CP and CDN to decide jointly on the recommendations, and favor cached contents (or contents whose retrieval cost for the CDN is low). In exchange, the CDN offers a price reduction to the CP on the delivery fees[1]. In order to derive these recommendations that will be called *cooperative recommendations*, and to provide incentives to these two entities to exchange information between them, we rely on cooperative game theory. In particular, we employ the celebrated Nash Bargaining Solution (NBS) [8] that achieves a Pareto optimal solution with proportionally fair gains. Our model takes into account the independent gains each entity could have achieved outside of any collaboration. That is, the gains when the CP derives the recommendations based on its own economic criteria (*standard recommendations*), and when the CDN cannot reduce its operational/retrieval costs without knowing how recommendations shape requests. The proposed collaboration (depicted in Fig. 1) could be managed either by the two entities or by a third party company.

Our contributions are the following:

- We propose a novel approach for using recommendations in OTT services using a hybrid revenue - cost criterion.

This work has been supported by the French National Research Agency under the "5C-for-5G" JCJC project with ref. number ANR-17-CE25-0001.

[1]Another interesting direction would be to model a collaboration where the two entities decide on both the recommendations and caching allocation. However, such an approach would require a better (and non-trivial) coordination among the two entities that we plan to address in future work.

978-1-7281-8104-2/21/$31.00 ©2021 IEEE

- We introduce a bargaining framework for modeling and addressing the tradeoffs between revenues (for CPs) and costs (for CDNs). We formulate the optimization problem which allows us to derive the cooperative recommendations in a way that the allocation of economic gains between the two entities is Pareto optimal and fair.
- We show that the formulated optimization problem can be solved efficiently and discuss the properties of the solution.
- We conduct numerical evaluations in realistic scenarios using a real dataset. Our results show that a significant increase in revenue, of up to 30%, can be achieved. We also present a sensitivity analysis that reveals the role of different input parameters.

To the best of our knowledge, this is the first work studying how to leverage recommendations to achieve a fair and efficient allocation of expected viewing gains and delivery cost savings among CPs and CDNs.

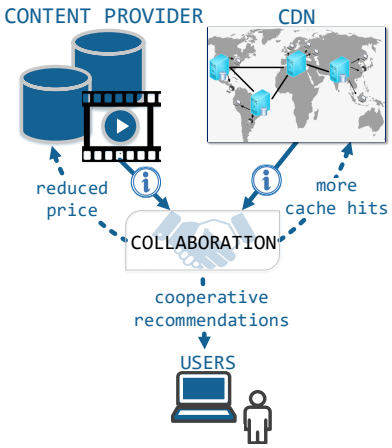

Fig. 1. The collaboration proposed in this work. When the CP and the CDN collaborate, they share information directly or through a third party. The users receive "cooperative" recommendations that benefit both parties: reduced price on the delivery fees for the CP, and more cache hits that lead to lower retrieval costs for the CDN.

II. Problem Setup

A. Business models, Revenue, and Costs

Before modeling the CP-CDN collaboration, it is important to discuss the different sources of revenue and costs for the two main actors[2] that are depicted in Fig. 2.

1) Content Provider (CP) of an OTT streaming service: The sources of revenue for the CP depend on the business model on which it is built. In fact, there are 4 main revenue models for streaming platforms [9]:

- ad-based (*e.g.,* YouTube (non-Premium))
- subscription-based (*e.g.,* Netflix)
- transaction-based or pay-per-view (*e.g.,* iTunes)
- hybrid model (*e.g.,* Amazon Prime Video).

A CP (without an in-house CDN) subscribes to a CDN provider through a Service Level Agreement (SLA) or delivery contract for the delivery of the contents to the users.

[2]We focus here on the revenues and costs that are relevant in our model. Obviously, additional costs are incurred by both the CP and CDN, *e.g.*, fixed business costs.

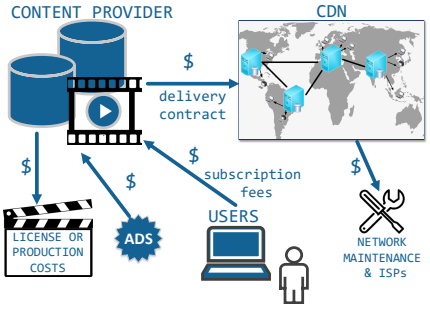

Fig. 2. Sources of revenue and cost for the CP and the CDN as described in Sec. II. The delivery contract is the foundation of their economic relationship.

Apart from this cost, the CP pays for the purchase of contents (through licensing or production).

2) Content Delivery Network provider (CDN): The main source of revenue for a CDN from the OTT industry is the delivery contract with the CPs. The related costs of the CDN are transit and maintenance costs. In particular, the CDN pays transit fees to transit networks or Internet Service Providers (ISPs) to retrieve the content from the origin servers of the CPs and make it available to the users. Moreover, the CDN needs to pay costs that are related to storage capacity, hardware, estate, energy, etc [3].

B. Recommendations, Content Requests, and Utility Functions

In this work, we will extend the aforementioned economic model to allow additional cooperation between the CP and CDN on the basis of the recommendations the former offers to its users. We first elaborate on the incomes and costs that are the result of the recommendations.

The CP owns a content catalogue $\mathcal{K}$ that is accessible to a set $\mathcal{U}$ of users through the CP's OTT service. A recommendation system is implemented within the OTT service (by the CP). A list of N_u recommended contents appears to every user $u \in \mathcal{U}$. As is common in related works (*e.g.*, [5], [6], [7]), the user makes a content request according to the following model:

- with probability α_u, the user requests a recommended content. Each of the N_u recommended items could be chosen with equal probability;
- with probability $(1-\alpha_u)$, the user ignores the recommendations and requests a content $i \in \mathcal{K}$ of the catalogue with probability p_{ui}.

The quantity α_u represents the percentage of time a user u tends to follow the recommendations. For example, it is estimated, on average, that $a_u = 0.8$ on Netflix [1]. The assumption that the user u may click each recommended content with equal probability $1/N_u$ can hold in scenarios where she cannot evaluate the utilities of the recommended items before requesting them. Parameter p_{ui} captures the probability of user u requesting the content i outside of recommendations (*e.g.*, through the search bar), and could relate to the aggregate interest in this content by users.

We denote by $r_{ui} \in [0, 1]$ the (predicted) utility or relevance of a content i to the user u. These values could be, for example, the result of collaborative filtering algorithms based on user data or other state-of-the-art algorithms employed by today's

recommender systems [1]. When a user u requests a content i, this content is associated with an expected revenue R_{ui} that is estimated by the CP as a result of its revenue model and the associated costs related to production or to licences. This expected revenue depends on the content utilities r_{ui} in a non-trivial way. For example, when the relevance of recommended contents is not high to the user, this can lead to user abandonment [1]. For this reason, we write

$$R_{ui} = f_{ui}(r_{ui}), \tag{1}$$

where f_{ui} is an increasing function of r_{ui} and describes the impact of the user's (predicted) interest in a content on the CP's revenues[3]. For example, f_{ui} could be related to the probability of user abandonment as a function of r_{ui}. In fact, the dependency of R_{ui} to r_{ui} inhibits the CP from recommending contents based solely on the expected revenue without taking into account the user's interests.

The delivery of a requested content is done by the CDN that charges the CP on a basis of the amount of transferred data (as is the case in today's CDNs [10]). We assume that the CP has to pay λ currency units per Gb requested[4]. The size of content $i \in \mathcal{K}$ in Gbs is denoted by $\sigma_i > 0$.

Therefore, if content i is recommended to user u, the expected net revenue for the CP will be $\frac{\alpha_u}{N_u}[R_{ui} - \sigma_i\lambda]$. We let $y^0_{ui} \in \{0,1\}$ denote the input variable for the content i being recommended to user u ($y^0_{ui} = 1$) or not ($y^0_{ui} = 0$) before any collaboration. These will be called *standard recommendations*. We define the initial utility of the CP (before any collaboration) as the expected revenue (per view) minus the expected price it has to pay to the CDN:

$$U^0_{CP} = \sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{K}} \frac{\alpha_u}{N_u} y^0_{ui}[R_{ui} - \sigma_i\lambda]. \tag{2}$$

Although the precise models on how the CP chooses the standard recommendations, *i.e.,* the values of y^0_{ui}, are not known (since it constitutes business information), these models seem to follow two main directions [1], [2]: they either aim for long-term user engagement (by recommending contents of high utility to the user) or for short-term user engagement (by recommending contents that, although not necessarily close to the user's preferences, can bring high revenue). In the first case, this means priority to r_{ui}, while in the second case priority to R_{ui}. The definition of U_{CP} is generic enough to capture both directions.

Remark 1. Note that we do not account for the revenue that comes from the content requests that are not a result of recommendations, *i.e.*, when, with probability $(1 - \alpha_u)$, the user does not follow any of the recommendations. Since this work proposes a collaboration scheme on the grounds of the recommendations, the requests outside of the recommendations do not affect the collaboration.

[3]CPs could estimate such functions based on user behavior. Such functions are often concave since they capture diminishing returns beyond a minimum content utility.

[4]In order to capture different pricing schemes where the CDN charges the CP per Gb *delivered* (and not requested), the delivery price λ could be multiplied by the probability of abandonment by the user.

From the CDN's perspective, its revenue from the CP depends on users' requests and on the price λ. If a user u requests a recommended content i, then the CDN receives $\lambda\sigma_i$ currency units from the CP. The associated cost that the CDN has to pay is the retrieval cost that depends on the CDN's caching decisions. For this reason, we present a generic cache network model in order to capture a variety of scenarios.

The CDN manages a set of C caches with capacity $\mathcal{C}_j$, $j = 1,\ldots,C$, where $\mathcal{C}_j \ll \sum_{i\in\mathcal{K}}\sigma_i$, as is common in most caching setups [11], [12]. We let x_{ij} be the input variables, where $x_{ij} = 1$ when the content i is cached in cache j, and $x_{ij} = 0$ otherwise. We denote the corresponding matrix by $X = \{x_{ij}\}_{i,j}$. We let $\mathcal{C}(u)$ be the subset of caches that a user u has access to. A request for content i by user u is served by one of the caches belonging to $\mathcal{C}(u)$ where the requested content is stored, *i.e.*, by one of the caches of the set $\{j : j \in \mathcal{C}(u) \text{ and } x_{ij} = 1\}$. Based on the analysis in Sec. II-A, every link between user u and the caches in the set $\mathcal{C}(u)$ is characterized by a cost. Specifically, we let k_{uj} denote the cost of retrieval (for the CDN) per Gb for user u by the cache j. The CDN serves each request through the lowest-cost cache that has the requested item, as is common in most caching setups [11], [12]. We denote the sequence of increasing user-cache costs by $k_{u(1)}, k_{u(2)}, \ldots, k_{|\mathcal{C}(u)|}$. Then, the retrieval cost for content i by user u is:

$$K_{ui}(X) := \sum_{j=1}^{|\mathcal{C}(u)|} \big[\sigma_i k_{u(j)} x_{i(j)} \prod_{l=1}^{j-1}(1 - x_{i(l)})\big]. \tag{3}$$

Given the caching and recommendation decisions before the CP-CDN collaboration, the initial utility of the CDN is expressed as the expected revenue (from the delivery contract) minus the expected retrieval cost:

$$U^0_{CDN} = \sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{K}} \frac{\alpha_u}{N_u} y^0_{ui}[\sigma_i\lambda - K_{ui}]. \tag{4}$$

C. Towards Cooperative Recommendations

As outlined in Sec. I-B, the CP-CDN collaboration is based on the recommendations. The CP would collaborate with the CDN in exchange for a reduction on the content delivery fees. In particular, this reduction will apply to the requests that are the result of cooperative recommendations. We let ϱ denote this reduction on the price λ, where $0 < \varrho < 1$. The value of ϱ is either set by the CDN or by a regulatory authority (who acts as a mediator for their collaboration). We denote by y_{ui} the cooperative recommendation (control) variables (that are the result of the collaboration), and by $Y = \{y_{ui}\}_{u,i}$ the corresponding matrix. We assume that $y_{ui} \in [0,1]$, *i.e.*, the variables are continuous, and thus, y_{ui} expresses the probability the content i features in the recommendations list of user u. This captures the fact that a user is not likely to receive the same recommendations during two different sessions.

The new estimated price that the CP would have to pay to the CDN is $\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{K}}\frac{\alpha_u}{N_u}\sigma_i\lambda y_{ui}[1 + \varrho(y^0_{ui} - 1)]$. Specifically, if a content i is likely to be recommended now but it was not before the collaboration (*i.e.*, $y_{ui} > 0$ and $y^0_{ui} = 0$),

then the reduction ϱ applies. If, on the contrary, the content continues to be recommended as before (*i.e.*, $y_{ui} > 0$ and $y^0_{ui} = 1$), no reduction applies. We note that the requests that do not come through recommendations (see Remark 1) are not subject to any reduction. Therefore, the new utilities are:

$$U_{CP}=\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{K}}\frac{\alpha_u}{N_u}y_{ui}\Big[R_{ui}-\sigma_i\lambda[1+\varrho(y^0_{ui}-1)]\Big], \quad (5)$$

$$U_{CDN}=\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{K}}\frac{\alpha_u}{N_u}y_{ui}\Big[\sigma_i\lambda[1+\varrho(y^0_{ui}-1)]-K_{ui}\Big]. \quad (6)$$

We implicitly assume that during the CP-CDN collaboration, the caching allocation remains the same. In fact, if the caching allocation changes, the terms of the collaboration change, and thus, a new collaboration should be negotiated.

III. Problem Formulation and Analysis

In this section, we will formulate the optimization problem that will allow us to find the cooperative recommendations (*i.e.*, the variables Y). In fact, the new recommendations should be selected in a way that the collaboration is interesting/beneficial for both the CP and the CDN. Firstly, it is clear that the collaboration should provide incentives for both entities. This means that the cooperative recommendations should satisfy: $U_{CP} \geq U^0_{CP}$ and $U_{CDN} \geq U^0_{CDN}$. Moreover, both entities would like to benefit as much as possible from the collaboration. In particular, the CDN that would receive reduced delivery fees needs to compensate this loss by reducing its internal costs (term K_{ui}). That would imply that the cooperative recommendations should include cached items. At the same time, the CP would not like to deteriorate its revenue (term R_{ui}), which implies that the cooperative recommendations should be contents that bring high revenue to the CP. In conclusion, the cooperative recommendations should lead to a Pareto optimal allocation of gains.

Having in mind the aforementioned desired properties of the cooperative recommendations, we model the optimization problem of the collaboration as a Nash Bargaining Solution (NBS) [8], [13]. The NBS is a collaboration mechanism that has been employed, among others, in problems of spectrum access coordination [14], bandwidth allocation [15], and content caching [16]. To the best of our knowledge, this is the first time such a mechanism is employed for selecting recommendations within OTT streaming services.

The NBS is defined as the maximization of the product of payoffs (*i.e.,* the utility gains) of the two entities, or equivalently the maximization of the logarithm of this product [13]. We thus formulate the collaboration problem as follows:

Collaboration Problem.

$$\max_{Y} \ [\log(U_{CP}-U^0_{CP})+\log(U_{CDN}-U^0_{CDN})] \quad (7)$$

$$\text{s.t.} \sum_{i\in\mathcal{K}} y_{ui} = N_u, \ \forall u\in\mathcal{U}; \quad (8)$$

$$U_{CP}-U^0_{CP}\geq 0; \quad (9)$$

$$U_{CDN}-U^0_{CDN}\geq 0; \quad (10)$$

$$y_{ui}\in[0,1], \ \forall u\in\mathcal{U}, i\in\mathcal{K}, \quad (11)$$

where $U_{CP}, U^0_{CP}, U_{CDN}, U^0_{CDN}$ are given in (5), (2), (6), and (4) respectively. The constraints in (8) suggest that each user receives N_u recommendations[5], and those in (11) capture the probabilistic nature of the recommendation variables. The inequalities (9) and (10) guarantee that the collaboration will benefit both parties. In fact, (U^0_{CP}, U^0_{CDN}) is the "disagreement point" of the collaboration: if $U_{CP} < U^0_{CP}$ or $U_{CDN} < U^0_{CDN}$, there will be no agreement on collaboration and the CP will keep its standard recommendations while the CDN will not offer a price discount. Note that (9) and (10) are implied by the logarithms in (7) and could be omitted.

Nash proved that, by formulating the optimization problem in this way, the solution uniquely satisfies some interesting properties (also called Nash's axioms) [8], [13]. First, the solution is Pareto optimal, *i.e.*, there should be no other feasible solution that is better than the solution for one entity and not worse than the solution for the other entity. Furthermore, due to the constraints (9) and (10), the payoff of every entity is no worse than the payoff it would get without collaboration, *i.e.*, (U^0_{CP}, U^0_{CDN}). Finally, if the positions of the two entities (in terms of utility functions and the disagreement point) are symmetric, then the solution treats them symmetrically.

From the perspective of the users, it is in the CP's best interest to recommend contents of high utility r_{ui}, as modeled in (1). For this reason, the CP may impose thresholds on the utility (*i.e.,* relevance to the users) of the cooperative recommendations as additional constraints in the Collaboration Problem. For example, for every $u\in\mathcal{U}$, adding the constraint $\sum_i r_{ui}y_{ui} \geq 0.8\sum_i r_{ui}y^0_{ui}$ ensures that the aggregate utility of the cooperative recommendations will not be less than 80% of the aggregate utility of the initial recommendations y^0_{ui}.

In the following lemma, we show that the problem can be solved efficiently.

Lemma 1. The objective function of the Collaboration Problem (7)-(11) is concave.

Proof. Since the logarithmic function is concave and non-decreasing, by the rules of composition of concave functions, it suffices to show that $U_{CP}-U^0_{CP}$ and $U_{CDN}-U^0_{CDN}$ are concave. In fact, they are both linear in Y, and therefore, concave. We also note that the set of feasible solutions as defined by the constraints (8)-(11) is a convex set. □

Remark 2. Within this collaboration, the two entities need to exchange information about their utility functions, *i.e.,* cost and revenue functions, that, one could argue, constitutes sensitive business information. We highlight here that our framework provides incentives to both entities to exchange such information (since $U_{CP}-U^0_{CP}\geq 0$ and $U_{CDN}-U^0_{CDN}\geq 0$). Furthermore, similar proposed and implemented cooperative mechanisms, *e.g.*, between ISPs and CDNs can also be implemented through trusted third parties [18]. Finally, we note that the user privacy is preserved as the CP does

[5]In practice, the solution of the problem might contain more than N_u positive y_{ui} values for a user u. In this case, a way of choosing *exactly* N_u contents to appear in the recomm. list while respecting the probabilities y_{ui} is described in [17]. Similar to our case, the control (caching) variables in [17] are continuous and subject to a limited number of memory slots.

not need to share user data or user history with the CDN. In fact, without explicit knowledge of the functions f_{ui}, the CDN cannot know the predicted content utilities r_{ui} for the users.

IV. Performance Evaluation

In this section, we evaluate numerically the payoffs that can be achieved through the proposed collaboration scheme. First, we present the scenario considered and the default input parameters used across the different simulations.

Catalogue and Recommendations: Our scenario consists of 100 users who have access to a catalogue of 6000 contents. Without loss of generality, we consider equal-sized contents of 1Gb. Every user receives $N_u = 10$ recommendations and the probability of following the recommendations varies in $[0.6, 1)$ (as in Netflix, where the average is equal to 0.8 [1]). For the matrix of content utilities r_{ui}, a subset of the Movielens dataset [19] containing 5-star ratings of movies was used. The ratings were mapped in the interval $[0, 1]$ and we performed matrix completion to obtain the missing ratings (as in [7]). As for the values R_{ui} (revenue per content), we derived them through the equation $R_{ui} = A_{ui} tanh(2r_{ui})$, where A_{ui} varies in different price ranges (the choice of f_{ui} is justified in footnote 3). The resulting range of R_{ui} will vary across the simulations and will be specified for each different case. The standard recommendations (before any collaboration), *i.e.,* the values y^0_{ui}, were set to be the ones maximizing U^0_{CP} in (2).

Caching Topology: We consider a network of 9 caches and a large cache containing all contents (this could be the CP's origin server or a cache deep in the network). Every user has access to 2 of the caches (chosen randomly) and to the large cache. The cost of retrieval for the CDN varies in the range of $[0, 0.002]$ for the connected caches and is 0.05 for the large cache. The CDN charges the CP \$0.12 per Gb (according to [10]) for the delivery. We assume that the caching allocation as decided by the CDN is based on a popularity distribution over the catalogue as observed by every cache in a time period that precedes the optimization problem. For this, the content popularities observed by cache j are equal to the normalized content utilities r_{ui} aggregated over the connected users, *i.e.*, $r_{ui}/\sum_{u\in\mathscr{C}_j} r_{ui}$, where $\mathscr{C}_j$ is the set of connected users to the cache j. Note that this is a rather optimistic scenario since a lot of contents that would initially be recommended (standard recommendations) are already cached.

For different values of reduction ϱ that will be specified in what follows, we solve the Collaboration Problem and derive the cooperative recommendations. The problem has a smooth and concave objective, with linear constraints, hence standard projected gradient or dual methods work quite fast.

First, we fix the revenue values R_{ui} in the range $[0.2, 0.25)(\$)$. In Fig. 3, we plot the relative increase in utility, *i.e.*, the quantity $100 \cdot (U_P - U^0_P)/U^0_P$, where $P = \{CP, CDN\}$, for different relative cache sizes, *i.e.*, as a fraction of the total size of the catalogue $\sum_{i\in\mathcal{K}} \sigma_i$, and for different values of reduction ϱ. We observe that, for low discount $\varrho = 0.05$ (blue solid lines), the CDN can increase its net revenue by up to 27%, while the CP only by 2%. It is important to highlight here that these points are Pareto optimal points. As explained in Sec. III, this means that there is no other solution

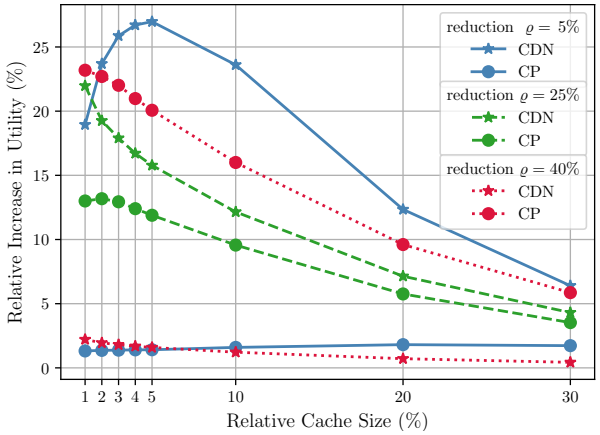

Fig. 3. Relative increase in utility for CP and CDN achieved through collaboration for different relative cache sizes $(1 - 30\%)$ and different price reduction values $\varrho = 0.05, 0.25, 0.4$.

that would benefit one party more without deteriorating the other's gains. The reason behind this difference in gains is the fact that the cooperative recommendation policy favors the cached items, but the CP can only save a small fraction on the delivery fees. Moreover, this is reflected in the CDN's payoffs when the cache size is $1 - 4\%$. In this case, its gains are smaller in comparison with the case where the cache size is equal to 5%. This is because the difference $U_{CP} - U^0_{CP}$ is so small (because of the low discount), that the recommendations should include contents that provide high revenue for the CP, while they are not necessarily among the cached items (whose number is limited). In other words, providing incentives for the CP can potentially limit the gains of the CDN. For moderate discount $\varrho = 0.25$ (green dashed lines), we observe that both parties benefit from important gains (up to 22% and 13%). For high discount $\varrho = 0.40$ (red dotted lines), the CP can enjoy a significant increase of utility (from the saving on the delivery fees), while the CDN's increase is much smaller (since the reduction ϱ is so large that it becomes less profitable).

In general, we observe that the utility gains decrease as the cache size increases. Note that when the cache size is large, it is likely that the standard recommendations, *i.e.,*$\{i \in \mathcal{K} | y^0_{ui} = 1\}$, are among the cached items. Therefore, as the cache size increases, the gains decrease since fewer and fewer recommendations need to be adjusted to favor cached contents.

We stress here the fact that even payoffs of 3% already correspond to very large absolute monetary sums saved (if one extrapolates to a much larger pool of users and requests, as in practice). Note that CPs, like Netflix, report annual profits of more than 2 billion US dollars [20].

Next, Fig. 4 depicts how the revenue values R_{ui} affect the payoffs from the collaboration. Here, we fixed the reduction ϱ to 0.3, and each of the caches is randomly assigned a size in the range of $1 - 5\%$ of the catalogue size. We have plotted the relative increase in utility for both entities for 5 different price ranges from $[0.1, 0.2)$ to $[0.1, 1)(\$)$. We observe that, for the range $[0.1, 0.2)$, the CP could have an increase of 28% of its utility. Then, as the range widens, the payoff for the CP is decreasing. In fact, when the CP's average revenue R_{ui}

is much larger than the delivery fee, a reduction on the fee will not have a significant impact on the CP's utility. On the other hand, the CDN's payoff is not affected as much as the range changes since its utility function does not contain the parameters R_{ui}. Nevertheless, when $R_{ui} \in [0.1, 1)$, the CDN's payoff tends to zero. In particular, as the range widens, both payoffs tend to 0.

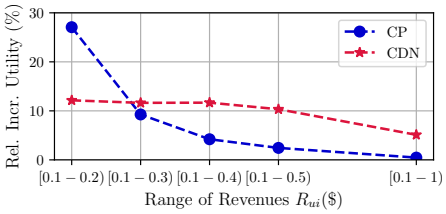

Fig. 4. Relative increase in utility for the CP and CDN achieved through collaboration when the values of R_{ui} vary in different ranges.

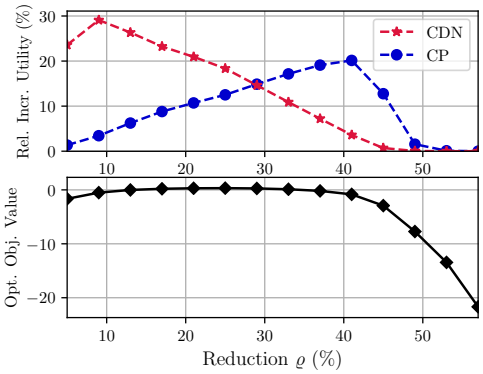

Fig. 5. Relative increase in utility for CP and CDN (top) and optimal objective function values (bottom) achieved through collaboration for different values of reduction $\varrho \in [0.05, 0.57]$.

Finally, for the fixed revenue range of $[0.2, 0.25)$, Fig. 5 depicts the relative payoffs (top subplot) and the optimal objective values (bottom subplot) for different reduction values ϱ. The top subplot shows what was already implied by the results of Fig. 3, *i.e.*, as the reduction ϱ increases, the gains of the CDN decrease, while the gains of the CP increase. Note that, for $\varrho > 0.5$, the payoffs tend to zero as at least one of $U_{CP} - U^0_{CP}$ and $U_{CDN} - U^0_{CDN}$ tends to zero. As a result, the value of the objective function is dropping significantly, as we see in the bottom subplot. Furthermore, in the interval $[0, 0.5]$, the optimal objective function values are not high because of the slow growth of the logarithmic function for arguments greater than 1, and because the scenario is relatively small. In fact, our scenario roughly corresponds to 280 hours of streaming, assuming that a content of 1Gb corresponds to 1.4 hours of video streaming [21]. We note that Netflix alone streams more than 100 million hours per day [1].

V. Conclusion and Future Work

We explored the financial implications of the recommendations and their impact on key economic metrics and arising tradeoffs that involve different OTT market players (CP and CDN). We proposed a novel collaboration scheme in which the CP and the CDN jointly decide on the recommendations in order to favor cached contents. The optimization problem of the collaboration was formulated in such a way that the cooperative recommendations lead to a fair and efficient allocation of financial gains between the two entities. Our numerical evaluations show that, in realistic scenarios, the two entities can benefit of an increase in their expected net revenue of up to 30%. They also revealed that key caching policy parameters, such as caching allocation and cache deployment made by the CDN, have an impact on the resulting economic gains. Therefore, a direction for future work would be to model a collaboration scheme where the CP and CDN jointly decide on caching and recommendations while splitting the gains fairly.

References

[1] C. A. Gomez-Uribe and N. Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems (TMIS)*, vol. 6, no. 4, p. 13, 2016.

[2] Z. Tufekci, "Youtube, the great radicalizer," *The New York Times*, 2018. [Online]. Available: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

[3] E. Gourdin, P. Maillé, G. Simon, and B. Tuffin, "The economics of CDNs and their impact on service fairness," *IEEE Transactions on Network and Service Management*, vol. 14, no. 1, pp. 22–33, 2017.

[4] D. K. Krishnappa, M. Zink, C. Griwodz, and P. Halvorsen, "Cache-centric video recommendation: an approach to improve the efficiency of YouTube caches," *ACM Trans. on Multimedia Comput., Comm., and Applications (TOMM)*, vol. 11, no. 4, p. 48, 2015.

[5] T. Giannakas, P. Sermpezis, and T. Spyropoulos, "Show me the cache: Optimizing cache-friendly recommendations for sequential content access," in *Proc. IEEE WoWMoM 2018*, 2018, pp. 14–22.

[6] L. E. Chatzieleftheriou, M. Karaliopoulos, and I. Koutsopoulos, "Jointly optimizing content caching and recommendations in small cell networks," *IEEE Trans. Mob. Comput.*, vol. 18, no. 1, pp. 125–138, 2019.

[7] D. Tsigkari and T. Spyropoulos, "User-centric optimization of caching and recommendations in edge cache networks," in *Proc. IEEE WoWMoM*, 2020, pp. 244–253.

[8] J. F. Nash Jr, "The bargaining problem," *Econometrica: Journal of the econometric society*, pp. 155–162, 1950.

[9] Maz Systems. (2020) OTT business models. [Online]. Available: https://www.mazsystems.com/types-of-ott-business-models/

[10] Amazon CloudFront. (2020) Pricing. [Online]. Available: https://aws.amazon.com/cloudfront/pricing/

[11] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, 2013.

[12] S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[13] R. B. Myerson, *Game theory*. Harvard University Press, 2013.

[14] Y. Wu and W.-Z. Song, "Cooperative resource sharing and pricing for proactive dynamic spectrum access via Nash bargaining solution," *IEEE Trans. on Par. and Distr. Sys.*, vol. 25, no. 11, pp. 2804–2817, 2013.

[15] H. Yuan, X. Wei, F. Yang, J. Xiao, and S. Kwong, "Cooperative bargaining game-based multiuser bandwidth allocation for dynamic adaptive streaming over http," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 183–197, 2017.

[16] L. Wang, G. Tyson, J. Kangasharju, and J. Crowcroft, "Milking the cache cow with fairness in mind," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 2686–2700, 2017.

[17] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *Proc. IEEE ICC*, 2015, pp. 3358–3363.

[18] B. Frank, I. Poese, Y. Lin, G. Smaragdakis, A. Feldmann, B. Maggs, J. Rake, S. Uhlig, and R. Weber, "Pushing CDN-ISP collaboration to the limit," *ACM SIGCOMM Comp. Comm. Review*, vol. 43, no. 3, pp. 34–44, 2013.

[19] F. M. Harper and J. A. Konstan, "The Movielens datasets: History and context," *ACM Trans. on Inter. Intell. Sys.*, vol. 5, no. 4, p. 19, 2016.

[20] Fortune. (2021) Netflix company profile. [Online]. Available: https://fortune.com/company/netflix/fortune500/

[21] Netflix Help Center. (2021) How to control how much data Netflix uses. [Online]. Available: https://help.netflix.com/en/node/87