Virtual CDN Providers: Profit Maximization through Collaboration

Thanasis G. Papaioannou*, Kostas Katsalis†, Vasilis Sourlas‡, Angelos Amditis‡
*Athens University of Economics & Business (AUEB), Greece.
†Huawei Technologies Duesseldorf Gmbh, Munich, Germany.
‡ Institute of Communication and Computer Systems (ICCS-NTUA), Athens, Greece.

***Abstract*—The proliferation of mobile devices with ubiquitous Internet access has made content access patterns highly volatile and spatio-temporally varying. At the same time, virtualization techniques have enabled the emergence of virtual Content Delivery Network (vCDN) providers, that bundle together a virtual infrastructure by utilizing storage resources anywhere along the path between the access network and the corresponding data center, where the requested content actually resides. In this context, efficient content placement in the various storage layers depends on accurately estimating content access patterns from the different access networks at any given time. At the same time, different vCDN providers compete against each other to provide low content retrieval latency to their users. However, there are some interesting synergies that might emerge among otherwise competing vCDN providers: (*i*) host content of another vCDN provider and, (*ii*) provide own content to the users/customers of other vCDN providers. In this paper, we formulate the overall problem of content placement for multiple vCDN providers that employ collaboration as an overall social-welfare maximization problem. The solution to this problem gives the optimal placement, achievable only in the case of full information. Alleviating the need for full information and considering vCDN providers separately as profit seekers, we also devise a distributed algorithm for content placement by exchanging limited information among them. Extensive simulation experiments show that business collaboration among competing vCDN providers is beneficial, as compared to isolated offerings, and allows them to adapt faster to content pattern changes.**

I. Introduction

The abundance of multimedia content-sharing applications resulted in an enormous increase in the demand for video streaming. A large portion of this demand originates from a fast-growing number of wireless handheld devices. Meeting this demand by increasing the capacity of the network is prohibitively expensive. A preferable solution is to deploy content-storages as close as possible to the mobile edge network, so that users' requests are locally served and network traffic at the core infrastructure is reduced, whereas the overall QoE is increased.

Research on edge caching usually focuses on where to deploy the caches and what to store on them [1], [2], [3]. In the new virtualized era and towards integrated 5G communications, a strong coupling between edge caching and Multi-access Edge Computing (MEC) concepts is increasingly important. MEC offers application developers and content providers cloud-computing capabilities and an IT service environment at the edge of the mobile network [4], [5].

In this context, the concept of virtual elastic Content Delivery Networks (vCDNs) is now feasible and is gaining momentum as a way to realize edge caching concepts by exploiting cloud computing technologies and bridge the gap between the service providers and the telecom operators. The vCDN concept is described for example in the SONATA EU-H2020 project, where CDN providers seek to exploit virtualization capabilities and leverage NFV platforms to deploy caches as VNFs (vCaches) instead of hardware appliances. This is also known as elastic CDN, as proposed by AKAMAI and JUNIPER. Caching services can be now offered by both traditional CDN and vCDN providers. The former offers a distributed network of proxy servers that are deployed in multiple points of an end-to-end network. The latter also offer CDN services, however these can rapidly exploit cloud, NFV and SDN technologies and build their infrastructure over virtualized and possibly shared systems.

In this paper, we focus on the vCDN providers, and we consider that storage space is deployed from the edge cloud network up to the remote data center(s). From the modeling perspective, we depart from the traditional approach where a CDN provider is the owner and the only entity that exploits the cached data and the storage space, and we extend it with synergies among vCDNs where content and services are efficiently placed and shared. Our approach is built on the following two concepts: (*i*) whenever a vCDN receives a request for an object, it can serve it from its own storage space or can request it from other vCDNs directly (more cost-effective and/or closer). (*ii*) Virtual storage space can potentially be shared between vCDNs.

With the goal of maximizing each vCDN's profit, we formulate an optimization problem where in addition we also consider that every vCDN provider is concurrently offering content services to multiple mobile network operators (MNOs). Since an end-user is associated with a MNO, in our modeling approach different contracts between the vCDN providers and the MNOs can be established. The contributions of this paper can be summarized as follows:

- We describe the concept of vCDN providers and the various stakeholders in the new virtualized edge network.
- We formulate a storage update optimization framework, where collaboration through data and storage exchange can be established between vCDN providers. We also consider different pricing schemes depending on the bilateral agreement between the MNO that the end user is associated with and the respective vCDN provider.
- We model both the centralized social welfare maximization problem and the more-practical distributed case where each vCDN provider independently takes its own caching decisions so as to maximize its profit, with minimal input information from the other providers.
- We provide extended simulation results and assess them on the benefits of exploiting vCDN collaboration.

II. Motivation & Background Information

A. Motivation

This work stems directly from the cloud industry and the need of service providers to bring data and services as closer as possible to end-users. The main features of the mobile edge computing are proximity to end-users; direct access to real-time network information; spatio-temporal context awareness; mobility support, and exploitation of the RAN agnostic network application distribution platform. Before issuing a request for a service or data, the MEC systems check if the request can be served locally; otherwise the traffic is delivered through the gateway systems to the Internet.

For the edge-content caching use case, the benefits in the backhaul and transport network traffic can be significant and this also impacts positively the users perceived QoE. The reason is that by having a cloud infrastructure at the edge network, we can avoid the dominant delay factor in the end-to-end path, which is the latency between the core network and the data centers [6], [7], [8].

Indeed, large content providers want to implement their own caching systems at the edge network, inside the domain of the ISP provider [9]. However, this approach is still not mature, because of conflict of interests between the service providers, the content providers and the ISPs [10]. This is mainly attributed to the fact that CDN providers control both the placement of content, as well as the decision on where to serve user requests. These decisions are taken without knowledge of the network topology and state in terms of traffic load, and can result in network performance degradation affecting the service quality experienced by the end users.

Although the approach is appealing, in practice the closer to the end-users, storage space becomes more limited, contended and expensive. Furthermore (*i*) different vCDN providers may have access to different content and at different cost; (*ii*) different vCDN providers have different and dynamic demand for different content objects; and (*iii*) different business relationships can be build between vCDN providers, as well as between vCDN providers and MNOs. In this complex environment a non-trivial formulation is required, when it comes to an efficient storage placement.

In our approach, for any vCDN provider, the demand may arise directly due to own customers' requests or indirectly due to requests coming from other vCDN providers. Actually, when a vCDN doesn't have the content, can either request it from a neighboring vCDN or forward the request to a remote data center or the cloud. In addition, storage space of the virtualized underlay network can be shareable with other vCDN providers. If storage is not used or is not profitable to use it, it can be shared with other vCDN providers. In this work we exploit all possible synergies between vCDN providers targeting the maximization of each vCDNs profit.

B. Related work

There is already some prior work in opening RAN and virtualization technologies for efficient CDN offerings. In [11], authors propose a new data plane LTE design in which they rely on L2 or L3 transport and SDN to perform all the required QoS, mobility management and security. As a result of using SDN, the GTP tunnel from the current LTE networks can be removed, thus allowing the placement of storages at any node or switch at the edge network.

In [12], authors describe the pros and cons when caching is performed in the RAN, in the EPC, in both the RAN and the EPC while they also describe CCN caching and they provide a performance comparison of these schemes for delay minimization.

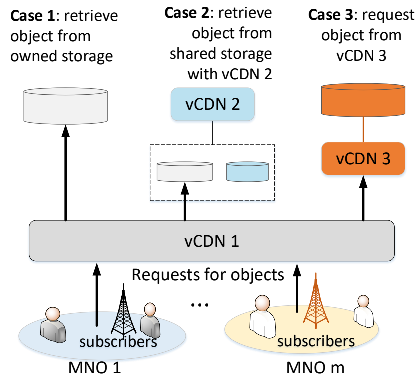

Fig. 1. Interactions between vCDNs.Three cases when a request for object comes from some MNO.

In [13], authors exploit cloud orchestration frameworks and NFV and design a scheme which offers the flexibility to a telecom operator to lease its CDN infrastructure in a dynamic manner, offering a virtual CDN service that can be deployed on demand over the operators private cloud. In [14], authors are also dealing with the vCDN orchestration to optimize the use of resources and improve the performance of the overall SDN/NFV-based CDN function in terms of network operator cost reduction and high streaming quality. Therein [14], they are mainly interested in the migration costs and they do not consider the case of collaboration as in our scenario.

There is also some work that mention collaboration among different CDN providers, albeit differently than considered in this paper. The EPCache framework for caching services at the packet core network with the goal of delay minimization is presented in [15]. Although the authors in [15] consider collaboration between the PDN service gateways, our approach is fundamentally different, since in [15] there is no notion of service differentiation between the service and the content providers.

Transparent caching, where the end-user is not aware of the exact object address, and collaborative caching at the edge networks are considered in [16]. The Access Node in [16] performs on-demand caching according to object popularity and then communicates this information to a centralized controller that holds the cached object and its location information in a hash table (content directory). The controller also provides a look-up service for requested objects. According to this lookup, the routing redirection is made using SDN techniques. The authors in [16] also propose a method to reduce the cache hit fault-positives that trigger unnecessary packet exchange between the edge nodes and the controller. However, collaboration in [16] solely means jointly considering multiple access networks for object caching, while the practical economic implications of collaboration are ignored, as opposed to our approach.

To the best of our knowledge, our approach is the first one that considers practical types of synergies among different vCDN providers that attempt to maximize their individual profits, while increasing the users' perceived QoE.

III. System model and Problem Statement

A. Definitions, Player Roles and the vCDN Provider

For clarity, we provide the following definitions that are necessary for the model articulation.

Service Provider (SP): is an entity that provides services to its customers in exchange of payment.

CDN and vCDN providers: both the CDN and vCDN offer a distributed network of proxy servers that are deployed in multiple points of an end-to-end network to provide storage services. A SP is utilizing these services. By exploiting cloud technologies and methodologies, a vCDN provider can utilize shared virtualized infrastructures, operate its storage network over both self-owned and shared infrastructures and new business models can be created. As elaborated later, we are considering that content services can be realized even in the total absence of owned storage space.

Mobile Network Operators (MNOs): is the communications service provider that offers the RAN infrastructure and supports all the connectivity services.

B. Modeling Assumptions

A1: We consider that all types of content (e.g., video, music, etc.) are stored in the form of objects.

A2: Every vCDN is also the service provider that directly interacts with the MNOs. When a request is made from a MNO user to vCDN, we consider the following cases regarding content retrieval (see Fig. 1), which are irrelevant to the MNO (as long as the requested content is delivered):

- `Case 1`: vCDN retrieves the content from its own storage.
- `Case 2`: vCDN rents storage space from another vCDN. This can be the case of expensive storage in the edge network. The object is retrieved from this storage space.
- `Case 3`: vCDN doesn't have the content and requests the object from another vCDN. The content delivery can be made using web services.

A3: For every vCDN provider, the storage space can be deployed in multiple geographical areas ranging from the access network, up to the remote data center (see Fig. 2); the closer is the storage to the access network the more limited and expensive usually is. To simplify modeling, we consider a grid representation of the storage space, with the rows representing the areas and columns the access layer (e.g., for an $n \times m$ grid, column 1 is the base station, whereas column m is the data center).

A4: We consider that mobile end-users are randomly distributed, while in the same cell multiple MNOs can co-operate. Every end-user is associated with a single MNO.

C. System Model

We assume a system where a set of $\mathcal{V}$ vCDN providers have a network of distributed storages that are used to store a set of objects $\mathcal{O}$. Each vCDN is associated with a number of MNOs belonging in a set $\mathcal{M}$. For each access area (each row in the grid, column equal to zero), every vCDN receives an aggregated request rate for objects from every MNO m. The vCDN associates every request with a MNO, the object requested and the access area. Furthermore, we consider that for every vCDN i every cell in the grid (row i represents an area, column c represents the distance from the access layer) has different unit placement cost represented as $c_i^{r,c}$, has different retrieval cost $h_{n,i},^{o,r,c}$ for object o requested by vCDN n and also different renting cost $w_{i,n},^{o,r,c}$. Intuitively, the closer to the access layer (column zero), the more expensive the storage space and more expensive the object retrieval cost.

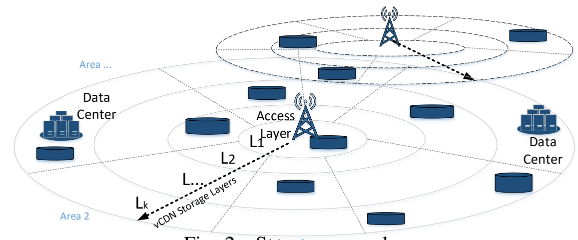

Fig. 2. Storage space layers

TABLE I
NOTATION SUMMARY

- $\mathcal{R}$: set of access areas. An area spans horizontally up to a data center. Every area is associated with a *row id of the grid*.
- $\mathcal{C}$: the set of layers as we move from the access network to the data center. The layer id equals the *column id of the grid*.
- $\mathcal{V}$: set of Virtual CDN Providers (vCDNs).
- $\mathcal{M}$: set of Mobile Network Operators (MNOs).
- $\mathcal{O}$: set of objects.
- s_o: size of object o in number of units.
- $s_i^{r,c}$: storage capacity of vCDN i in row r and column c
- $r_i^{m,o,r}$: request rate for object o send to vCDN i by users associated to MNO m at row r (column $c = 0$).
- $b_i^{j,o,r,k,c}$: benefit of vCDN i for object o retrieved by MNO j from column c and row k for requests originating by area r.
- $h_{i,n}^{o,r,c}$: benefit (price) that vCDN i gets from vCDN n for selling object o from column c and row r.
- $\psi_{i,n}^{r,c}$: rental price per unit storage at cache column c and row r that vCDN i rents to vCDN n.
- $c_i^{r,c}$: cost of placing unit storage by vCDN i in storage space at column l and row r of the grid.
- $x_i^{o,r,c}$: decision variable in $\{0, 1\}$ for vCDN i of placing object o in storage located in row r and column c.
- $w_{i,n}^{o,r,c}$: decision variable in $\{0, 1\}$ for vCDN i of placing object o in storage space of vCDN n in row r and column c.
- $y_{i,n}^{o,r,c}$: decision variable for vCDN i obtaining object o from vCDN n from storage space in row r and column c.

D. Social Welfare Maximization

In this section we formulate the social welfare maximization problem. In this formulation the optimization problem is solved centrally by a single entity that knows all the requests rates for every object in the system and all the vCDN storage capacities. For each vCDN i we identify the following benefits ($B_{i,x}$) and costs ($C_{i,x}$):

- $B_{i,1}$: compensation for serving object o for a request from a MNO m. The object can be actually retrieved by owned storage, rented storage or requested from another vCDN:

$$B_{i,1} = \sum_{o \in \mathcal{O}} \sum_{r \in \mathcal{R}} \sum_{m \in \mathcal{M}} r_i^{m,o,r} \Big(\sum_{d \in \mathcal{R}} \sum_{c \in \mathcal{C}} (x_i^{o,d,c} + \sum_{n \in \mathcal{V} \setminus i} (w_{i,n}^{o,d,c} + y_{i,n}^{o,d,c})) \cdot b_i^{m,o,r,d,c} \Big)$$

- $B_{i,2}$: benefit from renting storage to other vCDNs:

$$B_{i,2} = \sum_{o \in \mathcal{O}} \sum_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} \sum_{n \in \mathcal{V} \setminus i} s_o \cdot w_{n,i}^{o,r,c} \cdot \psi_{i,n}^{r,c}$$

- $B_{i,3}$:benefit for requests vCDN i serves to vCDN n:

$$B_{i,3} = \sum_{o \in \mathcal{O}} \sum_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} \sum_{n \in \mathcal{V} \setminus i} \sum_{m \in \mathcal{M}} r_n^{m,o,r} \cdot y_{n,i}^{o,r,c} \cdot h_{i,n}^{o,r,c}$$

- $C_{i,1}$: cost for storing or retrieving object o:

$$C_{i,1} = \sum_{o\in\mathcal{O}} \sum_{r\in\mathcal{R}} \sum_{c\in\mathcal{C}} \Bigg(s_o (c_i^{r,c} \cdot x_i^{o,r,c} + \sum_{n\in\mathcal{V}\setminus i} w_{i,n}^{o,r,c} \cdot \psi_{n,i}^{r,c}) + \sum_{n\in\mathcal{V}\setminus i} y_{i,n}^{o,r,c} \cdot h_{n,i}^{o,r,c} \Bigg)$$

- $C_{i,2}$: cost for renting storage from other vCDNs:

$$C_{i,2} = \sum_{o\in\mathcal{O}} \sum_{r\in\mathcal{R}} \sum_{c\in\mathcal{C}} \sum_{n\in\mathcal{V}\setminus i} s_o \cdot w_{n,i}^{o,r,c} \cdot c_i^{r,c}$$

The social welfare maximization problem is formulated as:

$$\textbf{maximize} \quad \sum_{i\in\mathcal{V}} (B_i - C_i) \tag{1}$$

$$\textbf{subject to} \quad \sum_{o\in\mathcal{O}} s_o \cdot \left(x_i^{o,r,c} + \sum_{n\in\mathcal{V}\setminus i} w_{n,i}^{o,r,c} \right) \leq s_i^{r,c}, \tag{2}$$

$$\sum_{r\in\mathcal{R}} \sum_{c\in\mathcal{C}} \left(x_i^{o,r,c} + \sum_{n\in\mathcal{N}} (w_{i,n}^{o,r,c} + y_{i,n}^{o,r,c}) \right) \leq 1, \tag{3}$$

$$y_{i,n}^{o,r,c} \leq x_n^{o,r,c} + \sum_j w_{n,j}^{o,r,c} \tag{4}$$

$$x_i^{o,r,c} \in \{0,1\}, w_{i,n}^{o,r,c} \in \{0,1\}, y_{i,n}^{o,r,c} \in \{0,1\} \tag{5}$$

$$\forall\ i \neq n \in \mathcal{V}, o \in \mathcal{O}, r \in \mathcal{R}, c \in \mathcal{C}$$

$x_i^{o,r,c}, w_{i,n}^{o,r,c}, y_{i,n}^{o,r,c}$ are the control variables of the problem and take values in $\{0,1\}$. In the constraints above, the capacity constraints are articulated by Eq.(2), while Eq.(3) formally dictates that for every vCDN an object can only be retrieved disjointly by either owned storage, rented storage or requested. Eq.(4) is used to guarantee that an object is available at vCDN n, before being requested by vCDN i.

IV. Distributed Optimization

Centrally solving the problem in Eq.(1) would maximize the social welfare of the overall ecosystem. However, the solution to this problem necessitates that information on request rate per content object is publicly known for the users of each MNO, which is not realistic. Also, solving the problem in Eq.(1) would necessitate that the storage topology and the storage capacities at each layer and access network are known for each vCDN provider. Moreover, vCDN providers are in fact *competitive* to each other; each of them seeks to place content objects appropriately, so as to maximize their own profit. Thus, the vCDN providers cannot be considered to exchange information on the demand patterns for their content. To make things more complex, prices for content exchange and content placement among vCDNs may not be fixed. In this context, it makes more sense to consider each vCDN as an individual profit-maximizing agent in a competitive market. Each vCDN provider cannot know in advance the demand from other vCDN providers either for its content objects or for its storage at each layer and access network cell. Consider a distributed setting where each vCDN provider i seeks to place its content objects, so as to maximize its profit (Eq.(1) for the specific i) in an iterative fashion. Whenever request rates or vCDN topology change significantly, a vCDN provider employs the following distributed algorithm:

STEP 1: Each provider i solves its own profit-maximization problem and determines $\mathbf{x}_i = \{x_i^{o,r,c}\}$, while assuming $y_{j,k}^{o,r,c} = 0$, $w_{j,k}^{o,r,c} = 0$, $\forall\ i,j,k \in \mathcal{V}, r \in \mathcal{R}, c \in \mathcal{C}$.

STEP 2: Each provider i announces $\mathbf{x}_i$ to others in quest for available stored objects, any possible spare storage and associated costs.

STEP 3: Each provider i finds $\mathbf{y}_i = \{y_{i,n}^{o,r,c}\}$ and $\mathbf{w}_i = \{w_{i,n}^{o,r,c}\}$, $\forall n \in \mathcal{V}, o \in \mathcal{O}, r \in \mathcal{R}, c \in \mathcal{C}$, independently from the others based on $\mathbf{x}_n$, $\forall n \in \mathcal{V}$ and the announced spare storage.

STEP 4: Each provider i announces $\mathbf{y}_i$ and $\mathbf{w}_i$ to others. In case of conflict, each provider chooses where and which object(s) to place in the storage of neighboring providers.

STEP 5: Each provider i, based on $\mathbf{x}_{-i}$, $\mathbf{y}_{-i}$ and $\mathbf{w}_{-i}$ of others, re-determines its own $\mathbf{x}_i$, $\mathbf{y}_i$ and $\mathbf{w}_i$ as follows: local objects that are not requested by others may be discarded in favor of hosting them remotely or fetching them from others, and new objects maybe stored.

STEP 6: Resulting tentative $\mathbf{x}_i^?$, $\mathbf{y}_i^?$ and $\mathbf{w}_i^?$ are communicated to others and conflicts are resolved on a local-basis in a profit-maximizing manner (according to Eq.(1) for the specific i). Final $\mathbf{x}_i^*$, $\mathbf{y}_i^*$ and $\mathbf{w}_i^*$ per vCDN provider i are communicated back to them.

STEP 7: If all $\mathbf{x}_i, \mathbf{y}_i$ and $\mathbf{w}_i$ for each provider $i \in \mathcal{V}$ have not been changed after an iteration, then the process is completed. Otherwise, changed values are communicated to others and the process is repeated from STEP 5.

Notice that, since, at each iteration of the above algorithm, fewer control variables change value (by respecting the current commitments to other vCDNs for hosting or providing objects), it is expectable (and experimentally found in Section V) that the algorithm converges. The effectiveness of this algorithm is experimentally assessed in Section V.

One should consider that price formation for hosting objects of others or providing own objects to others depend on the relative scarcity of the constrained resources (i.e., storage, network links, etc.), the relative demand for storage or objects and the number of players in the market. For simplicity, we consider a pure-competition setting in which the vCDN providers are considered to be *price takers*.

V. Evaluation

The performance evaluation is used for the assessment of the benefits of using collaboration schemes between the vCDN providers and the identification of trade-offs. The benchmarking process for our approach will consider: (*i*) demonstration of performance on how system and statistical parameters affect net benefit and (*ii*) overhead and trading-off analysis to face limitations on communication and processing overhead of the proposed approach.

A. Experimental setup

We use a custom Python simulator and IBM CPLEX in order to solve the relevant MILP problems. For the storage space, we consider a $r \times c$ grid topology, where the access layer is in column zero (the base stations) and different geographical areas are expressed by different rows. Data centers are located in the last column of the grid. Each vCDN can store content in any cell of the grid or rent storage to other vCDNs as long as its own storage limits are not violated.

TABLE II
BASELINE SETUP PARAMETRIZATION

Parameters:
$\|\mathcal{V}\| = 3$, $\|\mathcal{M}\| = 1$, $\|\mathcal{O}\| = 1000$, grid 1×3, $\lambda_i = 2000$,
$s_o = 5$, $z_{pop} = 2.2$, $s_i^{r,0} = 0.1 \cdot \|\mathcal{O}\| \cdot s_o$
Benefits & Costs:
$b_i^{j,o,src,dst,0} = 1000$, $\psi_{i,n}^{r,c} = 1.1 \cdot c_i^{r,c}$, $h_{i,n}^{r,c} = 0.1 \cdot \bar{b}$, $c_i^{r,c} = 2$

Every MNO (m) in the system sends an aggregate request traffic to each vCDN i that follows a Poisson distribution with average λ_i. The request rate for each object ($r_i^{m,o,r}$) is determined from its popularity. As the file popularity in the Internet follows Zipf distribution, we approximate the popularity of the objects by a Zipf law of parameter z_{pop}. The total storage capacity in each cell for each vCDN is a tunable parameter ($s_i^{r,c}$), similar to the rental cost ($c_i^{r,c}$), the benefit of object delivery $b_i^{j,o,r,k,c}$ from area k when the request originates from area r, the rental price ($\psi_{i,n}^{r,c}$) and the benefit of selling object o ($h_{i,n}^{o,r,c}$). For the baseline setup we assume same object size s_o. The baseline setup parametrization is depicted in Table II.

B. Simulation results

We compare performance and optimality between the following three schemes:

- *Centralized With Collaboration (CWC):* A centralized entity has knowledge of the entire system, traffic rates, pricing, storage availability and retrieval costs. This entity solves the total welfare maximization problem in Eq.(1) considering all the vCDNs at the same time.
- *Non-collaborative (NC):* Each vCDN i solves its own local maximization problem (Eq.(1) for specific i), without renting storage or retrieving objects from other vCDNs.
- *Distributed With Collaboration (DWC):* Each vCDN i solves independently the relevant net benefit maximization problem (Eq.(1) for specific i), considering however the ability to rent storage or retrieve objects from other vCDNs. Each vCDN follows the heuristic algorithm proposed in Section IV.

In Fig.3(a), we present a comparison between the three schemes for 100 slots, by means of overall net benefit obtained. In this experiment we use the baseline configuration with Poisson request rate and Zipf object popularity, different for each vCDN. As we can see the benefits of exploiting collaboration schemes can dramatically change the overall net benefit. The reason is that objects that are not stored in local storage can be retrieved from other vCDNs or placed in rental storage, contributing to the overall gain when serving customer requests. Note that the actual net benefit obtained depends on the system parameterization and fine tuning is required. However, in all cases both CWC and DWC outperform the case of NC (which is actually the optimal solution when no collaboration is in effect).

In Fig.3(b), we present the allocation of the decision variables x, y and w for the case of CWC and DWC. Although the distributed algorithm performs with an average deviation of $\approx 10\%$ from the CWC optimal, the decision variables allocation on average may vary up to 50%. Note that the DWC heuristic policy relies on an iterative scheme that in transient state firstly decides local x values (STEP 1), and in the following steps for remote storage opportunities and remote requests. The deviation from the goal progressively is minimized, however policy improvements are required to minimize the re-allocations necessary at each step.

In Fig.3(c), the effects on increasing the number of objects are presented. Again the benefits of vCDN collaboration are obvious and as expected both the CWC and DWC outperform the case where each vCDN is taking placement decisions (NC policy) without any form of collaboration. In the case of vCDN collaboration, the system is able to confront to the objects number increase smoothly up to a point where all the capacity limitations are reached and all fetching opportunities are satisfied. This saturation point for the case of CWC is reached earlier, since no collaboration is exploited and thus no extra objects can be fetched.

In Fig.3(d), we investigate the effects in performance while increasing the number of vCDNs and keeping the same storage constraints and number of objects. By increasing the number of vCDNs more collaboration opportunities are possible and for both *CWC* and *DWC* a linear net benefit increase can be observed. In order to stress the system further, average results are also presented for the case where even more objects are placed by the vCDNs providers. We highlight that for each vCDN a different Zipf distribution was used, meaning that each content may had different popularity.

In Fig.3(e), we varied $\psi_{i,n}^{r,c}$ (rental price per unit storage) and $h_{i,n}^{o,r,c}$ (benefit that vCDN i earns for serving content to vCDN n) respectively. As we can observe from Fig.3(e), lower values for $\psi_{i,n}^{r,c}$ lead to better overall system net benefit, while the ratio between dwc and cwc is essentially the same. The reason is twofold. In one hand for the overall system, the storage rental terms are canceled out; the benefit a vCDN is getting from renting storage (B_2 term) is a cost for the other vCDN that is using the extra storage. On the other hand, in both cases of low and high rental prices, the overall benefit is dominated by the b factor; if the storage is expensive then every vCDN tries to maximize the available objects either by storing them locally or just retrieving them from other vCDNs (thus y is increased). If the rental costs are cheap, then every vCDN tries to store as many objects as possible remotely.

Finally, in Fig.3(f), we varied the h value to investigate the effects of changing the remote retrieving cost/benefit. In the Y axis the ratios for the number of y's allocation is depicted and the corresponding net benefit ratio between the DWC and CWC. In the X-axis we increased h by a factor $h = b \cdot a$ where a varied form 0.1 to 1.9 with a step of 0.2. As we can observe, while $h < b$ the y allocation ratio equal to 0.8 is preserved, whereas when $h > b$ then the ratio dramatically falls to 0.1. The reason is that in this case every vCDN tries to store locally or store remotely, rather than requesting for objects. Due to additional storage limitations, the overall *DWC* performance is decreasing as less collaboration opportunities are desirable.

C. Practical Considerations

In the devised model even in the distributed case, computation complexity is exacerbated with the increase of parameters like the number of objects or the grid size. Although the storage update period can be large, simple heuristics can be used to solve the relevant MILP problems in near polynomial time. In practical deployment scenarios, updating the object placement at each step imposes extra configuration and network cost. However, in our model the popularity of each object is determined according to a Zipf distribution and the request

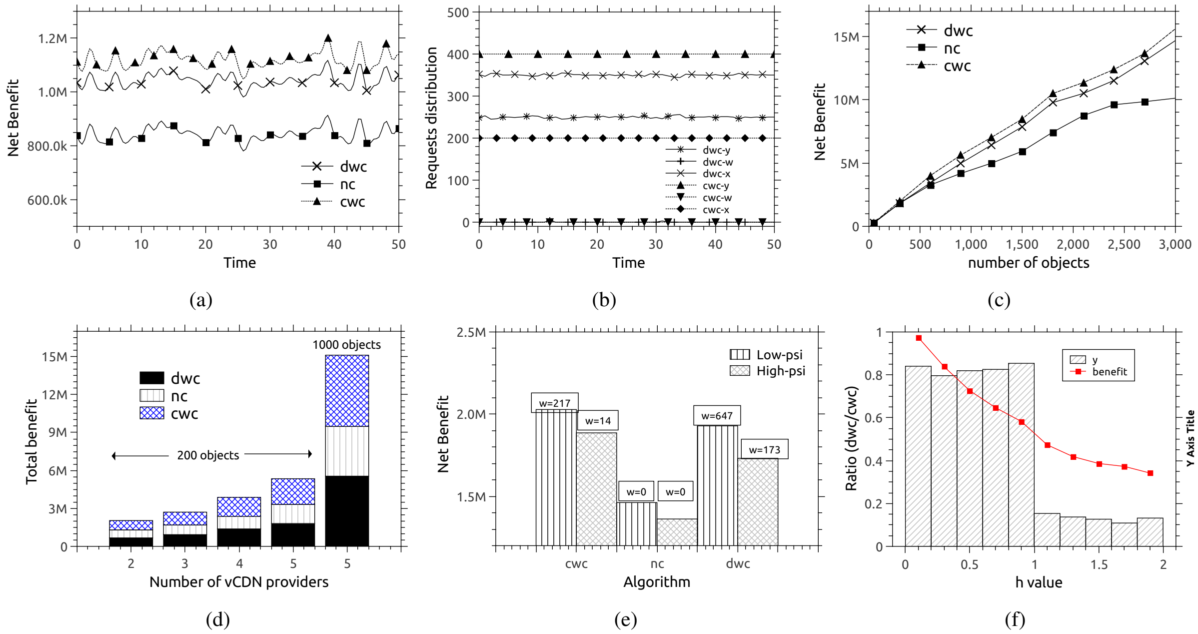

Fig. 3. The examined aspects of vCDN collaboration in overall Net Benefit

rate according to a Poisson distribution. In practice, these are stationary and are not radically changed at each slot. Thus, the solution to the relevant optimization problems will keep relatively unchanged both the sorting of the objects by means of net benefit contribution and object allocation. Regarding the system architecture and the end-to-end storage orientation, in mobile networks like for example LTE, using SDN routing, one can remove the GTP tunnels and actually place caches in any switch in the network path. In principle, because of the GTP tunneling (without any patches and modifications), one could just place storage space in the eNodeB or after the PGW. By considering a multiple-storage -layers architecture, our model can accommodate both design principles.

VI. Conclusions and Future Work

In this paper, we addressed the problem of collaborative content caching among multiple competing virtual CDN providers, the storage resources of which may span end-to-end between the edge networks and the data centers. We considered two interesting synergies that might emerge among competing vCDN providers: (*i*) host content of another vCDN provider and, (*ii*) provide own content to the users/customers of other vCDN providers. We first considered full information and formulated the problem of optimal content placement for all vCDN providers as a social-welfare maximization problem. Then, moving more to the practical side, we decoupled the overall problem as profit-maximizing ones for each individual vCDN provider; vCDN providers exchange limited amount of information on content/storage availabilities and prices for object hosting/relaying. Experimental results indicate that business collaboration among competing vCDN providers is beneficial, as compared to isolated offerings, and allows them to adapt faster to content pattern changes. These CDN synergies and information exchanges can amend existing standards for CDN interaction, i.e., RFC 7336. Future work directions include theoretical analysis of the distributed case and further investigation of its convergence process. Furthermore, we plan to investigate different collaboration schemes, taking also into account network constraints besides storage limitations.

Acknowledgements

Part of this work has been conducted as part of ICT4CART project which has received funding from the European Unions Horizon 2020 research innovation programme under grant agreement No. 768953.

References

[1] V. Sourlas *et al.*, "Distributed Cache Management in Information-Centric Networks," *IEEE TNSM*, vol. 10, no. 3, Sept. 2013.

[2] ——, "Storage planning and replica assignment in content-centric publish/subscribe networks," *Computer Networks*, vol. 55, no. 18, 2011.

[3] N. Giatsoglou *et al.*, "D2d-aware device caching in mmwave-cellular networks," *IEEE JSAC*, vol. 35, no. 9, pp. 2025–2037, 2017.

[4] M. Patel *et al.*, "Mobile-edge computing introductory technical white paper," *MEC industry initiative*, 2014.

[5] O. Ascigil *et al.*, "On Uncoordinated Service Placement in Edge-Clouds," in *IEEE CloudCom*, Dec. 2017, pp. 41–48.

[6] L. Tong *et al.*, "A Hierarchical Edge Cloud Architecture for Mobile Computing," in *IEEE INFOCOM*, 2016.

[7] K. Katsalis *et al.*, "Content Placement in Heterogeneous End-to-end Virtual Networks," in *ACM SAC*, 2015, pp. 602–608.

[8] H. Tan *et al.*, "Online Job Dispatching and Scheduling in Edge-Clouds," in *IEEE INFOCOM*, 2017, pp. 1–9.

[9] A.-J. Su *et al.*, "Drafting behind Akamai," *SIGCOMM CCR*, vol. 36, no. 4, pp. 435–446, Aug. 2006.

[10] A. Antonopoulos *et al.*, "Shedding light on the internet: Stakeholders and network neutrality," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 216–223, July 2017.

[11] J. Costa-Requena *et al.*, "SDN optimized caching in LTE mobile networks," in *IEEE ICTC*, 2014.

[12] X. Wang *et al.*, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Comm. Mag.*, vol. 52, no. 2, Feb. 2014.

[13] P. A. Frangoudis *et al.*, "CDN-As-a-Service Provision Over a Telecom Operator's Cloud," *IEEE TNSM*, vol. 14, no. 3, pp. 702–716, 2017.

[14] H. Khedher *et al.*, "Optimal and Cost Efficient Algorithm for Virtual CDN Orchestration," in *IEEE LCN*, Oct. 2017, pp. 61–69.

[15] J. Zhu *et al.*, "EPCache: In-network video caching for LTE core networks," in *WCSP*, Oct. 2013.

[16] Z. Cao and P. Papadimitriou, "Collaborative Content Caching in Wireless Edge with SDN," in *ACM CCDWN*, 2015.