IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 12, 2018, accepted December 21, 2018, date of publication December 25, 2018,
date of current version January 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889787

Edge Cache-Based ISP-CP Collaboration Scheme for Content Delivery Services

CHAO FANG[1,2], HAIPENG YAO[3], ZHUWEI WANG[2], PENGBO SI[2], YIXIN CHEN[4], (Senior Member, IEEE), XIAODONG WANG[5], (Fellow, IEEE), AND F. RICHARD YU[6], (Senior Member, IEEE)

[1]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2]Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China
[3]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[4]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA
[5]Electrical Engineering Department, Columbia University, New York City, NY 10027, USA
[6]Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: Chao Fang (fangchao@bjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601330, in part by the Natural Science Foundation of Beijing Municipality under Grant 4174081, and in part by the Scientific Research Plan of Beijing Municipal Commission of Education under Grant KM201910005026.

ABSTRACT With the explosive increase of mobile data traffic, content delivery issue on the Internet is a growing concern for both Internet service providers (ISPs) and content providers (CPs). To improve content transmission efficiency and reduce network delay, many ISP-CP cooperation schemes are designed, parts of which are trying to introduce the idea of in-network caching. However, the combination influence of edge cache and content popularity is largely ignored in the existing solutions. Therefore, we propose a novel edge cache-based ISP-CP collaboration model for content delivery services, where the two important factors are simultaneously taken into account. Then, the model is analyzed to obtain the maximal network profit from the perspective of online and offline. The simulation results show that the profit gains of the proposed solution over the existing Internet models only considering cooperation between ISPs and CPs in the heterogeneous network environments.

INDEX TERMS Edge cache, ISP-CP collaboration, content popularity, content delivery.

I. INTRODUCTION

As more and more end users access the Internet via mobile devices, such as mobile-connected PCs and Smartphones (including phablets), mobile traffic in the Internet is explosively increasing. According to the 2017 Cisco Visual Networking Index (VNI), global mobile data traffic has grown 18-fold over the past 5 years, and will increase 7-fold in the next 5 years, reaching 49 exabytes per month by 2021 [1]. Therefore, it has been a challenging problem to address the increasing mobile traffic represented by video contents [2], [3].

In order to cope with the rapid growth of mobile traffic, both Internet Service Providers (ISPs) and content providers (CPs) deploy a massive network equipments (e.g., servers and bandwidth) with different considerations [4]–[6]. For an ISP, its focus is to solve the traffic engineering (TE) problem, i.e., modifying the routing table to achieve minimal congestion probability in the network, which makes users have lower packet loss and latency, and the network gracefully absorb flash crowds. For a CP, its aim is to solve the server selection (SS) problem, i.e., choosing which server should deliver content to each client, which makes user demand satisfied as soon as possible and realizes load balance of servers. However, the growth of demand for contents and the resulting deployment of content delivery infrastructures pose the pressure for cost reduction and customer satisfaction to ISPs and CPs, which motivates them to cooperatively work to provide better content delivery services [4].

Recently, several works have studied on this new field. Jiang *et al.* [6] propose a mathematical framework, which considers three cooperative models between ISPs and CPs. Hande *et al.* [7] design a scheme to obtain the precise trade-off of benefits distributed among ISPs, CPs, and end-users. To dynamically adapt to the traffic demand for contents hosted on CPs, Frank *et al.* [4] propose Content-aware Traffic Engineering (CaTE) by using ISP network information and end-user locations during the SS process. Moreover, given that the prominent advantages of in-network caching in content distribution [8]–[11], researchers are trying to introduce the idea of in-network caching to further improve network performance while considering the cooperative work of ISPs and CPs. Pham *et al.* [12] model and analyze the interactions on prices and caching investments among transit ISPs, access ISPs and CPs in Information-Centric Networking (ICN) interconnection. Based on extensive in-network caching ability of ICN, Arifuzzaman*et al.* [13] use game theory to analyze the decision making problem of caching contents to realize monetary compensation and revenue sharing among network players. Douros *et al.* [14] propose to deploy a cache in an ISP and make the ISP and CPs cooperate to share deployment cost and additional profit. Ndikumana *et al.* [15], [16] propose a joint incentive policy and a price-based cache replacement scheme for paid content in named data networking, which can increase the ISP's and CPs' profits. To deliver content of a CP over the "last mile" of an ISP's infrastructure, Le *et al.* [17] partition the buffer space at the BS and lease each cache slice to the service providers. Then, Mitra *et al.* [18] explore pricing and capacity decisions for bandwidths and caches, and model the ISP-CP interaction. Mangili *et al.* [19] analyze the allocation problem of unused bandwidth and available cache at the wireless access networks of ICN from the perspective of a CP's revenue.

Although parts of these excellent works exploit the thought of in-network caching to improve content transmission efficiency and reduce network delay, the combination influence of edge cache and content popularity is largely ignored, and not analyzed theoretically in the current solutions. In this paper, therefore, we propose a novel edge cache-based ISP-CP collaboration scheme for content delivery services, which can achieve the optimal network profit. The main contributions of this paper are as follows.

- We theoretically analyze the cooperative working scheme of ISPs and CPs in content delivery services by simultaneously considering the effects of content popularity and edge cache.
- We formulate the edge cache-based ISP-CP collaboration problem to be a maximal network profit model, and analyze the model from the perspective of online and offline, respectively.
- Extensive simulations are done to evaluate the performance of the proposed model in the heterogeneous network environments. The results demonstrate that the profit gains of our proposed solution performs better than the existing Internet models only considering cooperation between ISPs and CPs.

The rest of this paper is organized as follows. In Section II, the optimization problem of network profit to resolve is formulated and analyzed from the perspective of online and offline. Simulation results are presented and discussed in Section III. Finally, we conclude this study in Section IV.

TABLE 1. Notations of the key parameters.

Symbols	Notations
$\mathcal{M}$	Set of mobile clients
$\mathcal{B}$	Set of base stations
$\mathcal{K}$	Set of different network contents
$\mathcal{R}$	Set of content requests
M	the number of mobile clients
B	the number of base stations
K	different network contents
R	the total number of content requests of mobile clients
α	Zipf skewness parameter
q_i^k	Request rate for content k at mobile client i
s^k	Size of content k
$x_{i,j}^k$	Boolean variable to indicate whether mobile client i requests content k from BS j
y_j^k	Boolean variable to indicate whether BS j caches content k
$H_{j,CP}$	Shortest hops between BS j and CP
C_j	Size of a cache deployed at BS j
P_{CP}	Membership fee of each mobile client charged by CP
P_{ISP}	Average profit that a unit of bandwidth generates for ISP
C_{CP}	Average processing cost per unit of network flow, paid by CP
C_n	Average processing cost per unit of network flow, paid by the nodes along the path between a BS and CP
C_l	Average transmission cost per unit of network flow, paid by the links along the path between a BS and CP
C_0	Average cost per unit of bandwidth bought by ISP
C_{ca}	Average cost per unit of cache size
C_{re}	Average retrieval cost per request
I	Caching incentive of ISP, paid by CP

II. SYSTEM MODEL AND ANALYSIS

In this section, we formulate the maximal profit problem as an ISP-CP collaboration model for content delivery services, where both edge cache and content popularity are considered. Then, this model is analyzed theoretically to obtain the optimal and realistic solutions, respectively. Without loss of generality, only one ISP and CP scenario is considered in our model to simplify this problem. The notations of key parameters used in this paper are summarized in the Table 1.

A. CONTENT POPULARITY MODEL

Due to that network content popularity follows the Zipf distribution [20], we assume that the number of different contents provided by servers of a CP is K, whose rank is from 1 to K based on content popularity [21]. Thus, the content corresponding to the rank 1 has the highest popularity while the one labeled by the rank K is the leat popular content. Let the total number of requests be R, therefore, the number of requests about content k, R_k, can be written as

$$R_k = R\frac{k^{-\alpha}}{\sum\limits_{k=1}^{K} k^{-\alpha}}, \quad k = 1, 2, \cdots, K. \tag{1}$$

where the number of popular contents increases with the growth of the value for skewness factor α.

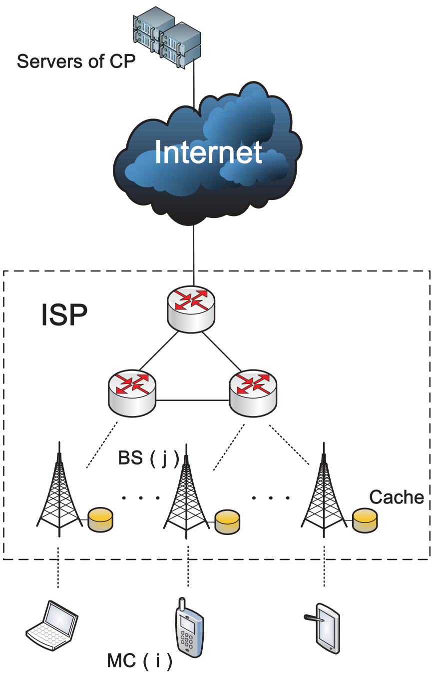


FIGURE 1. **Network model. Each BS has a caching capability, and content requests of a mobile client are satisfied either by cached contents of BSs or by the servers of CP.**

B. NETWORK MODEL

As shown in Fig. 1, caches are deployed at the edge of the ISP, e.g., base stations (BSs), which can improve network performance. Based on the rank of content popularity mentioned above, the vast majority of user requests can be satisfied by edge caches if popular contents that mobile clients (MCs) are interested in are buffered at the BSs. Although incurring additional cost for the ISP, the introduced caches can satisfy users' requests at the wireless access points, which can reduce transmission traffic and work load of network equipments, and then increase network profit.

C. PROBLEM FORMULATION AND OBJECTIVE

1) REVENUE MODEL

In our system, the revenue model of the whole network consists of two parts: membership fees of MCs charged by the CP and the traffic fees charged by the ISP. The ISP charges traffic fees to provide network support for mobile users and the CP. Due to that parts of network requests can be coped with at edge caches of the ISP, the traffic income of the ISP can be formulated as $\sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k s^k P_{ISP} + \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) P_{ISP}$, where the latter refers to the bandwidth income to deal with the unsatisfied network requests at edge caches of the ISP. In the above expression, q_i^k is the request rate for content k at mobile client i, s^k is the size of content k, P_{ISP} is the average profit that a unit of bandwidth generates for the ISP, and $x_{i,j}^k$ and y_j^k are boolean variables. $x_{i,j}^k$ takes the value of 1 if the content k that mobile user i requests is cached in BS j, and 0 otherwise. y_j^k takes the value of 1 if BS j caches content k, and 0 otherwise. Moreover, to encourage the ISP to buffer more popular contents, an incentive I must be provided by the CP to the ISP due to its caching behavior. As for the CP, its revenue $M \cdot P_{CP}$ is the sum of the membership fee P_{CP} of each mobile client. Therefore, the total network revenue model can be written as:

$$M \cdot P_{CP} + \sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k s^k P_{ISP} + \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) P_{ISP} + I. \tag{2}$$

2) COST MODEL

The total network cost is composed of expenses of the ISP and CP, which is a mixed cost for both of them. The expenditures of the CP and traditional ISP mainly depends on processing cost, bandwidth cost and retrieval cost, which are caused by network devices (e.g., routers, switches and servers), network links, and users' requests, respectively. Due to that the majority of content requests can be satisfied at the storage equipments of access networks, the introduced edge caches can reduce network transmission traffic, which reshapes the conventional cost models of the ISP and CP.

Specifically, after deploying a memory at each BS, the corresponding bandwidth input and processing cost of the ISP can be written as $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) C_0 + \sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k s^k C_0$ and $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) \left(H_{j,CP}-1\right) (C_n + C_l)$, where $H_{j,CP}$ is the shortest hop between BS j and CP, C_n and C_l are average processing and transmission cost per unit of network flow paid by the nodes along the path between a BS and CP, and C_0 is average cost per unit of bandwidth bought by ISP. Although the deployed memory can reduce network transmission traffic, it makes the ISP incur content caching cost $\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} y_j^k s^k C_{ca}$ and retrieval cost of content requests in the caches $\sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k C_{re}$, where C_{ca} is Average cost per unit of cache size. Similarly, the processing cost, bandwidth cost, incentive cost, and retrieval cost of the CP can be written as $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) C_{CP}$, $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) P_{ISP}$, I, and $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k x_{i,j}^k \left(1-y_j^k\right) C_{re}$, respectively, where C_{CP} is average processing cost per unit of network flow paid by CP, and C_{re} is average retrieval cost per request. Therefore, the cost model of our system can be written as:

$$\begin{aligned}
&\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right) \left[C_{CP} + \left(H_{j,CP}-1\right)(C_n+C_l)\right] \\
&+ \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k (1-y_j^k)(C_0+P_{ISP}) + \sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} y_j^k s^k C_{ca} \\
&+ \sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} \left(q_i^k s^k C_0 + q_i^k C_{re}\right) + \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k x_{i,j}^k \left(1-y_j^k\right) \\
&\times C_{re} + I.
\end{aligned} \tag{3}$$

3) PROFIT MODEL

In this paper, our objective is to maximize network profit of ISP-CP collaboration for content delivery by deploying caches at the BSs, which is equal to network revenue minus its cost. Therefore, the profit maximum problem can be formulated as the non-linear programming problem, which is written as (4), shown at the bottom of this page.

In the optimization problem, the first constraint forces each request of a mobile user to be assigned to only one BS or the content source to obtain the data. The second inequality ensures the maximal size of contents buffered in a BS j to be below cache capacity C_j of the BS. The third constraint represents that the variable $x_{i,j}^k$ and y_j^k can only take the value of 0 or 1.

D. MODEL ANALYSIS

In this part, we analyze the optimal solution of the proposed system under the following three assumptions, which will provide an optimal bound to evaluate future solutions able to obtain near-optimal results under different network scenarios. The first assumption is that each BS has the same cache size. The second one is that the top N contents of content popularity rank are buffered in the cache, which can achieve the highest cache hit rate. The last one is that the average hop between a BS and the CP, and the average size of each content are represented by $\overline{H}$ and $\overline{S}$, respectively.

Under these assumptions, $\sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k s^k (P_{ISP}-C_0)$, $\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} y_j^k s^k C_{ca}$ and $\sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k C_{re}$ can be rewritten as $R\cdot\overline{S}\cdot(P_{ISP}-C_0)$, $B\cdot N\cdot\overline{S}\cdot C_{ca}$, $R\cdot C_{re}$, respectively. Besides, based on the zipf model of content popularity, the number of requests unsatisfied at an edge cache can be calculated. That is to say, $\sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k x_{i,j}^k \left(1-y_j^k\right)$ can be rewritten as $R\cdot \sum_{k=N+1}^{K} k^{-\alpha}/\sum_{k=1}^{K} k^{-\alpha}$. Moreover, the second constraint $\sum_{k\in\mathcal{K}} y_j^k s^k \le C_j, \forall j\in\mathcal{B}$ also can be simplified to be $\sum_{k\in\mathcal{K}} y_j^k \le N, \forall j\in\mathcal{B}$. Therefore, the profit model (4) can be rewritten as the expression (5), shown at the bottom of this page, where B is the number of base stations in an ISP.

Recall that M, B, $\overline{H}$ are parameters related to network topology, R, α, K, N, $\overline{S}$ are parameters related to network contents and requests, and P_{CP}, P_{ISP}, C_{ca}, C_{re}, C_n, C_l, C_0 are the given network parameters, which are all with fixed values. Hence, we can come to the conclusion that network profit of ISP-CP collaboration largely relies on cache size, content popularity, and average hops of data dissemination, which will be further discussed in the simulation.

III. SIMULATION RESULTS AND DISCUSSIONS

In this section, we use computer simulations to evaluate the performance of the proposed model. We first describe the simulation settings, and then present the simulation results.

$$\begin{aligned}
\max\ & M\cdot P_{CP} + \sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k s^k (P_{ISP}-C_0) - \sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} y_j^k s^k C_{ca} - \sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}} q_i^k C_{re} - \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k x_{i,j}^k \left(1-y_j^k\right) C_{re} \\
& - \sum_{i\in\mathcal{M}}\sum_{j\in\mathcal{B}}\sum_{k\in\mathcal{K}} q_i^k s^k x_{i,j}^k \left(1-y_j^k\right)\left[C_0 + C_{CP} + \left(H_{j,CP}-1\right)(C_n + C_l)\right] \\
s.t.\ & \sum_{j\in\mathcal{B}} x_{i,j}^k \le 1, \forall i\in\mathcal{M},\quad k\in\mathcal{K} \\
& \sum_{k\in\mathcal{K}} y_j^k s^k \le C_j,\quad \forall j\in\mathcal{B} \\
& x_{i,j}^k \in \{0,1\},\quad y_j^k \in \{0,1\}, \forall i\in\mathcal{M},\ j\in\mathcal{B},\ k\in\mathcal{K}
\end{aligned} \tag{4}$$

$$\begin{aligned}
\max\ & M\cdot P_{CP} + R\cdot\overline{S}\cdot(P_{ISP}-C_0) - B\cdot N\cdot\overline{S}\cdot C_{ca} - R\cdot C_{re}\cdot\left(1 + \sum_{k=N+1}^{K} k^{-\alpha}/\sum_{k=1}^{K} k^{-\alpha}\right) \\
& - \left[C_0 + C_{CP} + \left(\overline{H}-1\right)(C_n + C_l)\right]\overline{S}\cdot R\cdot \sum_{k=N+1}^{K} k^{-\alpha}/\sum_{k=1}^{K} k^{-\alpha} \\
s.t.\ & \sum_{j\in\mathcal{B}} x_{i,j}^k \le 1,\quad \forall i\in\mathcal{M},\ k\in\mathcal{K} \\
& \sum_{k\in\mathcal{K}} y_j^k \le N,\quad \forall j\in\mathcal{B} \\
& x_{i,j}^k \in \{0,1\},\quad y_j^k \in \{0,1\}, \forall i\in\mathcal{M},\ j\in\mathcal{B},\ k\in\mathcal{K}
\end{aligned} \tag{5}$$

TABLE 2. Value information of the key simulation parameters.

Symbol/Attribute	Value/Value Range
M	12000
B	100
K	1000
R	100000
C_j	[0.1,1]
α	[0.6,1.5]
P_{CP}	198 CNY/Year
P_{ISP}	0.29 CNY/MB
C_{ca}	0.2 CNY/MB
Topology	Power-Law, Transit-Stub, Waxman
Num. of network nodes	64

A. SIMULATION SETTINGS

1) SIMULATION ENVIRONMENTS

In this part, the simulation environments are presented to evaluate the proposed model, and the key simulation parameters are summarized in the Table 2.

In the simulation, cache size of each BS is abstracted as a proportion that cache size is defined as relative size to the total number of different contents in the network. Given that the cache size is small in realistic networks [22]–[26], we evaluate the network performance of each solution when cache size varies from 0.1% to 1% [25]–[27]. Moreover, we set the range of skewness factor α in Zipf distribution to be 0.6 [28] - 1.5 [20], [21]. Besides, the values of some important variables on the network charges are chosen or set by considering actual network situations. For example, the membership fee of each mobile client charged by iQIYI is 198 China Yuan (CNY) per year, and the traffic charge of each mobile user paid to China Mobile Ltd., the world's largest carrier, is generally more than 0.29 CNY per MB. Finally, the simulation evaluation is carried out in the Power-Law topology generated by using Inet topology generator [29], Transit-Stub and Waxman topologies generated by using GT-ITM library [30], respectively. These topologies are randomly generated and includes 64 network routers in the simulation, where the average hop $\overline{H}$ varies. Take the hypothetical United States backbone network US64 [21] for example, the value of $\overline{H}$ is about 7.

2) COMPARATIVE POLICY

To demonstrate the advantages of our scheme, we compare the offline solution and online solution of the proposed model with the existing Internet model only considering cooperation between ISPs and CPs [5], [6], which are named as "Offline with Cache", "Online with Cache" and "OPT without Cache", respectively. Offline with Cache refers to the solution of the simplified profit model (5), where the most popular N contents are cached in the edge caches. Compared with Offline with Cache, Online with Cache uses the Least Recently Used (LRU) [31] as its cache replacement policy and then obtains the realistic solution. In the OPT without Cache scheme, ISPs and CPs can share their network information (e.g., network topology, traffic load of each network equipment and so on) to make optimal routing decision, which can realize load balance and minimal network delay.

Algorithm 1 Process to Solve the Online Model

```
Input: Network topologies $\{\mathcal{B}\}$, network contents $\{\mathcal{K}, s^k, \alpha\}$, mobile users $\{\mathcal{M}, \mathcal{R}\}$, network caches $\{C_j\}$, network charges $\{P_{CP}, P_{ISP}, C_{ca}, C_{re}, C_0, C_{CP}, C_n, C_l\}$
Output: $\{q_i^k\}$, $\{x_{i,j}^k\}$, $\{y_j^k\}$, $\{H_{j,CP}\}$
for each request in $\mathcal{R}$ do
    if $x_{i,j}^k = 1, y_j^k = 0$ then
        Calculate and record the shortest hops $\{H_{j,CP}\}$;
        CP transmits content $k$ to user $i$;
        if $\sum_{k\in\mathcal{K}} y_j^k s^k < C_j$ then
            BS $j$ directly caches content $k$;
        else
            Content $k$ is cached to replace the least recently used object buffered at BS $j$;
        end if
    else
        if $x_{i,j}^k = 1, y_j^k = 1$ then
            BS $j$ transmits the cached content $k$ to user $i$;
        end if
    end if
    Update cache status of BS $j$ according to LRU policy;
    Record the information about $\{q_i^k\}$, $\{x_{i,j}^k\}$, and $\{y_j^k\}$;
end for
Compute the solution of expression (4) based on the recorded statistics.
```

Algorithm 1 describes the detailed process to achieve the solution of Online with Cache in the simulation. We first input values of the parameters related to expression (4) in the given simulation environments, e.g., network topologies $\{\mathcal{B}\}$, network contents $\{\mathcal{K}, s^k, \alpha\}$, mobile users $\{\mathcal{M}, \mathcal{R}\}$, network caches $\{C_j\}$, network charges $\{P_{CP}, P_{ISP}, C_{ca}, C_{re}, C_0, C_{CP}, C_n, C_l\}$. Then, the real-time solution can be calculated by inputting the recorded parameter information, e.g., $\{q_i^k\}$, $\{x_{i,j}^k\}$, $\{y_j^k\}$, $\{H_{j,CP}\}$. In Offline with Cache, the profit maximum problem in expression (4) can be simplified to be the expression (5), where the optimal solution varies with the change of cache size N, content popularity α and average hop $\overline{H}$. Therefore, the optimal solution can be obtained by inputting the values of related variables when the simulation environments are given.

B. PERFORMANCE EVALUATION RESULTS

Fig. 2 shows network profits of the three solutions under different cache sizes when Zipf skewness parameter α is 0.8. Moreover, the simulation is carried out in the Power-Law topology. In spite of additional cost incurred by the deployed in-network caches at the BSs, due to the obviously reduced resource consumption and network traffic, the introduction

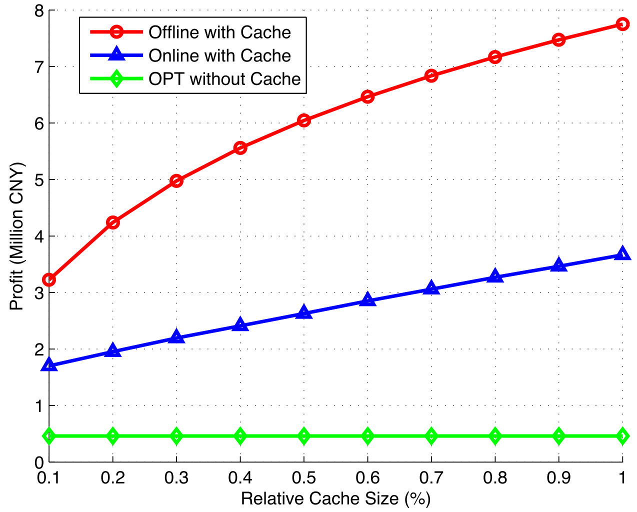


FIGURE 2. **Network profit versus cache size.**

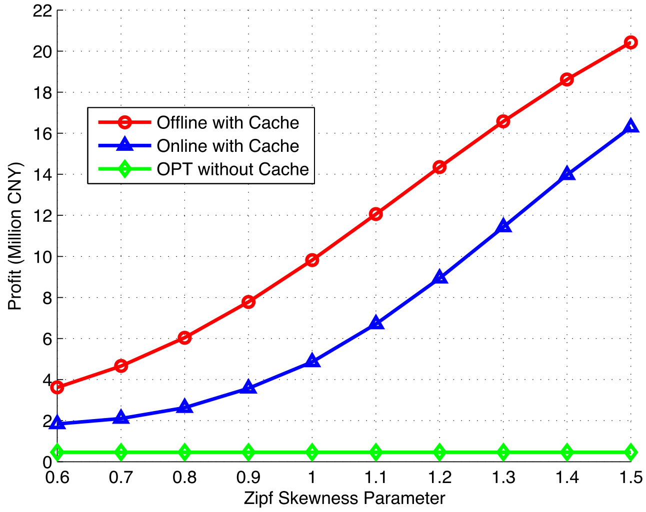


FIGURE 3. **Network profit versus content popularity.**

of edge caches can efficiently improve network performance. Moreover, network profit of the proposed model is growing as cache size increases, because a larger cache can make more popular contents cached at the edge of Internet and more requests of users satisfied by the contents buffered at the BSs. Similarly, the gap between offline and online solution of our model is becoming larger when the cache size is increasing. The reason is that the cached contents are frequently updated when edge caches adopt the LRU cache replacement policy, which can not realize the optimal caching of network contents. However, a larger buffer makes the growth of network profit decline, because the cost caused by in-networking cache increases faster than its brought benefits.

Fig. 3 shows network profits of the three solutions under different content popularities when cache size is 0.6% and simulation is conducted in the Power-Law topology. From Fig. 3, we can observe that network profit of the designed model is growing as content popularity increases. The reason is that a larger Zipf skewness parameter generates a larger proportion of content requests for the popular contents in all the data source, which makes more MCs' requests directly satisfied by the contents buffered at the edge caches and then brings a higher network profit.

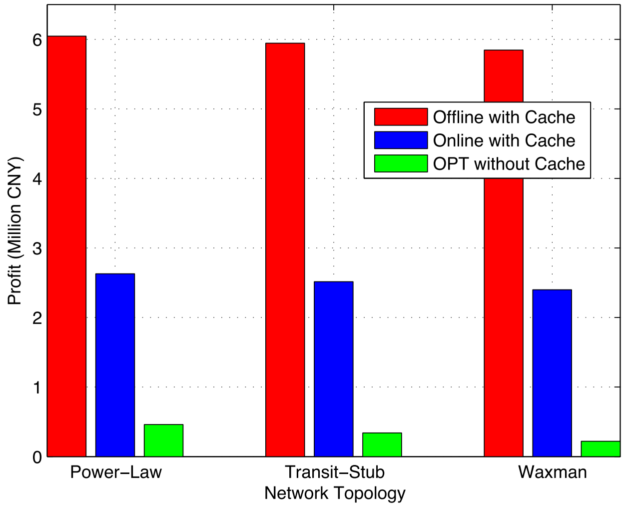


FIGURE 4. **Network profit versus network topology.**

Fig. 4 shows network profits of the three solutions under different network topologies when cache size is 0.6% and Zipf skewness parameter α is 0.8. Compared with the traditional scheme "OPT without Cache", our proposed solutions can always achieve better performance when network topology varies. Therefore, the proposed model has a universal characteristic, which can be widely applied to heterogeneous network environments.

IV. CONCLUSIONS

In this paper, we propose a novel edge cache-based ISP-CP collaboration model for content delivery services by simultaneously considering the effect of edge cache and content popularity on network profit. Moreover, the proposed model is analyzed to obtain the maximal network profit from the perspective of online and offline, respectively. Simulation results show that profit gains of the proposed scheme can achieve better performance than the existing Internet models only considering cooperation between ISPs and CPs in the heterogeneous network conditions.

The offline solution of the proposed model provides an optimal bound to evaluate the future solutions able to obtain near-optimal results under different network scenarios. In future work, we intend to develop an more efficient caching policy in the network scenario of multiple ISPs and CPs, e.g., cooperative caching and predictive caching, to achieve the online solution and increase network profits. Moreover, the access behaviors of mobile users and its related model will be fully discussed to calculate the proposed solution. Finally, the thought of profit split between ISPs and CPs will also be taken in account to improve the proposed model and realize a win-win ISP-CP collaboration.

REFERENCES

[1] *Cisco Visual Networking Index: Forecast and Methodology: 2016–2021*, Cisco, San Jose, CA, USA, Mar. 2017.

[2] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of green information-centric networking: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1455–1472, 3rd Quart., 2015.

[3] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2353–2371, 3rd Quart., 2018.

[4] B. Frank, I. Poese, G. Smaragdakis, S. Uhlig, and A. Feldmann, "Content-aware traffic engineering," in *Proc. 12th ACM SIGMETRICS/PERFORMANCE Joint Int. Conf. Meas. Modeling Comput. Syst.*, London, U.K., Jun. 2012.

[5] T. Leighton, "Improving performance on the Internet," *Commun. ACM*, vol. 52, no. 2, pp. 44–51, Feb. 2009.

[6] W. Jiang, R. Zhang-Shen, J. Rexford, and M. Chiang, "Cooperative content distribution and traffic engineering in an ISP network," in *Proc. 7th Int. Joint Conf. Meas. Modeling Comput. Syst.*, Seattle, WA, USA, Jun. 2009, pp. 239–250.

[7] P. Hande, M. Chiang, R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 990–998.

[8] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 26–36, Jul. 2012.

[9] J. Choi, J. Han, E. Cho, T. T. Kwon, and Y. Choi, "A survey on content-oriented networking for efficient content delivery," *IEEE Commun. Mag.*, vol. 49, no. 3, pp. 121–127, Mar. 2011.

[10] K. Thar, N. H. Tran, S. Ullah, T. Z. Oo, and C. S. Hong, "Online caching and cooperative forwarding in information centric networking," *IEEE Access*, vol. 6, pp. 59679–59694, Oct. 2018.

[11] M. Hajimirsadeghi, N. B. Mandayam, and A. Reznik, "Joint caching and pricing strategies for popular content in information centric networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 3, pp. 654–667, Mar. 2017.

[12] T.-M. Pham, S. Fdida, and P. Antoniadis, "Pricing in information-centric network interconnection," in *Proc. IFIP Netw.*, Brooklyn, NY, USA, May 2013, pp. 1–9.

[13] M. Arifuzzaman, K. Yu, and T. Sato, "Content distribution in information centric network: Economic incentive analysis in game theoretic approach," in *Proc. ITU Kaleidoscope Acad. Conf.*, Saint Petersburg, Russia, Jun. 2014, pp. 215–220.

[14] V. G. Douros, S. E. Elayoubi, E. Altman, and Y. Hayel, "Caching games between content providers and Internet service providers," *Perform. Eval.*, vol. 113, pp. 13–25, Aug. 2017.

[15] A. Ndikumana *et al.*, "In-network caching for paid contents in content centric networking," in *Proc. IEEE Globecom*, Singapore, Dec. 2017, pp. 1–6.

[16] A. Ndikumana, N. H. Tran, T. M. Ho, D. Niyato, Z. Han, and C. S. Hong, "Joint incentive mechanism for paid content caching and price based cache replacement policy in named data networking," *IEEE Access*, vol. 6, pp. 33702–33717, Jun. 2018.

[17] T. H. T. Le, N. H. Tran, P. L. Vo, Z. Han, M. Bennis, and C. S. Hong, "Contract-based cache partitioning and pricing mechanism in wireless network slicing," in *Proc. IEEE Globecom*, Singapore, Dec. 2017, pp. 1–6.

[18] D. Mitra, Q. Wang, and A. Hong, "Emerging Internet content and service providers' relationships: Models and analyses of engineering, business and policy impact," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, May 2017, pp. 1–9.

[19] M. Mangili, F. Martignon, S. Paris, and A. Capone, "Bandwidth and cache leasing in wireless information-centric networks: A game-theoretic study," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 679–695, Jan. 2017.

[20] K. Katsaros, G. Xylomenos, and G. C. Polyzos, "MultiCache: An incrementally deployable overlay architecture for information-centric networking," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–5.

[21] N. Choi, K. Guan, D. C. Kilper, and G. Atkinson, "In-network caching effect on optimal energy consumption in content-centric networking," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 2889–2894.

[22] Z. Li, G. Simon, and A. Gravey, "Caching policies for in-network caching," in *Proc. 21st Int. Conf. Comput. Commun. Netw. (ICCCN)*, Maui, Hawaii, Aug. 2011, pp. 1–7.

[23] D. Rossi and G. Rossini, "Caching performance of content centric networks under multi-path routing (and more)," Télécom ParisTech, Paris, France, Tech. Rep., 2011.

[24] H. Xie, G. Shi, and P. Wang, "TECC: Towards collaborative in-network caching guided by traffic engineering," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 2546–2550.

[25] J. Li, B. Liu, and H. Wu, "Energy-efficient in-network caching for content-centric networking," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 797–800, Apr. 2013.

[26] N. Laoutaris, S. Syntila, and I. Stavrakakis, "Meta algorithms for hierarchical Web caches," in *Proc. IEEE ICPCC*, Phoenix, AZ, USA, Apr. 2004, pp. 445–452.

[27] C. Fang *et al.*, "An energy-efficient distributed in-network caching scheme for green content-centric networks," *Comput. Netw.*, vol. 78, pp. 119–129, Feb. 2015.

[28] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE INFOCOM*, New York, NY, USA, Mar. 1999, pp. 126–134.

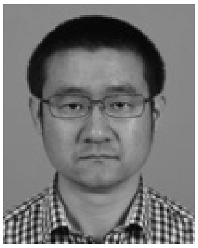
[29] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "Towards capturing representative AS-level Internet topologies," *Comput. Netw.*, vol. 44, no. 6, pp. 737–755, 2004.

[30] E. W. Zegura, K. L. Calvert, and M. J. Donahoo, "A quantitative comparison of graph-based models for Internet topology," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 770–783, Dec. 1997.

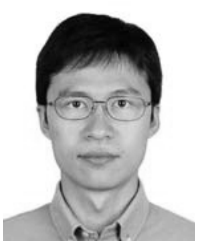
[31] Y. Kim and I. Yeom, "Performance analysis of in-network caching for content-centric networking," *Comput. Netw.*, vol. 57, no. 13, pp. 2465–2482, 2013.

CHAO FANG received the B.S. degree in information engineering from the Wuhan University of Technology, Wuhan, China, in 2009, and the Ph.D. degree from the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2015. From 2013 to 2014, he was with Carleton University, Ottawa, ON, Canada, as a Visiting Scholar. He joined the Beijing University of Technology, in 2016, where he is currently an Associate Professor. Moreover, he is also a Visiting Scholar with the University of Technology Sydney and The Hong Kong Polytechnic University. His current research interests include future network architecture design, information centric networking, software-defined networking, big data for networking, mobile edge computing, resource management, and content delivery. He has served as the Session Chair for the IEEE ICC NGN 2015 and the Poster Co-Chair for the IEEE HotICN 2018. He has also served the Technical Program Committee of the IEEE ISCIT 2017, ISCIT 2018, and CMST 2019.

HAIPENG YAO is currently an Associate Professor with the School of Information and Communication Engineering, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His main research interests include future network architecture, big data for networking, and the architecture and key technology of the new generation mobile communication systems.



ZHUWEI WANG received the B.E. degree from the School of Telecommunication Engineering, Beijing University of Posts and Telecommunications, in 2005, and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, in 2011. He is currently an Associate Professor with the Beijing University of Technology. In 2008 and 2010, he was a Visiting Scholar with the Department of Electronic and Computer Engineering, University of California at San Diego, San Diego, La Jolla, CA, USA, where he was supervised by Prof. L. B. Milstein. In 2012 and 2014, he was a Postdoctoral Fellow with the Department of Electrical Engineering, Columbia University, New York City, where he was supervised by Prof. X. Wang. His research interests include artificial intelligence, space–ground integrated networks, radar technology, and high-data rate transmission.



PENGBO SI received the B.E. degree in communications engineering and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2009, respectively. From 2007 to 2008, he was a Visiting Ph.D. Student with Carleton University, Ottawa, Canada. He joined the Beijing University of Technology, Beijing, in 2009, where he is currently a Professor. From 2014 to 2015, he was a Visiting Research Scholar with the University of Florida, Gainesville, FL, USA. He received the Best Paper Runner-Up Award from the 2015 International Conference on Wireless Algorithms, Systems, and Applications. He served as the Technical Program Committee Co-Chair for the IEEE ICCC-GMCN 2013 and the Program Vice-Chair for the IEEE GreenCom 2013. He serves as an Associate Editor for the *International Journal on AdHoc Networking Systems*, the Editorial Board Member for *Ad Hoc and Sensor Wireless Networks*, and the Guest Editor for *Advances in Mobile Cloud Computing* and the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTINGwith a Special Issue on Advances in Mobile Cloud Computing.

YIXIN CHEN received the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, in 2005. He is currently a Professor of computer science with Washington University in St. Louis. His research interests include nonlinear optimization, constrained search, planning and scheduling, data mining, and data warehousing. His work on planning has received the First Class Prize from the International Planning Competitions (2004 and 2006). He has received the Best Paper Award from ICTAI (2005) and AAAI (2010), and the Best Paper Nomination from KDD (2009). He has also received the Early Career Principal Investigator Award and a Microsoft Research New Faculty Fellowship from the Department of Energy, in 2006 and 2007, respectively. He is a Senior Member of the IEEE. He serves as an Associate Editor for the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, and *ACM Transactions on Intelligent Systems and Technology*.

XIAODONG WANG (S'98–M'98–SM'04–F'08) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA. He is currently a Professor of electrical engineering with Columbia University, New York, NY, USA. He has authored a book *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice Hall, in 2003). He has published extensively in his research areas. His research interests include in the general areas of computing, signal processing, communications, wireless communications, statistical signal processing, and genomic signal processing.

Dr. Wang received the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He is listed as an ISI Highly Cited Author. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY.

F. RICHARD YU (S'00–M'04–SM'08) received the Ph.D. degree in electrical engineering from The University of British Columbia, in 2003. From 2002 to 2006, he was with Ericsson, Lund, Sweden, and a start-up company in CA, USA. He joined Carleton University, in 2007, where he is currently a Professor. His research interests include cross-layer/cross-system design, connected vehicles, security, and green ICT. He is a registered Professional Engineer in the province of Ontario, Canada, and a Fellow of the IET. He is currently a Distinguished Lecturer, the Vice President Membership, and an Elected Member of the Board of Governors of the IEEE Vehicular Technology Society. He has received the Leadership Opportunity Fund Award from the Canada Foundation of Innovation in 2009, the Excellent Contribution Award from the IEEE/IFIP TrustCom 2010, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award), in 2011, the Carleton Research Achievement Award, in 2012, the IEEE Outstanding Leadership Award, in 2013, the IEEE Outstanding Service Award, in 2016, and the Best Paper Award from the International Conference on Networking 2005, the IEEE/IFIP TrustCom 2009, GLOBECOM 2012, ICC 2014, and the IEEE VTC 2017 in Spring. He has served as the TPC Co-Chair for numerous conferences. He serves on the Editorial Board of several journals, including as the Co-Editor-in-Chief for *Ad Hoc and Sensor Wireless Networks* and the Lead Series Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

• • •