IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 14, 2018, accepted June 11, 2018, date of publication June 18, 2018, date of current version July 6, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2848231

Joint Incentive Mechanism for Paid Content Caching and Price Based Cache Replacement Policy in Named Data Networking

ANSELME NDIKUMANA[1], NGUYEN H. TRAN[1,2], (Member, IEEE), TAI MANHO HO[1,3], DUSIT NIYATO[1,4], (Fellow, IEEE), ZHU HAN[1,5], (Fellow, IEEE), AND CHOONG SEON HONG[1], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea
[2]School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia
[3]Faculty of Information Technology, Posts and Telecommunications Institute of Technology, Ho Chi Minh 700000, Vietnam
[4]School of Computer Engineering, Nanyang Technological University, Singapore 639798
[5]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA

Corresponding author: Choong Seon Hong (e-mail: cshong@khu.ac.kr)

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the Grand Information Technology Research Center Support Program supervised by the Institute for Information and Communications Technology Promotion under Grant IITP-2018-2015-0-00742 and in part by the National Research Foundation of Korea (NRF) Grant through the Korean Government (MSIT) under Grant NRF-2017R1A2A2A05000995.

ABSTRACT Internet traffic volume is continuing to increase rapidly. Named data networking (NDN) has been introduced to support this Internet traffic growth through caching contents close to consumers. While caching in NDN is beneficial to both Internet service providers (ISPs) and content providers (CPs), ISPs serve cached contents independently without any coordination with CPs. By authorizing the ISPs to cache and distribute the contents accessible on payments, it becomes impractical for CPs to control content access and payments. In this paper, we address these challenges by proposing a joint incentive mechanism and a price-based cache replacement (PBCR) policy for paid content in NDN that improves the ISP's and CPs' profits. We use an auction theory, where the ISP earns profits from caching by alleviating traffic load on transit links and participating in contents selling. Therefore, before the ISP starts selling cached contents, it needs to cache them first. Furthermore, the ISP cache capacity is limited; therefore, we propose PBCR, where the PBCR triggers the content that needs to be replaced when the cache storage is full based on both content price and link cost. The simulation results show that our proposal increases the profits of all the network players involved in paid content caching and improves cache hit ratio.

INDEX TERMS Named data networking, paid content caching, ISP network, reverse auction.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATIONS

Since the last decade, many projects such as EU FIA, and US NSF GENI were funded to research on the definition of new future Internet architectures [1]. In the proposed future Internet architectures, Named Data Networking (NDN) is one of them that has its origin from Content Centric Networking (CCN) [2]. NDN changes the existing host-oriented Internet architectures to an information oriented architecture, where contents are requested and retrieved by their names rather than using IP addresses.

NDN uses two types of packets, namely Interest packet and Data packet. For requesting content, the Interest packet is used. On the other hand, by using the reverse path of Interest packet, the Data packet is transmitted to requester as a response to the submitted Interest packet [2]. Furthermore, for the first time requested content, the node retrieves that content from the origin server of Content Provider (CP), while the later requests of the same content can be served from the caches implemented in the intermediate routers between consumers and the origin server [2]. Caching content inside the ISP network reduces the traffic volume crossing its network and transmission delay, while also improving the Quality of Experience (QoE) for the customers (consumers of the contents) of both Internet Service Providers (ISPs) and CPs [3].

Content caching has been extensively studied for Peer-to-Peer (P2P) networks [4], [5], and Content Delivery Network (CDN) [6], [7]. However, caching in NDN/CCN is totally different from P2P/CDN caching due to the following reasons:

- P2P and CDN operate at the application layer as an overlay on TCP/IP architecture, while NDN/CCN works directly at the network layer [8].
- CDN is location dependent, where cache servers are geographically distributed all over the world. Based on the geographical location of the consumer, the requested content by the consumer will be fetched from the nearest CDN server. On the other hand, in NDN, routers in transmission paths keep the copies of the contents passing through them. In other words, the content can be cached and retrieved anywhere in the networks. Since contents can be retrieved by their names, not by IP addresses, NDN is location independent [9].
- In CDN, when the consumer's identity needs to be verified in order to guarantee its access to the requested content, the cache server needs to contact the appropriate content server of the producer/CP. In other words, CDN uses the session based security model. In NDN/CCN, each content is signed by its CP. This enables caching routers to match securely the request (Interest packet) and the content object (Data packet) without always contacting the CP's server [10].

In addition to the above highlighted differences between NDN/CCN and CDN/P2P caching, also, many NDN/CCN proposals in literature such as [11]–[13], and [14] focused on content caching. However, given that the previous literatures in both CCN and NDN have not investigated and specified how the content accessible on payment can be handled, a new incentive mechanism and cache replacement policy for paid content in NDN/CCN is needed. Furthermore, caching and selling cached content (content accessible on payment) will overcome the following issues raised in literature: i) ISP's lack of incentives to cache the contents [15], and ii) the ISP's profit loss due to content caching (specifically, for small ISPs, where these ISPs do not have customer-provider links) [12]. To the extent of our knowledge, we are the first to investigate paid content caching defined below in NDN/CCN by using auction theory.

B. WHAT IS PAID CONTENT?

Paid content is also referred to as ''pay-for-content'', it is defined as a type of content such as text, audio, graphics, animation, and video that is shared and distributed over the Internet based on payment. Paid contents are usually copyrighted, where the consumers have to pay fees for accessing or downloading them [16]. Furthermore, high quality of paid contents help the CPs to attract consumers and charge them downloading or subscription fees [17]. Even though there are many types of contents (paid and non-paid contents) that can be cached in NDN, we especially focus on paid contents available on downloading fees.

Paid content has become the main source of revenue for many CPs such as Netflix, Hulu, Amazon Prime, and Playstation Vue [18], [19]. Furthermore, paid content models can be classified into two types: (i) Subscription Video on Demand (SVOD), where subscribers need to pay monthly fees for unlimited access to a library, as with Hulu, Netflix, Amazon Prime, and Playstation Vue, and (ii) Transactional Video on Demand (TVOD), where consumers buy or rent individual titles a la carte, as with iTunes and Vudu [20].

Many newspapers (e.g., Telegraph, Financial Times, and Times) also have already started to move into the paid content business with different strategies such as: i) a mixed approach, where some contents are available on payments while others are available without payments, ii) paid access with the ability to share the content, but with limited access to search engines and links, and (iii) paid access with denied access to search engines, links, and without the ability to share the content [21].

C. CHALLENGES IN PAID CONTENT CACHING

While caching paid contents inside the ISP network increases customer satisfaction, i.e., customers of both ISP and CPs can get cached content with reduced delay. However, authorizing the ISP to cache and distribute paid content affects the profits of CPs in the following ways:

- First, the ISP manages the cached content inside its network (in cache-enabled routers) independently, i.e., without any coordination with the CPs. This results in making content access and payment monitoring more impracticable for CPs.
- Second, for CPs, the lack of access control mechanism for paid contents cached inside the ISP network may complicate the task of increasing paid content quality and diversity.
- Third, both the ISP and CP profits are determined via consumers' payments, where the customers of CPs need to pay fees to the ISP for Internet access and to the CPs for paid content access. However, between the CPs and ISP, there is no coordination and profit sharing approach for cached contents.

D. OUR CONTRIBUTIONS

In order to attract both the ISP's and CPs' attention in implementing the NDN architecture and address the above challenges, a joint incentive mechanism and price based cache replacement policy for paid content caching in NDN is proposed in this paper.

As an extended version of our earlier work published in [22], the main contributions of this work are summarized as follows:

- For paid content caching, we propose a new incentive mechanism that improves both the ISP's and CPs' profits via the use of auction theory. In our proposal, the ISP earns profits from caching by alleviating the volume of transit traffic and selling cached contents to its customers, where the contents are from multiple CPs. Our incentive mechanism encourages the ISP and CPs to be truthful when participating in the designed auction, through the Vickrey-Clarke-Groves (VCG) approach [23].

- In the designed auction, in order to prevent the bidders from bidding and delivering content sizes that are different than the content sizes needed by the ISP, Frequency-Based Content Size (FBCS) checking is proposed. FBCS, which is based on Apriori Algorithm in Data Mining [24], is used to find frequent content size over all submitted content sizes in the auction.
- Before the ISP starts selling the cached contents to its consumers, it needs to cache them first. Thus, since the ISP cache capacity is limited, we propose a Price Based Cache Replacement (PBCR) as a cache replacement policy for paid content. In PBCR, the ISP keeps the most requested and expensive contents or contents retrieved from the most expensive link (in terms of monetary price and delay) in the cache for longer periods than other contents.
- The simulation results demonstrate that our proposal increases the profits of both the ISPs and CPs and also improves the cache hit ratio up to 35.29%.

Specifically, the novelties of this paper over [22] are: *(i)* In [22], network delay was not considered in the incentive mechanism. Therefore, in order to reflect the realistic Internet peering structure [25], in this work, we have enhanced the incentive mechanism by considering new system model and including network delay in our analysis. *(ii)* From the enhanced incentive mechanism, we propose PBCR as new cache replacement policy for paid contents, where PBCR is based on network delay, the cost of transit bandwidth, and content price. *(iii)* In order to ensure smooth implementation of our proposal, in this journal version, we propose Content Controller (CC). The CC is used to run the auction model for determining paid contents that need to be cached inside the ISP network, and the payment that the ISP has to pay for each paid content cached. *(iv)* In [22], we focused only on paid content. However, in this work, even though we focus on paid content caching, we assume that the ISP may cache free contents in order to minimize the delay and save bandwidth cost. Therefore, for coordinating paid content and non-paid content caching, and joining incentive mechanism with cache replacement policy, we propose computation module that is used at the CC for computing cache freshness parameter and attaching it to each content need to be cached. Cache freshness parameter defines a time period that determines how long to keep content in cache storage. *(v)* In this work, the effectiveness of our proposal is demonstrated via simulation in realistic and mesh ISP topology.

The rest of the paper is organized as follows: We discuss related works in Section II, while Section III describes in detail our system model. We present our joint incentive mechanism and price based cache replacement policy for paid content caching in Section IV, while Section V presents our performance evaluation. We conclude the paper in Section VI.

II. RELATED WORK

In this section, we discuss some important related works, which are grouped into three categories: (i) interaction between CPs and ISPs through the use of game theory [14], [15], [26]–[28], (ii) the economic point of view of caching in ISP networks [12], [29], and (iii) cost aware caching [11], [13].

(i) Interaction between CPs and ISPs: First, Hajimirsadeghi *et al.* in [14] developed an analytical framework for a joint caching and pricing in CCN via a non-coperative game between ISPs and CPs. They showed that caching cost varies with content type and popularity. Furthermore, Kocak *et al.* [26] analyzed the effect of caching on Internet model. The authors introduced a non-cooperative game played by the CP and ISP for adjusting their prices. They considered three prices: (*i*) the price between the subscribers and the ISP, (*ii*) the prices between the subscribers and the CP, and (*iii*) the price between the CP and the ISP. A positive price between the CP and ISP, means that there is a direct payment from the CP to the ISP for bandwidth to transfer contents to the consumers. A negative price between the CP and ISP, means that the ISP caches contents and needs to pay the CP for the cached contents. The same system model was examined by Kesidis [15], where the authors analyzed the payments between the ISP and CP. They considered also a non-cooperative game in two directions between the ISP and the CP. They showed that the introduction of payment between the CP and ISP for cached content will cause the CP to increase the content price, and this results in lowering the overall demands for content. For the ISP, the authors highlighted that there is no incentive to cache the content in the Internet model. In other words, caching contents in the Internet model causes the ISP to pay the CPs for the cached contents.

Due to the complexity of the problem in the Internet model, other approaches have also been proposed. In addition to the ISP and the CP, Arifuzzaman *et al.* [27] included a CDN provider in non-cooperative game. In their proposal, the CP selects the CDN provider that can cache and distribute its content, which means that there is a direct payment between the CP and the CDN provider. Furthermore, the CDN provider makes agreements with the ISPs on behalf of the CP, where the ISPs have cache storages that are used to cache and distribute the contents to the customers. Furthermore, in [28], Douros *et al.* introduced a new business model where the ISP and CPs collaborate via sharing cache deployment cost and profit. To analyze their business model, the authors used a coalitional game, where the Shapley value was applied to split the cache deployment cost and profit.

(ii) Economic point of view of caching in ISP network: N. Kamiyama [12] investigated the effectiveness of introducing CCN in ISP networks, where the introduction of CCN in ISP networks will affect the ISP's business. The author modeled an inter-autonomous system by using a three-layer structure, where layer 2 and layer 3 have customer-to-provider links, while layer 1 has no customer-to-provider links.

Introducing CCN in layers 2 and 3 increases the ISP profits by reducing the transit traffic volume. On the other hand, introducing CCN in layer 1 reduces the ISP's profit.

Another alternative in this second category was proposed by Oueslati *et al.* [29], in which the authors highlighted the need of establishing a business model for CCN that can motivate the network players to invest in CCN implementation. They proposed a payment scheme, where Interest packets are used to buy content. The consumers pay the ISP for the contents received, and thus it is the responsibility of the ISP to return contents and provide QoS to the consumers. This results in protecting the consumers by not being penalized monetarily for requesting more contents than the ISP network can handle.

(iii) Cost aware caching: Araldo *et al.* [13] proposed cost-aware caching, where the ISP caches the most popular content objects, as well as a content object retrieved from the most expensive transit links, in terms of the monetary price. Furthermore, Ma and Towsley [11] proposed time-based caching, where the ISP monetizes its cache through negotiating the caching contract with the CP. In their proposal, the price that the CP has to pay to the ISP is based on the request rate, and the time that content is kept in the cache. The caching in both [11] and [13] was implemented in the decision policy, which is used to define the rules of either caching or not caching new incoming Data packets.

The key differences between our approach with these prior approaches are as follows: (i) In NDN, we assume that a consumer requests paid content through sending Interest packet, without any information on where to get the content. Thus, NDN is location independent, contents can be cached and retrieved from any cache-enabled routers located in different networks [30]. Therefore, for paid content, we need a payment approach, which is based on the consumer's request and cache hit. (ii) The ISP can monetize its cache storage by caching most requested contents and selling cached contents to its customers rather than selling routers' cache spaces to CPs. For both the ISPs and CPs, this approach can be an incentive to adopt the NDN architecture. (iii) The PBCR is introduced in the cache replacement policy rather than introducing it in the cache decision policy. (iv) Finally, FBCS prevents the bidders from bidding and delivering content sizes that are different than the needed content sizes.

III. SYSTEM MODEL

A. ACCESS ISP (A-ISP)

As shown in Fig. 1, we consider the A-ISP as an organization that provides Internet access services to individuals and organizations. Here, to reduce network delay and transit traffic, we consider A-ISPs as access NDN providers, where each A-ISP caches most requested contents and serves them to its consumers based on payment as a retailer of cached contents.

For the A-ISP, we model its network as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is used to denote the set of routers and $\mathcal{E}$ is used to denote the set of links. In addition, we consider that each router $R_v \in \mathcal{V}$ has a cache storage c_v for caching the contents, while each link $l \in \mathcal{E}$ has a bandwidth capacity of c_l. Furthermore, we use $\mathcal{A} \subset \mathcal{E}$ to denote a set of the internal A-ISP links (solid black lines in Fig. 1) and $\mathcal{L} \subset \mathcal{E}$ to denote a set of the transit links (dashed red line in Fig. 1), where $\mathcal{E} = \mathcal{A} \cup \mathcal{L}$.

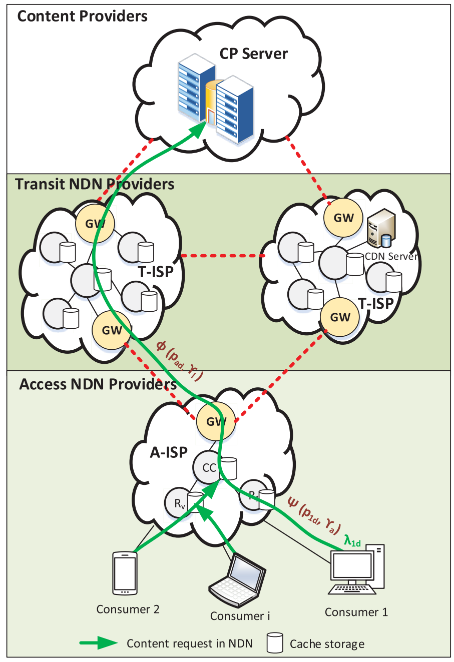


FIGURE 1. Illustration of our system model.

B. TRANSIT ISP (T-ISP)

We consider T-ISP as an Internet transit service provider sometimes called "upstream ISP" that provides Internet transit service to A-ISPs. Here, we consider that T-ISP provides NDN transit service to A-ISP. In other words, transit traffic is NDN traffic, where contents are retrieved by using their names rather than IP addresses. In addition to NDN transit traffic, T-ISPs participate in content distribution on behalf of a CPs as content distributors through caching the contents. Furthermore, T-ISP can have CDN servers in its network. We assume that the A-ISP is connected to one or more T-ISPs via transit links, where $\mathcal{N}$ is considered as a set of T-ISPs. More detail on the difference between access Internet service provider and transit Internet service provider is in [25].

C. CONTENT PROVIDER (CP)

We use CP in our system model to denote the CP's server, which is responsible for the actual distribution of contents that are not cached in both the A-ISP and T-ISP networks, where we use $\mathcal{D}$ to denote a set of contents. Furthermore, for effective caching and distributing content, and thus the CP does not have its own network, the CP can collaborate with and use T-ISP as the content distributor. However, content selling and distribution agreements between the T-ISP, CDN, and the CP are outside the scope of this paper. Hereinafter, we focus on interaction between A-ISP and T-ISPs, where CPs are represented by T-ISPs (distributors of contents from CPs).

D. CONTENT CONTROLLER (CC)

To implement our incentive mechanism, we use a Content Controller (CC), proposed and described in detail in [31], as a special functionality that can be implemented in any node inside the A-ISP network to run the auction model, which will be described in detail in Subsection IV-B. Furthermore, the CC attaches a caching parameter ω_d to each paid content $d \in \mathcal{D}$ that needs to be cached. However, any node $R_v \in \mathcal{V}$ on the transmission path can cache and replace content $d \in \mathcal{D}$ based on Price Based Cache Replacement Policy (PBCR) that use ω_d, which will be described in detail in Subsection IV-E.

E. GATEWAY (GW)

We consider the A-ISP network to be equipped with GW routers connecting its network to the external networks through the use of transit links. For any request/Interest packets that reaches the GW, GW assumes that the content is not cached inside the A-ISP network and forwards content request to the transit links.

F. CONTENT DISTRIBUTION

We consider $\mathcal{I}$ to be a set of consumers for contents $\mathcal{D}$, and λ_{id} as request for content $d \in \mathcal{D}$ from customer i of A-ISP. λ_{id} arrives at the GW router $R_v \in \mathcal{V}$ when the content $d \in \mathcal{D}$ can not be retrieved from the cache storages of A-ISP. As described in Fig. 1, for the content $d \in \mathcal{D}$ that is not available/cached in the A-ISP's network (e.g., content requested by consumer 1), the A-ISP forwards request λ_{id} to the transit links via GW routers. Any T-ISP $n \in \mathcal{N}$, which has requested content d, returns to the consumer $i \in \mathcal{I}$ the content for the free content or the price p_{id} for the content accessible based on payment.

If the ISP-A caches the paid content, it has to pay to T-ISP $n \in \mathcal{N}$ the price γ_l for transit bandwidth, and price p_{ad} for caching the paid content (the A-ISP caches the paid contents and sells them to its consumers). Furthermore, the later content requests from consumers can be served from cache-enabled routers implemented inside the A-ISP network through the use of any internal link $a \in \mathcal{A}$ (e.g., the content requested by consumers 2 and i is cached inside the A-ISP).

For each internal link $a \in \mathcal{A}$, the internal traffic volume of content $d \in \mathcal{D}$ [13] is expressed as:

$$\rho_a = \sum\nolimits_{i \in \mathcal{I}(a)} \sum\nolimits_{d \in \mathcal{D}} r_{id} \lambda_{id} h_{id}, \quad (1)$$

where $\mathcal{I}(a)$ is used to denote the set of all consumers using the link a and $\mathcal{I}(a) \subset \mathcal{I}$. Furthermore, for the content $d \in \mathcal{D}$ requested by customer $i \in \mathcal{I}(a)$, $h_{id} \in \{0, 1\}$ is used to denote the cache hit indicator. $h_{id} = 1$ if the content $d \in \mathcal{D}$ is retrieved from the cache storage of the A-ISP, and 0 otherwise. Moreover, we consider r_{id} as the size of content d requested by consumer $i \in \mathcal{I}(a)$.

TABLE 1. Key notations used in this paper.

Symbol	Definition
$\mathcal{V}$	Set of routers, $\lvert\mathcal{V}\rvert = V$
$\mathcal{E}$	Set of links, $\lvert\mathcal{E}\rvert = E$
$\mathcal{A}$	Set of internal A-ISP links, $\lvert\mathcal{A}\rvert = A$
$\mathcal{L}$	Set of exeternal A-ISP links, $\lvert\mathcal{L}\rvert = L$
c_v	Cache storage of each router $R_v \in \mathcal{V}$
p_c	Price per unit of cache storage
c_l	Capacity of external link $l \in \mathcal{L}$
$\mathcal{N}$	Set of T-ISPs, $\lvert\mathcal{N}\rvert = N$
$\mathcal{I}$	Set of users, $\lvert\mathcal{I}\rvert = I$
$\mathcal{D}$	Set of contents, $\lvert\mathcal{D}\rvert = D$
λ_{id}	Request for content $d \in \mathcal{D}$ from customer $i \in \mathcal{I}$
ρ_a	Traffic volume for each internal A-ISP link $a \in \mathcal{E}$
ρ_l	Traffic volume for each each transit link $l \in \mathcal{E}$
γ_a	Access bandwidth price for internal link $a \in \mathcal{A}$
γ_l	Access bandwidth price for transit link $l \in \mathcal{L}$
p_{ad}	Price for caching and distributing content $d \in \mathcal{D}$
h_{id}	Cache hit for content $d \in \mathcal{D}$ requested by customer i
ω_d	Cache freshness parameter
U_A	Utility of A-ISP
U_T	Utility of T-ISP
U_P	Utility of CP
r_{id}	Size of content $d \in \mathcal{D}$ requested by customer $i \in \mathcal{I}$
r_{nd}	Size of content $d \in \mathcal{D}$ that can be supplied by T-ISP n
b_{nd}	Bid of T-ISP n for content d

We define the payment from consumers to the A-ISP for access bandwidth and content as follows:

$$\Psi(p_{id}, \gamma_a) = \sum\nolimits_{a \in \mathcal{A}} \gamma_a \rho_a + \sum\nolimits_{i \in \mathcal{I}(a)} \sum\nolimits_{d \in \mathcal{D}} p_{id} \lambda_{id} h_{id}, \quad (2)$$

where γ_a is the access bandwidth fee per unit of Data and p_{id} is the price that each customer i has to pay for content d. In other words, as described in [32], we use usage pricing model, which has more advantages over flat and cap pricing models. Usage pricing model does not force light consumers to subsidize heavy consumers. It allows consumers to spend their money for premium service rather than standard or lower service on cost-serving [32].

For each external link $l \in \mathcal{L}$, the transit traffic volume is given by:

$$\rho_l = \sum\nolimits_{i \in \mathcal{I}(l)} \sum\nolimits_{d \in \mathcal{D}} r_{id} \lambda_{id} (1 - h_{id}), \quad (3)$$

where $\mathcal{I}(l) \subset \mathcal{I}$ is used to denote the set of consumers using transit link l. For content $d \in \mathcal{D}$, which is not available/cached in the A-ISP network, $1 - h_{id}$ is used to denote the cache miss indicator.

In order to reduce the network delay and alleviate the load on transit link(s), A-ISP can cache the paid contents on its routers, where we use $\lambda_d = \sum_{i \in \mathcal{I}(l)} \lambda_{id}$ to denote the total number of the requests for content d from inside the A-ISP network. Furthermore, inspired by bandwidth delay product for congestion control [33], we defined demand delay product β_d as follows:

$$\beta_d = \lambda_d (1 - h_d) VRTT_d, \quad (4)$$

where $h_d = \sum_{i \in \mathcal{I}(a)} h_{id}$ is the total cache hits inside the A-ISP network, and $VRTT_d = \frac{\sum_{i \in \mathcal{I}} RTT_{id}}{|\mathcal{I}|}$ is Virtual Round

Trip Time (VRTT) for content d. Traditional TCP/IP based Round Trip Time (RTT_{id}) cannot work properly in NDN, and this motivates us to use VRTT proposed in [1] and [33], which is considered as an average time for sending request/Interest packets for content d and receiving correspondent content d. Therefore, the amount $\Phi(p_d, \gamma_l)$ that the A-ISP needs to pay the T-ISP for Internet transit traffic $\sum_{l\in\mathcal{L}} \gamma_l \rho_l$ and for caching content is given by:

$$\Psi(p_{ad}, \gamma_l) = \begin{cases} \sum_{l\in\mathcal{L}} \gamma_l \rho_l + \sum_{i\in\mathcal{I}(l)} \sum_{d\in\mathcal{D}} p_{ad}\lambda_{id}(1-h_{id}), \\ \quad \text{if } \beta_d \geq \theta_d, \\ \sum_{l\in\mathcal{L}} \gamma_l \rho_l + \sum_{i\in\mathcal{I}(a)} \sum_{d\in\mathcal{D}} p_{id}\lambda_{id}h_{id}, \\ \quad \text{otherwise,} \end{cases} \tag{5}$$

where γ_l, for each transit link $l \in \mathcal{L}$, is used to denote the Internet transit fee per unit of Data, while $\theta_d > 0$ is a tolerance threshold (fixed by A-ISP) for demand delay product. In other words, the A-ISP caches content based on number demands and network delay. If $\beta_d \geq \theta_d$, A-ISP pays p_{ad} for caching the content d. Otherwise A-ISP doesn't cache and pay for content d, i.e., $p_{ad} = 0$ (the A-ISP only transfers its consumer payment p_{id} to the content distributor).

Both amounts $\Psi(p_{id}, \gamma_a)$ and $\Phi(p_{ad}, \gamma_l)$ are composed of two components separated by a plus sign, where the first component is related to the bandwidth fee and the second component is related to the content price. Splitting each amount into two components helps our incentive mechanism to fit into both a paid peering and a free peering model. Paid peering refers to the transit fee payment between A-ISP and T-ISP, while free peering is where A-ISP and T-ISP can exchange traffic without paying a transit fee ($\gamma_l = 0$) [12].

The A-ISP's utility $U_A(p_{id}, p_{ad}, c_v)$ depends on: (*i*) the payment $\Psi(p_{id}, \gamma_a)$ from customers that the A-ISP receives via selling cached contents and access bandwidth, (*ii*) the payment $\Phi(p_{ad}, \gamma_l)$ that the A-ISP has to pay the T-ISPs for Internet transit bandwidth and for caching content, and (*iii*) the cache deployment cost $\sum_{v\in V} c_v p_c$, where p_c is used to denote the price per unit of cache storage, while c_v is used to denote the size of cache storage at each router v. Therefore, the A-ISP's utility is expressed as follows:

$$U_A(p_{id}, p_{ad}, c_v) = \Psi(p_{id}, \gamma_a) - \Phi(p_{ad}, \gamma_l) - \sum_{v\in V} c_v(\mathbf{r}^*_{\mathbf{d}})p_c, \tag{6}$$

where $c_v(\mathbf{r}^*_{\mathbf{d}})$ is the size of cache storage needed to cache contents of size $\mathbf{r}^*_{\mathbf{d}}$. The A-ISP maximizes its utility $U_A(p_{id}, p_{ad}, c_v)$ by caching more contents, i.e., minimizing the volume of transit traffic $\Phi(p_{ad}, \gamma_l)$ and selling cached contents. Moreover, for A-ISP, reducing the volume of transit traffic requires $\sum_{v\in V} c_v(\mathbf{r}^*_{\mathbf{d}})p_c$ investment in cache storage deployment.

Both the A-ISP and T-ISPs profits are determined via consumers' payments. However, when the A-ISP sells the cached content at a high price, consumers will end up not buying the cached contents. This will affect the business of both the A-ISP and T-ISPs. To overcome the above challenge, for $\beta_d \geq \theta_d$, our incentive mechanism described below encourages the A-ISP and T-ISPs to be involved truthfully in the auction for finding the optimal content price p_{ad}. In other words, the price p_{ad} should help both the A-ISP and T-ISPs to reduce churn and retain customers. Even if caching content reduces the delay experienced by customer, but customer has to continue to receive the content d at a price p_{id} regardless of whether A-ISP caches or not caches the content d. To achieve this goal, as proposed by Kreuzer [34], we assume that the content price p_{ad} has to be less than p_{id}. In other words, p_{id} is the standard price fixed by CPs and T-ISPs that each consumer i of content d has to pay.

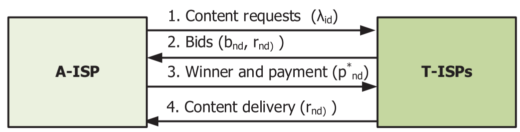


FIGURE 2. Illustration of our incentive mechanism workflow.

IV. INCENTIVE MECHANISM FOR PAID CONTENT AND PRICE BASED CACHE REPLACEMENT POLICY

In this section, we present in detail our incentive approach that improves both A-ISP's and CPs' profits via the use of auction theory in paid content caching. Subsection IV-A presents our incentive mechanism workflow, while Subsection IV-B describes in detail our auction model. Subsection IV-C discusses the problem formulation, while Subsection IV-D presents the proposed algorithms. Furthermore, as the A-ISP earns profits from caching, we will discuss PBCR as a cache replacement policy for paid content in Subsection IV-E.

A. INCENTIVE MECHANISM WORKFLOW

The workflow of our incentive mechanism for paid content caching is illustrated in Fig. 2 and summarized as follows:

- *Step 1:* For the requested paid content d by consumer $i \in \mathcal{I}$, which is not cached in the A-ISP network, the A-ISP forwards request as demand λ_{id} to all transit links.
- *Step 2:* Any T-ISP $n \in \mathcal{N}$ that has the requested paid content $d \in \mathcal{D}$ in its cache storage returns the bid (b_{nd}, r_{nd}). b_{nd} is used to denote the content price and r_{nd} is used to denote the size of content $d \in \mathcal{D}$ of each bidder $n \in \mathcal{N}$.
- *Step 3:* The A-ISP collects all the bids, evaluates the bids by using a reverse auction, and comes up with the winner, where the T-ISP $n \in \mathcal{N}$ that has lowest bid for content $d \in \mathcal{D}$ wins the auction. Then, after winner and price determination, the A-ISP informs the winner $n \in \mathcal{N}$ for delivering the content d.
- *Step 4:* For the delivered paid content d from the winner $n \in \mathcal{N}$, the A-ISP caches the content d and then forwards it to the consumer.

B. AUCTION MODEL

In our auction model, depicted in Fig. 3, we use a Reverse Auction for Paid COntent in NDN (RAPCON). In RAPCON, the roles of the buyer and the seller are reversed, where a single buyer can buy one or more products from multiple sellers. We consider two players: (1) the A-ISP as a content retailer/buyer, where the A-ISP buys the contents, caches them, and sells them to its consumers in order to satisfy its customer demands, save the transit bandwidth cost, and reduce network delay; (2) multiple T-ISPs as bidders/content distributors.

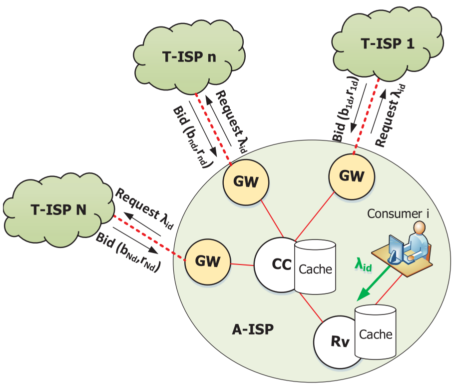


FIGURE 3. Illustration of our auction model.

RAPCON helps the A-ISP to request bids from multiple T-ISPs through flooding Interest packets, and then selects the bid that minimizes its total payment. Furthermore, the A-ISP cares about the content size, network delay, number of demands, and thus it has to buy the content that it is able to cache. In addition, the same content $d \in \mathcal{D}$ may have different sizes based on the encoding scheme utilized. In other words, T-ISP may have multiple sizes of the same content. However, the A-ISP needs to cache the most requested size by its customers. To overcome these challenges, the A-ISP collects the bids from multiple T-ISPs and examines if the content size r_{nd} in each bid (b_{nd}, r_{nd}) satisfy the most frequently requested content size r_{id} by its customers, (i.e., $r_{nd} \geq r_{id}$) and the cache capacity constraint. This condition helps the A-ISP to select a bid that meets its demand in terms of content size. However, in NDN, the content size is unknown by the A-ISP before receiving it. Thus, consumers request contents by names (without specifying the content sizes) through sending Interest packets.

To deal with the above issue, we propose the FBCS method. FBCS helps the A-ISP to find the most frequent requested content size by its customers over all the submitted content sizes by bidders via transit link(s). This helps the A-ISP to prevent each bidder $n \in \mathcal{N}$ from bidding and delivering content size which is smaller than the needed content size. FBCS is based on the Apriori Algorithm in Data Mining [24]. We choose the Apriori Algorithm due to its computational efficiency. In FBCS, based on the content sizes $\mathbf{r_d} = \{r_{nd}\}, \forall n \in N, d \in D$ submitted by bidders, the A-ISP calculates the number of occurrences of each content size r_{nd} as support count. Then, the content size with maximum support count will considered by the A-ISP as the baseline content size r_{id} mostly requested by its consumers. However, caching in NDN architecture is based on a packet level rather than a file level. Therefore, to differentiate contents and control content sizes, we consider that all Interest and Data packets belong to the same content d are characterized by a common prefix in their names [2], [35].

Definition 1 (RAPCON): In the RAPCON, the A-ISP wants to procure the content $d \in \mathcal{D}$ from T-ISP with the lowest cost. For each paid content $d \in \mathcal{D}$ requested by the A-ISP, each bidder (T-ISP) $n \in \mathcal{N}$ is allowed to submit bid b_{nd}, without knowing the bids of other bidders. Then, the A-ISP selects the bid that minimizes its total payment.

Each bidder n has its valuation function denoted by $V_n(r_{nd})$, where $V_n(r_{nd})$ is expressed as follows:

$$V_n(r_{nd}) = \begin{cases} v_{nd}, & \text{if bidder } n \text{ submit bid for content } d, \\ +\infty, & \text{otherwise,} \end{cases} \tag{7}$$

where v_{nd} is used to denote the true valuation of bidder n for content $d \in \mathcal{D}$. Therefore, content $d \in \mathcal{D}$ for which bidder n has not submitted a bid is assigned an infinity cost. Therefore, RAPCON becomes truthful bidding when $b_{nd} = v_{nd}$.

In our RAPCON, we apply the VCG approach, which guarantees an efficient and truthful outcome [23]. We choose VCG over other auction and game approaches, because VCG is computationally feasible and guarantees social-optimal solution over a set of possible outcomes. In addition, many of existing proposals in the literatures such as [15], [26], [27], and [28] focused on game theory for analyzing in-network caching.

In RAPCON, we denote v_N as the total minimum valuation for all bids with content size r^*_{nd} that satisfy the free disposal condition ($r^*_{nd} = r_{nd} \geq r_{id}$), caching condition ($\beta_d \geq \theta_d$), and customer retaining condition ($b_{nd} < p_{id}$), where v_N becomes:

$$v_N = \min_{\mathbf{r^*_d}} \sum_{n \in \mathcal{N}} b_{nd}(r^*_{nd}), \quad \forall d \in \mathcal{D}. \tag{8}$$

where $\mathbf{r^*_d} = \{r^*_{nd}\}, \forall n \in N, d \in D$.

For each bidder $n \in \mathcal{N}$, VCG computes the total minimum valuation v_{-n} without n, where v_{-n} is given by:

$$v_{-n} = \min_{\mathbf{r^{*\prime}_d}} \sum_{m \in \mathcal{N} \setminus \{n\}} b_{md}(r'_{md}), \quad \forall d \in \mathcal{D}, \tag{9}$$

where $\mathbf{r^{*\prime}_d} = \{r^*_{md}\}, \forall m \in N, d \in D$, is a vector of content sizes that satisfy the free disposal, caching, and customer retaining conditions bid by other bidders than n.

In VCG, each bidder $n \in \mathcal{N}$ should pay any harm that its participation in the auction may cause for other bidders. Furthermore, we use $\mathcal{W} \subset \mathcal{N}$ to denote the set of the winners. The winner is T-ISP j who provides contents r^*_{nd} and its bid minimizes total payment. Therefore, from (8) and (9), we calculate the price that the A-ISP has to pay to winner $n \in \mathcal{W}$ for each content d as follows:

$$p^*_{nd} = v_{-n} - \min_{\mathbf{r}^*_{\mathbf{d}}} \sum_{m \neq n} b_{md}(r^*_{md}). \tag{10}$$

In (10), bidder $n \in \mathcal{W}$ gets paid based on its social cost, i.e., bids of the other bidders if bidder $n \in \mathcal{W}$ was not participating in RAPCON, and bids of the other bidders from the chosen outcome when bidder n is participating.

Definition 2 (Bidder's Utility): Suppose bidder $n \in \mathcal{N}$ submits bid b_{nd} for content $d \in \mathcal{D}$. If bidder $n \in \mathcal{N}$ is a winner, it will be paid p^*_{nd} by the A-ISP. Otherwise, it will receive zero payment. Therefore, the utility U_n of any bidder $n \in \mathcal{N}$ for content d becomes:

$$U_n = \begin{cases} p^*_{nd} - v_{nd}, & \text{if bidder } n \in \mathcal{W}, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Definition 3 (Individual Rationality): The RAPCON is individually rational if and only if $U_n \geq 0$ holds for every bidder $n \in \mathcal{N}$.

A bidder $n \in \mathcal{N}$ will decide to participate in the RAPCON if and only if its payment p^*_{nd} is not less than its true valuation v_{nd} for content $d \in \mathcal{D}$. In other words, its utility is not negative ($p^*_{nd} \geq v_{nd}$).

Definition 4 (Truthfulness): The RAPCON is truthful if and only if each bidder $n \in \mathcal{N}$ would prefer to submit its truth valuation ($b_{nd} = v_{nd}$) rather than deviating from it (by lying). In other words, bidding ($b_{nd} = v_{nd}, r_{nd}$) maximizes each bidder's utility over any other possible bidding values ($b_{nd} \neq v_{nd}, r_{nd}$) that deviate from true valuation.

We assume that each bidder $n \in \mathcal{N}$ is strategic and selfish. Thus, it is possible for any bidder $n \in \mathcal{N}$ to submit a bid ($b_{nd} \neq v_{nd}, r_{nd}$) that deviates from its true valuation ($b_{nd} = v_{nd}$), so our goal is to ensure that the truthful bidding is a dominant strategy for every bidder $n \in \mathcal{N}$.

Theorem 1: The RAPCON is truthful.

Proof: Based on monotonicity and critical payment conditions defined in [36] as truthful bidding conditions, we approve that RAPCON is truthful by showing that RAPCON satisfies these conditions.

- *Monotonicity:* For content $d \in \mathcal{D}$, let us suppose that the bidder $n \in \mathcal{N}$ bids b'_{nd} and b_{nd}, where $b_{nd} < b'_{nd}$. In increasing order of the bids, the RAPCON chooses the winner as a bidder that has a bid that minimizes the total valuation. Furthermore, if bidder $n \in \mathcal{N}$ wins the RAPCON with bid b'_{nd}, and thus b_{nd} is less than b'_{nd}, it will also win for any bid $b_{nd} < b'_{nd}$.
- *Critical payment:* For achieving social optimal allocation, VCG pays each winner based on its social cost. Therefore, critical payment is achieved in RAPCON when bidder $n \in \mathcal{N}$ with minimum bid always wins the auction regardless of the other bidders' bid. Furthermore, we observe in (11) that no bidder will receive negative utility. Therefore, the winner's bidding value must be at least the critical value, which is equal to its payment p^*_{nd}. Let us consider bidder $n \in \mathcal{N}$ with bid b'_{nd} ($b_{nd} < b'_{nd}$) who wins the auction by using misreported valuation ($v_{nd} \neq b'_{nd}$). Hence, VCG in RAPCON pays the winner the 2nd-minimum bid value, the winner will get paid p^*_{jd} which is the same payment as for the true bid ($b_{nd} = v_{nd}$). Therefore, the bidder's utilities from these two bids b'_{nd} and b_{nd} will be the same, i.e., the bidder still wins, and its payment is still the same. In other words, submitting bid ($b'_{nd} \neq v_{nd}$) that deviates from its true valuation ($v_{nd} = b_{nd}$) will not be beneficial, and thus the bidder will not be better off. Therefore, RAPCON is truthful, where truth telling ($b_{nd} = v_{nd}$) is a dominant. □

Theorem 2: The RAPCON is individually rational.

Proof: The RAPCON becomes individually rational if each bidder $n \in \mathcal{N}$ has positive or zero utility for any choice that can be made with or without its participation in the auction. Based on the above Theorem 1 and the definition of individually rational in [36], the RAPCON chooses bidder $n \in \mathcal{N}$ that has a minimum bid as the winner. On the other hand, no matter what the other bidders bid, the loser gets paid 0 and the winner gets paid $p^*_{nd} \geq b_{nd}$, which is equivalent to $U_j \geq 0$ (11). □

C. PROBLEM FORMULATION

As mentioned in a previous subsection, the VCG mechanism guarantees an efficient outcome and truthful bidding by choosing a social optimal allocation that minimizes the social cost, i.e., the cost of all bidders. The RAPCON Social Cost Minimization (RAPCON-SCM) is formulated as follows:

$$\min_{\mathbf{x_d}} \sum_{n \in \mathcal{N}} x_{nd} v_{nd}(r_{nd}) \tag{12}$$

$$\text{subject to: } \sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} r_{nd} \leq c_l, \quad \forall\, l \in \mathcal{L}, \tag{12a}$$

$$\sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} \leq 1, \tag{12b}$$

$$\sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} r_{nd} \geq \sum_{d \in D} r_{id}, \;\; \forall i \in \mathcal{I}, \tag{12c}$$

$$\sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} v_{nd} < \sum_{i \in \mathcal{I}} \sum_{d \in \mathcal{D}} p_{id}, \tag{12d}$$

$$\sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} \beta_d \geq \sum_{d \in \mathcal{D}} \theta_d, \tag{12e}$$

$$\sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} x_{nd} r_{nd} \leq \sum_{v \in \mathcal{V}} c_v, \quad \forall v \in \mathcal{V}, \tag{12f}$$

$$x_{nd} \in \{0, 1\}, \;\; v_{nd} \in [0, +\infty), \;\; \forall n \in \mathcal{N}, \;\; d \in \mathcal{D}, \tag{12g}$$

where $\mathbf{x_d} = \{x_{nd}\}, \forall n \in N, d \in D$.

However, the true valuation v_{nd} of each bidder $n \in \mathcal{N}$ is private information that is not known by the A-ISP. Thus, we aim to design the RAPCON to choose a winning bid that minimizes the total payment that the A-ISP has to pay for the content $d \in \mathcal{D}$. To address this challenge, based on Theorem 1, we consider that each T-ISP $n \in \mathcal{N}$ bids its true valuation $v_{nd} = b_{nd}$. Therefore, the RAPCON-SCM problem (12) is transformed into the RAPCON Total Payment Minimization (RAPCON-TPM) problem, which is given as follows:

$$\min_{\mathbf{x_d}} \sum_{n \in \mathcal{N}} x_{nd} b_{nd}(r_{nd}). \tag{13}$$

The RAPCON-TPM (13) and RAPCON-SCM (12) problems have the same constants, constraints, and variables. Unless otherwise stated, hereinafter, we focus on the RAPCON-TPM problem.

1) CONSTANTS

The RAPCON-TPM problem takes a set $\mathcal{N}$ of bidders and vectors of bids $\mathbf{b_d} = \{b_{nd}\}$ and content sizes $\mathbf{r_d} = \{r_{nd}\}$, $\forall n \in N, d \in D$, where each vector takes nonnegative real values as inputs.

2) VARIABLES

We use a vector $\mathbf{x_d} = \{x_{nd}\}, \forall n \in N, d \in D$ of binary decision variables, where x_{nd} is given by:

$$x_{nd} = \begin{cases} 1, & \text{if bidder } n \in \mathcal{W}, \quad \forall d \in \mathcal{D}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

3) CONSTRAINTS

In (12a), the sum of all Internet transit traffic passing through each link $l \in \mathcal{L}$ has be less than or equal to c_l. The constraint in (12b) guarantees that each content $d \in \mathcal{D}$ is delivered by one T-ISP $n \in \mathcal{W}$. The constraint in (12c) is called free disposal condition, it ensures that each content size for bidder $n \in \mathcal{N}$ must be greater than or equal to the content size r_{id} required by the A-ISP. Furthermore, the constraint in (12d), refers to the customer retaining condition ($b_{nd} < p_{id}$). The constraint in (12e), all contents that need to be cached must allow A-ISP to alleviate transit traffic and reduce network delay. In others words, demand-delay product has to be greater than or equal to the tolerance threshold fixed by the A-ISP. The constraint in (12f), all of the delivered contents that need to be cached by the A-ISP must be less than or equal to the cache capacity.

Theorem 3: The RAPCON-TPM problem is NP-hard

Proof: The NP-hardness of the RAPCON-TPM problem is proved by reduction in the formulated conjunctive normal form [37], where we consider a special case of the RAPCON-TPM problem with a constant bid vector $\mathbf{b_d}$.

Let $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2 \cap \ldots \cap \mathcal{B}_m$ be a conjunctive normal form, where $\mathcal{B}_j$, $1 \leq j \leq m$ is a set of bids. Each $\mathcal{B}_j$ is a disjunction of j bids $\{b_{1d}\} \cup \ldots \cup \{b_{nd}\}$, where each bid has no negative value denoted v_{nd}. We turn $\mathcal{B}$ into an arithmetic form Γ as follows:

- We used an integer variable x_{nd} defined in (14) for every value v_{nd}. Therefore, by replacing $\cap$ with $*$, and $\cup$ with $+$, Γ becomes an NP-hard Integer Linear Programming (ILP), as $\Gamma \geq 1$, for $0 \leq x_{nd} \leq 1$.

□

Algorithm 1 Winner Determination

1: **Input:** $\mathcal{N}$, $\mathbf{b_d}$, $\mathbf{p_d}$, $\mathbf{r_d}$, β_d, θ_d;
2: **Output:** $\mathcal{N}$, $\mathcal{N}'$, $\mathcal{W}$, $\mathbf{x_d}$, v_N;
3: Initialization: $\mathcal{W} \leftarrow \emptyset$, $\mathbf{x_d} \leftarrow (0, \ldots, 0)$, $v_N \leftarrow 0$, $\boldsymbol{R} \leftarrow (0, \ldots, 0)$;
4: **while** $\mathbf{r_d} \neq (0, \ldots, 0)$, $\mathbf{b_d} < \mathbf{p_d}$, and $\beta_d \geq \theta_d$ **do**
5: Support count initialization for $n \in \mathcal{N}$: $count[r_{nd}] \leftarrow 0$;
6: **for all** $r_{nd} \in \mathbf{r_d}$ **do**
7: $count[r_{nd}] = count[r_{nd}] + 1$;
8: **end for**
9: $\boldsymbol{R} \leftarrow \{r_{nd}, (count[r_{nd}])\}$;
10: In $\boldsymbol{R}$, find content r_{nd} with maximum $count[r_{nd}]$; // Set r_{nd} as baseline content size $r_{id} \leftarrow r_{nd}$;
11: **for all** $b_{nd} \geq 0$, $n \in \mathcal{N}$, and $r_{nd} \geq r_{id}$ **do**
12: $\mathbf{r_d^*} \leftarrow r_{nd}$;
13: Sort $\mathbf{b_d}$ in ascending order;
14: //For $r^*_{nd} \in \mathbf{r^*_d}$, find a T-ISP $n \in \mathcal{N}$ which has the minimum bid $n = min(\mathbf{b_d})$;
15: $v_N \leftarrow v_N + b_{nd}(r_{nd})$;
16: $\mathcal{W} \leftarrow \mathcal{W} \cup \{n\}$;
17: $\mathcal{N}' \leftarrow \mathcal{N} \setminus \{n\}$; $\mathbf{r_d^{*\prime}} \leftarrow \mathbf{r^*_d} \setminus b_{nd}$;
18: $x_{nd} \leftarrow 1$;
19: $\mathbf{x_d} \leftarrow x_{nd}$;
20: **end for**
21: **end while**
22: **Return:** $\mathcal{N}$, $\mathcal{N}'$, $\mathcal{W}$, $\mathbf{x_d}$, $\mathbf{r_d^{*\prime}}$, v_N.

D. WINNER AND PRICE DETERMINATION

The formulated Integer Linear Programing (ILP) problem in (13) is NP-hard. Therefore, with the purpose of reducing the computational complexity of (13), we propose Winner Determination (WD) and Price Determination (PD) algorithms that can solve (13) in two stages. We use WD algorithm in stage 1 and PD algorithm in stage 2.

The Algorithm 1 is utilized for determining bidder $n \in \mathcal{N}$ that has a minimum bid for content delivery. Therefore, we consider a set of bidders $\mathcal{N}$, vectors of bids $\mathbf{b_d}$, standard prices $\mathbf{p_d}$, and content sizes $\mathbf{r_d}$ as the inputs of the Algorithm 1. In the Algorithm 1, the vector of content size $\mathbf{r_d}$, total valuation v_N, the set of winners $\mathcal{W}$, and the vector of content size counts $\boldsymbol{R}$ are initialized at line 3. At lines $4 - 9$, the Algorithm 1 calculates support count, i.e., the number of occurrences of each content size, and then considers at line 11 the content size with maximum support count as a baseline content size.

For the positive bids, content sizes that are greater than or equal to the baseline content size r_{id} needed by the A-ISP, and demand-delay product greater than or equal to the tolerance threshold fixed by the A-ISP, the Algorithm 1 starts winner determination process at line 11. The Algorithm 1

Algorithm 2 Price Determination

1: **Input:** $\mathcal{N}$, $\mathcal{N}'$, $\mathbf{x_d}$, $\mathcal{W}$, r_{nd}, v_N, $\mathbf{b_d}$, $\mathbf{r_d^{*'}}$, c_l, c_v;
2: **Output:** p^*_{nd};
3: Initialization: $\mathcal{W}' \leftarrow \emptyset$, $p^*_{nd} \leftarrow 0$, $v_{-n} \leftarrow (0, \dots, 0)$;
4: **for all** $b_{md} \geq 0$, $m \in \mathcal{N}'$, and $r^*_{md} \in \mathbf{r_d^{*'}}$ **do**
5: //Find T-ISP $m \in \mathcal{N}'$ which has the minimum bid $m = min(\mathbf{b_d})$;
6: $\mathcal{W}' \leftarrow \mathcal{W}' \cup \{m\}$;
7: $\boldsymbol{v_{-n}} \leftarrow \boldsymbol{v_{-n}} + b_{md}$;
8: **end for**
9: $p^*_{nd} = v_{-n} - \min\limits_{\mathbf{r_d^*}} \sum_{m \neq n} b_{md}(r^*_{md}))$;
10: **Return:** p^*_{nd}.

sorts $\mathbf{b_d}$ in ascending order at line 13 and chooses the bidder $n \in \mathcal{N}$ that has the minimum bid as the winner at line 14. Then, at lines $15 - 16$, it computes the total valuation and adds bidder n as a winner to the winner set $\mathcal{W}$. At line 17, the Algorithm 1 makes a new set $\mathcal{N}'$ of bidders, in which the winner $n \in \mathcal{W}$ is excluded in $\mathcal{N}'$. Finally, in the outputs, the algorithm comes up with the vector of decision variables $\mathbf{x_d}$ and the set $\mathcal{W}$ of winner at line 22.

The Algorithm 2 is utilized for computing the optimum price p^*_{nd}. In its inputs, we have the set of T-ISPs ($\mathcal{N}$ and $\mathcal{N}'$), the set $\mathcal{W}$ of winners, vector of decision variables $\mathbf{x_d}$, total valuation v_N, vector of bids $\mathbf{b_d}$, content sizes $\mathbf{r_d^{*'}}$, link capacity c_l, and cache capacity c_v. In the Algorithm 2, the optimal payment p^*_{nd}, the vector of total valuation $\boldsymbol{v_{-n}}$ without bidder $n \in \mathcal{W}$, and the set of winners $\mathcal{W}'$ for each bidder $m \in \mathcal{N}'$ are initialized at at line 3. Then, at lines $4 - 8$, the total valuation without winner $n \in \mathcal{W}$ is calculated. Finally, the Algorithm 2 computes the optimum price p^*_{nd} at line 9 and returns the optimum price p^*_{nd} as the output at line 10. In other words, A-ISP pays p^*_{nd} to T-ISP n for each content d that needs to be cached and/or sold, where $p_{ad} = p^*_{nd}$.

Theorem 4: The computational complexity of RAPCON is $O(n^3)$

Proof: In the Algorithm 1, we have while loop at lines $4 - 22$ that performs n iterations for checking submitted bids, where n is the size of the vector $\mathbf{b_d}$. Inside the while loop, we have the first for loop at lines $6 - 8$ for FBCS, which perform n^2 iterations to calculate the support count of each content size and attach the support count to the content size, where n is the size of the vector $\mathbf{r_d}$. Furthermore, for choosing bidder $n \in \mathcal{N}$ that has the minimum bid in all submitted bids, the second for loop (lines $11 - 20$) nested in while loop performs n iterations. As result, for the Algorithm 1, the computational complexity is $O(n^3 + n^2)$.

In the Algorithm 2, we have only one main loop at lines $4 - 8$, which runs $n - 1$ iterations to find the minimum valuation without bidder n ($\mathcal{N} \setminus \{n\}$). As a result, for the Algorithm 2, the computational complexity is $O(n - 1)$. Therefore, from both computation complexities of the Algorithm 1 and the Algorithm 2, we conclude that the computational complexity of RAPCON is $O(n^3)$. □

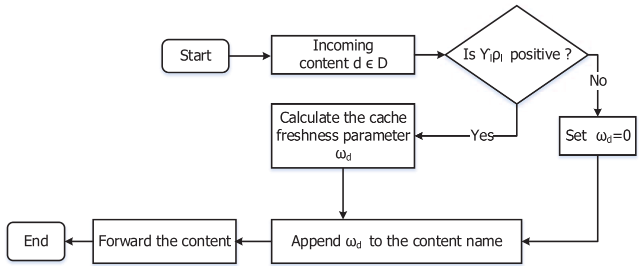


FIGURE 4. Illustration of computation module.

E. PRICE BASED CACHE REPLACEMENT POLICY (PBCR)

In PBCR, we consider that the A-ISP's network is equipped with cache-enabled routers for caching and distributing the contents. Since the A-ISP's cache capacity is limited, when the cache storage is full, to trigger the content that needs to be replaced from the cache storage, the A-ISP has to consider all the following factors:

- The number of demands/Interest packets λ_{id} for each content $d \in \mathcal{D}$ crossing the peering link $l \in \mathcal{L}$. Thus, each Interest packet is used to request one content $d \in \mathcal{D}$.
- The cost of the links in terms of $VRTT_d$, where the A-ISP keeps in cache storage for a long period of time the most requested contents retrieved from the transit links which have high delays.
- The transit link price $\gamma_l \rho_l$, in which the Data object is retrieved from.
- The price p^*_{nd} of each paid content $d \in \mathcal{D}$.

In PBCR, for the content $d \in \mathcal{D}$ requested through the paid peering link $l \in \mathcal{L}$, CC uses computation module described in Fig. 4 for computing the cache freshness parameter ω_d and attaching ω_d to each content $d \in \mathcal{D}$. ω_d defines a time period that determines how long to keep content $d \in \mathcal{D}$ in the cache, where ω_d is expressed as follows:

$$\omega_d = \begin{cases} VRTT_d(\lambda_d + \dfrac{p^*_{nd}}{\gamma_l \rho_l}), & \gamma_l \rho_l > 0 \,, l \in \mathcal{L}, \\ d \in \mathcal{D}, & n \in \mathcal{N}, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Therefore, even though we focus on paid content caching, we assume that the A-ISP may cache free contents in order to minimize the delay and save bandwidth cost. In (15), for the free content ($p^*_{nd} = 0$), the A-ISP caches the content based on $VRTT_d$ and the number of demands λ_d of content $d \in \mathcal{D}$ crossing the transit link $l \in \mathcal{L}$. We consider a content $d \in \mathcal{D}$ that has a high value of demands λ_d to be popular content, where caching that content contributes to the increase in cache hit and the reduction in transit traffic. For the free content ($p^*_{nd} = 0$) that is available on free peering link $l \in \mathcal{L}$ ($\gamma_l = 0$), there is no motivation for the A-ISP to use its caching resources for caching this content. The A-ISP and T-ISP can exchange traffic without paying any fee (by default ω_d is set to be 0). In addition, due to the content prices and link prices, the most expensive contents that are retrieved from the paid peering links stay in the cache storage for longer period than other contents. Each node $R_v \in \mathcal{V}$ caches the incoming content with its ω_d value in its Content Store (CS), where ω_d is measured in terms of milliseconds. The ω_d value is considered as a countdown timer. When ω_d value becomes 0, the node keeps content d until the cache becomes full or other content d with higher value of ω_d comes in cache. This results in making cache storage always used at the maximum.

On cache fullness, the node needs to replace some content from the CS in order to make free room in cache storage for new incoming content that needs be cached. The node uses the cache replacement module of PBCR described in Fig. 5 for replacing the contents. Therefore, during the content replacement, the node starts replacing the contents with those with a lower value of ω_d. However, for contents that have the same ω_d values, based on the access frequency counts, the Least Frequently Used (LFU) [38] cache replacement policy is utilized. LFU replaces the contents that have smallest least frequency number. Furthermore, the interoperability between PBCR and LFU is described in detail in Fig. 5.

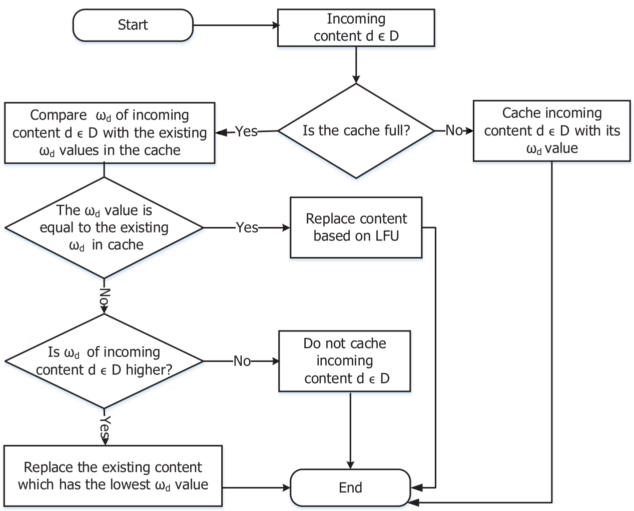


FIGURE 5. Illustration of cache replacement module.

V. SIMULATION RESULTS AND ANALYSIS

In this section, we present the performance evaluation of our proposal. During the evaluation, we use both numerical analysis and simulation. We use Julia language [39] for numerical analysis, and ndnSIM 2.1 (an ns-3 based simulator) [40], [41] for simulation.

A. BASELINE METHOD

As baseline methods, we use the following auction/game models and cache replacement policies:

1) AUCTION AND GAME MODELS

In order to analyze our proposal, and compare it with other similar schemes, we solve our ILP problem (13) by using the Gurobi optimization solver. We choose Gurobi optimization solver over other solvers, thus it can be used to solve integer and mixed-integer programming problems by using a linear programming approach called "branch and bound" [42]. Then, we compare optimal solution from Gurobi with the solutions computed through the use the VCG and coalition game described below:

- Auction model: We use the VCG mechanism in RAPCON, where the VCG approach can be applied to solve the winner determination problem once each bid is analyzed one after the other [43]. In VCG, a winner of the auction and price of the content are determined via Algorithm 1 and Algorithm 2, respectively.
- Game model: We compare the VCG mechanism in RAPCON with the caching game model proposed in [28], where the A-ISP and CPs split the cache deployment cost and profit through the use of the coalition game.

2) CACHE REPLACEMENT POLICIES

We compare our PBCR with other well-known cache placement policies [44], which are described below:

- Least Recently Used (LRU): In LRU, when the cache storage is full, in order to give room to new incoming content that needs to be cached, LRU replaces the least recently used content first.
- First-In First-Out (FIFO): FIFO replaces the first cache content. However, during content replacement, FIFO does not consider how often or how many times the content was used.
- Least Frequently Used (LFU): In LFU, the node keeps records on how many times the content is used. On cache fullness, LFU replaces the content with the smallest frequency number.
- Random Replacement: In the random cache replacement (RANDOM) policy, when the cache is full, the node randomly selects the content to replace in order to give room to new incoming content that needs to be cached.

B. EVALUATION SETUP

1) EVALUATION SETTINGS 1

In our numerical analysis, we consider three settings (one setting per each row), which are described in Table 2. In the first row, we consider that the number of customers of A-ISP varies in the range from $|\mathcal{I}| = 1,000$ to $|\mathcal{I}| = 10,000$. Furthermore, we consider that the A-ISP has cache-enabled routers $|\mathcal{V}| = 200$, and the cache size at each router $c_v = 100GB$. We use the number of bidders (T-ISPs) $|\mathcal{N}| = 50$. In the second row, the number of consumers is fixed, where $|\mathcal{I}| = 10,000$, and we vary the cache size at each

TABLE 2. Evaluation settings 1.

$\|\mathcal{I}\|$	$\|\mathcal{N}\|$	b_{nd}	r_{nd}	$\|\mathcal{V}\|$	c_v
$[1,000, 10,000]$	50	$[5, 10]$	$[5, 500]$	200	100
$10,000$	50	$[5, 10]$	$[5, 500]$	200	$[100, 1,000]$
$10,000$	50, 100	$[5, 10]$	$[5, 500]$	200	$1,000$

router from $c_v = 100GB$ to $c_v = 1,000GB$. In the last row, we fix the number of routers, cache size, and number of consumers, and we vary the number of bidders from $|\mathcal{N}| = 50$ to $|\mathcal{N}| = 100$. From the ranges given in Table 2, we generate randomly the bid b_{nd} and content size r_{nd} (in terms of GB) of each bidder $n \in \mathcal{N}$. Furthermore, for caching the contents, we assume that the average cost of each $1MB$ of cache storage is $p_c = 0.003625\$$.

We use $\gamma_a = 50\$$ per 1 $Gbps$ as monthly Internet access fee that each consumer $i \in \mathcal{I}$ has to pay to the A-ISP [12]. Moreover, for the Monthly Transit Fee (MTF) calculation, i.e., monthly transit bandwidth fee that the A-ISP has to pay to its T-ISP for each link $l \in \mathcal{L}$, we use $95th$ percentile measurement described in [25]. MTF is defined as the product of the Monthly Transit Volume (MTV) and the Transit Fee (TF) per unit of Data. TF for one month is set at $\gamma_l = 0.63\$$ per $1Mbps$ [25]. We adopt the calculation of the MTV based on a large amount of Data transmitted per second in one of two transmission directions, called uphill and downhill [12]. Furthermore, for calculating MTV, we assume that the number of days in each month is 30.

2) EVALUATION SETTINGS 2

During the simulation, we use a realistic and mesh A-ISP topology called the Rocketfuel topology [45]. The key settings of this topology are summarized in Table 3, but more details of these settings are available in [46].

TABLE 3. **Evaluation settings 2.**

Parameter	Value
$\|\mathcal{V}\|$	279 nodes
$\|\mathcal{I}\|$	30
$\|\mathcal{N}\|$	15
GW	45
CC	45
b_{nd}	$[5, 10]$
c_v	500 content objects
$\|\mathcal{D}\|$	10000 files
Popularity of Content	Follow Zipf distribution
Rank parameter (q)	$[0.6, 1.0]$
Exponent characterizing the distribution(s)	$[0.6, 1.0]$

In the Rocketfuel topology, we consider 45 GW routers, where each GW router $R_v \in \mathcal{V}$ has CC functionality. Consumers 1 to 9 request video content through use of the content name */video*, consumers 10 to 19 request music content through the content name */music*, and consumers 20 to 29 request books through the content name */book*. Furthermore, T-ISPs 1 to 5 are able to return video content to the A-ISP, T-ISPs 6 to 10 are able to return music content, and T-ISPs 11 to 15 are able to return book content. The bids b_{nd} range from \$5 to \$10 for each content d.

C. PERFORMANCE METRICS

1) NETWORK PLAYER'S PROFITS

Profit is one of our key evaluation metrics, and thus our proposal focuses on incentive mechanism and cache replacement policy that improves the profits of both A-ISP and T-ISPs/CPs. The A-ISP maximizes its profit (6) through minimizing its payment (13). When A-ISP sells content at p_{id}, its customers can receive the cached content with minimized delay. Otherwise, consumers will be experienced high delay. Moreover, receiving the contents in reduced delay can increase the customer satisfaction, and contribute to increasing the number of customers for both the A-ISP and T-ISPs/CPs. Furthermore, in case we separate the T-ISPs and the CP, the utility of a CP becomes:

$$U_p(p_{id}, p_{ad}) = \begin{cases} \sum_{i \in \mathcal{I}(a)} \sum_{d \in \mathcal{D}} p_{ad} \lambda_{id} (1 - h_{id}), \\ \text{if the A-ISP caches content}, \\ \sum_{i \in \mathcal{I}(a)} \sum_{d \in \mathcal{D}} p_{id} \lambda_{id} (1 - h_{id}), \ \text{otherwise}, \end{cases} \tag{16}$$

while the utility of T-ISP $n \in \mathcal{W}$ becomes:

$$U_T(\rho_l) = \sum_{l \in \mathcal{L}} \gamma_l \rho_l. \tag{17}$$

2) CACHE HIT RATIO

In PBCR, we trace the number of cache hits and misses. A cache hit h_d occurs when the requested content $d \in \mathcal{D}$ is cached and retrieved from the cache storage. On the other hand, a cache miss occurs when the requested content $d \in \mathcal{D}$ is not available in the cache storage. Therefore, the probability of a cache hit for content $d \in \mathcal{D}$ is expressed as follows:

$$P_{h_d} = \frac{\sum_{R_v \in \mathcal{V}} N_{R_v}^{h_d}}{\sum_{R_v \in \mathcal{V}} N_{R_v}^{h_d} + N_{R_v}^{1 - h_d}} \tag{18}$$

where P_{h_d} is the probability of cache hit for content $d \in \mathcal{D}$, $\sum_{R_v \in \mathcal{V}} N_{R_v}^{h_d}$ is the total number of cache hits, and $\sum_{R_v \in \mathcal{V}} N_{R_v}^{h_d} + N_{R_v}^{1-h_d}$ is the sum of the total number of cache hits and cache misses.

3) CONTENT POPULARITY AND ZIPF PARAMETERS

The caching performance depends on the popularity of the contents. As shown in [47], we consider that the popularity of the content follows a Zipf distribution. Therefore, in our simulation, the content requests follow a Zipf-Mandelbrot distribution with q and s as the parameters of the distribution [41]. During the evaluation, we fixed the rank parameter to be $s = 0.6$, and we varied the exponent parameter that describes the Zipf-Mandelbrot distribution from $q = 0.6$ to $q = 1.0$.

4) DELAY

We consider the total delay as time interval between the time of sending Interest packet and the time of receiving corresponding content d. In other words, consumer request for content d moves from place to place in the network until it reaches a cached-enabled node, which has requested content in its cache storage, where the node returns the requested content to the consumer. Otherwise, the content d has to be retrieved from CP server [33], [48].

D. EVALUATION RESULTS

In Figs. 6, 7, and 8, the solution of RAPCON-TPM (13) via VCG (the Algorithm 1 and the Algorithm 2) and the optimal solution computed via Gurobi optimization solver [42] are compared. VCG uses WD Algorithm 1 to choose a winner of the auction and PD Algorithm 2 to determine the price of the content. The simulation results in these Figs. show that our proposed algorithms (which use VCG) performs closely to the optimal approaches. In addition, we compare the profits from our VCG-based RAPCON with the profits from the game model (denoted CGCPISP) proposed in [28].

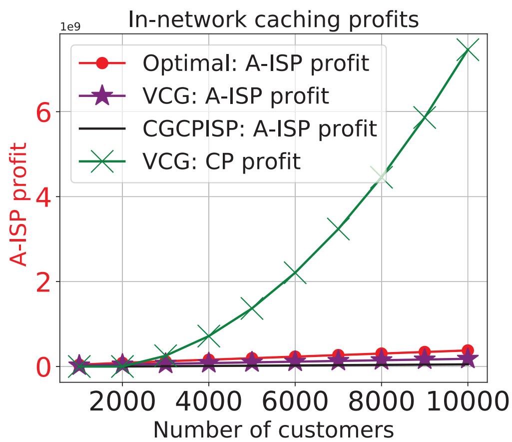


FIGURE 6. Monthly A-ISP profit vis-a-vis the CP profit.

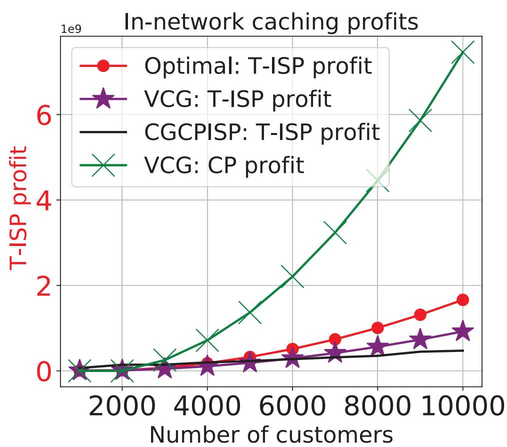


FIGURE 7. Monthly T-ISP profit vis-a-vis the CP profit.

In Fig. 6, the monthly profit of the A-ISP vis-Ã -vis the CP profit is shown, where the number of customers per day is the range from $1,000$ to $10,000$. We assume that customers of A-ISP pay both monthly Internet access fees and content fees (for downloaded paid contents). The A-ISP is equipped with cache-enabled routers for caching the contents, where the A-ISP caches the contents in its cache-enabled routers based on demands from its customers. In other words, A-ISP caching approach does not use content prefetching. This results in linearly increasing the profit of the A-ISP in our proposal. However, in the A-ISP network, without considering content selling and customer-to-provider links, caching more non-paid contents does not remarkably increase the A-ISP's profits for CGCPISP. This issue of reducing A-ISP's profits due to caching (without selling cached content) was also highlighted in [12] and [15].

Fig. 7 shows the monthly profit of the T-ISP vis-Ã -vis the CP profit. The T-ISP, as a content distributor, is able to serve the contents on behalf of CPs, i.e., the T-ISP caching is based on content prefetching. In our proposal, when the number of customers is less than or equal to $T = 3000$, the profit of T-ISP $n \in \mathcal{W}$ does not considerably increase, and thus the cache deployment cost needs to be compensated by transit bandwidth fees. However, once $T \geq 3000$, the T-ISP's profit starts to increase. Furthermore, T-ISP's profit from CGCPISP is lower than the profits from optimal and VCG, because CGCPISP does not consider paid content distribution in T-ISP networks.

Fig. 8 shows the profit of CP vis-à-vis T-ISP profit. For CP, the increase in profit is due to the payment that CP gets from T-ISP $n \in \mathcal{W}$ and the A-ISP for each content d sold. In other words, the A-ISP pays to T-ISP $n \in \mathcal{W}$, while T-ISP $n \in \mathcal{W}$ pays the CP. However, the CP needs to have central repository of its contents, in which involves storage cost. When the number of consumers of contents is less than or equal to $T = 2000$, the profit of CP does not increase, and thus the cost of central repository deployment needs to be compensated by content prices. Once $T \geq 2000$, the CP's profit considerably starts to increase. Furthermore, the CGCPISP does not consider content selling inside the ISP or T-ISP networks, each CP is in charge of selling its contents.

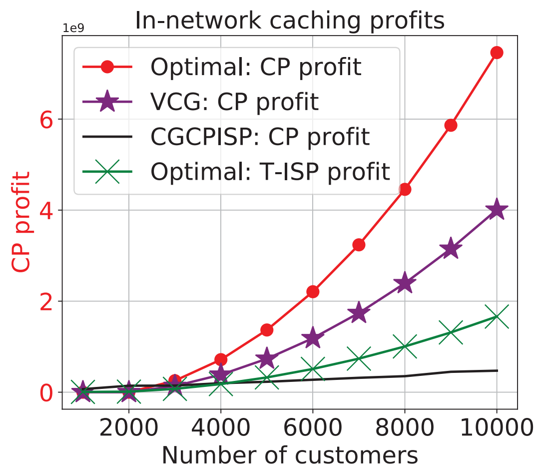


FIGURE 8. Monthly CP profit vis-a-vis the T-ISP profit.

Figs. 9, 10, and 11 show the performance evaluation of PBCR, where the A-ISP caches the most expensive contents or contents retrieved from the most expensive links, keeping them in the cache storage for a longer period than the others. The A-ISP caches each content $d \in \mathcal{D}$ with its ω_d value, and in the state of cache fullness, the A-ISP replaces

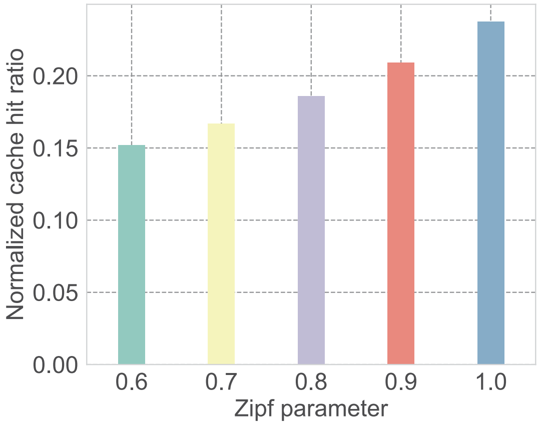


FIGURE 9. **Normalized cache hits for PBCR.**

the content based on the cache freshness parameter defined in (15).

Fig. 9 shows a normalized cache hit, while Fig. 10 shows Zipf parameter vis-a-vis the total number of packets sent over transit links in PBCR. In both figures, the Zipf exponent parameter q varies between 0.6 and 1.0. During the simulation, each node $R_v \in \mathcal{V}$ has a storage capacity c_v of caching 500 files, while the content catalog is 10, 000 files. Based on the cache capacity of each node, the total content catalog, and the Zipf parameter, the normalized cache hit ratio varies from 0.15 to 0.23, while the total number of packets range from 797810 to 1004267. Furthermore, when the Zipf exponent parameter q increases, the content becomes more popular. This results in caching more contents and reducing the number of packets needed to be sent over the transit links.

In Fig. 10, when $q = 0.6$, the nodes send more packets over the transit links, while for $q = 1.0$, the contents become more popular and the nodes keep the most requested and expensive contents in the cache storage for a longer period than the others. This increases the cache hits (Fig. 9) and reduces the number of packets needed to be sent over the transit links (Fig. 10).

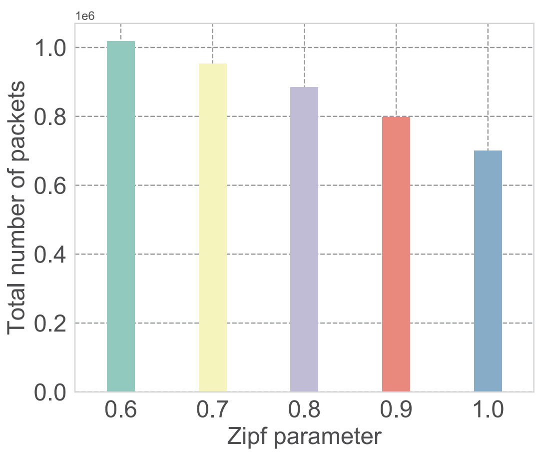


FIGURE 10. **Zipf parameter vis-a-vis the total number of packets sent over transit links.**

The simulation results in Fig. 11 shows a comparison between PBCR and other five well-known cache replacement policies, namely LFU, LRU, RANDOM, and FIFO [44] as implemented in [41]. PBCR yields a better performance than those of the other policies in terms of cache hits. This is due to the fact that when the cache is not full, the nodes cache all most requested contents and contents retrieved from most expensive links passing through them. This results in making cache always used at the maximum. However, when the cache is full, the nodes replace contents based on the ω_d parameter. Furthermore, in all five cache replacement policies compared to PBCR, they treat all contents in a coupled manner, where the link and content prices are not considered during the cache decision and cache replacement processes.

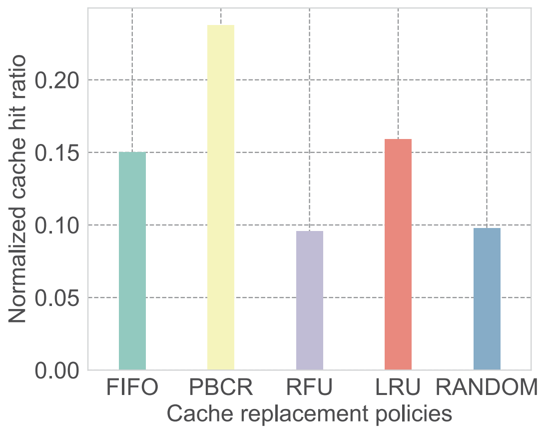


FIGURE 11. **Comparison of cache replacement policies.**

The proposed PBCR has computational module, which may cause network delay during the computation of cache freshness parameter ω_d. Therefore, we analyze the effect of PBCR on network delay defines in Section V-C.4 and compare the PBCR with LFU, LRU, RANDOM, and FIFO. In Fig. 12, the horizontal black line in the box represents median. The simulation results in this figure show that with PBCR consumers are experienced lower delay than

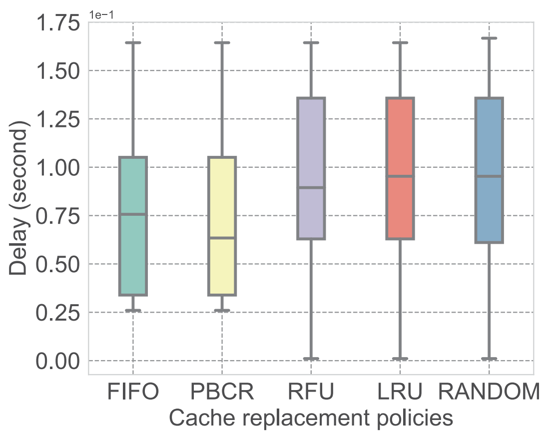


FIGURE 12. **The effect of PBCR over total network delay.**

other cache replacement policies. In other words, computation of cache freshness parameter ω_d does not increase the network delay. Furthermore, the good performance of PBCR is due to the fact that it allows cache storage to be always used at the maximum.

VI. CONCLUSION

The ability to cache the content in routers is one of the NDN features that can help the ISPs to reduce the network traffic crossing their networks and increase their profits, where the bandwidth usage costs will be reduced. For both the ISPs and CPs, caching improves their customer satisfaction, where customers can retrieve cached contents with minimized delays. However, ISPs serve cached contents independently without any coordination with CPs. Therefore, for CPs, by authorizing the ISPs to cache the paid contents, it may complicate the tasks of controlling content utilization and payments. As a solution, in this paper, we proposed an incentive mechanism, where the ISP generates revenue from caching by selling cached contents to its customers. In addition, we proposed PBCR as a cache replacement policy for paid content, where PBCR triggers the content needs to be replaced when the cache storage is full. The effectiveness of our proposal is demonstrated via simulation, where our proposal increases the cache hit ratio and the profits of both ISPs and CPs.

REFERENCES

[1] G. Carofiglio, M. Gallo, L. Muscariello, and D. Perino, "Modeling data transfer in content-centric networking," in *Proc. 23rd Int. Teletraffic Congr.*, San Francisco, CA, USA, Sep. 2011, pp. 111–118.
[2] L. Zhang *et al.*, "Named data networking," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 66–73, Jul. 2014.
[3] S. Ullah, K. Thar, and C. S. Hong, "Management of scalable video streaming in information centric networking," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21519–21546, 2017.
[4] G. Gao, R. Li, H. He, and Z. Xu, "Distributed caching in unstructured peer-to-peer file sharing networks," *Comput. Elect. Eng.*, vol. 40, no. 2, pp. 688–703, 2014.
[5] N. Kumar and J.-H. Lee, "Peer-to-peer cooperative caching for data dissemination in urban vehicular communications," *IEEE Syst. J.*, vol. 8, no. 4, pp. 1136–1144, Dec. 2014.
[6] M. Z. Shafiq, A. X. Liu, and A. R. Khakpour, "Revisiting caching in content delivery networks," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 567–568, 2014.
[7] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking: A survey," *Comput. Netw.*, vol. 57, no. 16, pp. 3128–3141, 2013.
[8] NDN-NSF. *NDN Frequently Asked Questions (FAQ)*. Accessed: Apr. 9, 2018. [Online]. Available: https://named-data.net/project/faq/
[9] M. Mangili, F. Martignon, and A. Capone, "A comparative study of content-centric and content-distribution networks: Performance and bounds," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 1403–1409.
[10] Y. Yu, "Public key management in named data networking," NDN, Univ. California, Los Angeles, Los Angeles, CA, USA, Tech. Rep. NDN-0029, 2015.
[11] R. T. Ma and D. Towsley, "Cashing in on caching: On-demand contract design with linear pricing," in *Proc. 11th ACM Conf. Emerg. Netw. Exp. Technol.*, Heidelberg, Germany, Dec. 2015, Art. no. 8.
[12] N. Kamiyama, "Analyzing impact of introducing CCN on profit of ISPs," *IEEE Trans. Netw. Service Manag.*, vol. 12, no. 2, pp. 176–187, Jun. 2015.
[13] A. Araldo, D. Rossi, and F. Martignon, "Cost-aware caching: Caching more (costly items) for less (ISPs operational expenditures)," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 5, pp. 1316–1330, May 2016.
[14] 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, Feb. 2017.
[15] G. Kesidis, "Side-payment profitability under convex demand-response modeling congestion-sensitive applications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 2785–2789.
[16] P. A. Bruck, A. Buchholz, Z. Karssen, and A. Zerfass, *E-Content: Technologies and Perspectives for the European Market*. Berlin, Germany: Springer, Jan. 2006.
[17] Z. Li and Y. Cheng, "From free to fee: Exploring the antecedents of consumer intention to switch to paid online content," *J. Electron. Commerce Res.*, vol. 15, no. 4, p. 281, 2014.
[18] Tomsguide. *Best Streaming Video Services 2018*. Accessed: Apr. 9, 2018. [Online]. Available: https://www.tomsguide.com/us/best-streaming-video-services,review-2625.html
[19] Digital Trends. *Battle of the Streaming Giants: Netflix vs. Hulu vs. Amazon Prime*. Accessed: Apr. 9, 2018. [Online]. Available: https://www.digitaltrends.com/home-theater/best-on-demand-streaming-services/
[20] Brightcovey. *From ADS to Paid Content: Monetizing Online Video*. Accessed: Apr. 9, 2018. [Online]. Available: https://www.brightcove.com/en/blog/2013/06/new-whitepaper-ads-paid-content-monetizing-online-video
[21] R. G. Picard, "New approaches to paid digital content," Reuters Inst. Study Journalism, Oxford Univ., Oxford, U.K., Tech. Rep., May 2014, p. 80. [Online]. Available: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Reuters%2520Institute%2520Digital%2520News%2520Report%25202014.pdf
[22] A. Ndikumana *et al.*, "In-network caching for paid contents in content centric networking," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
[23] M. B. Caminati, M. Kerber, C. Lange, and C. Rowat, "Vickrey–Clarke–Groves (VCG) auctions," College Inf. Sci. Technol., Pennsylvania State Univ., State College, PA, USA, May 2015. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.732.1804&rep=rep1&type=pdf
[24] Y. Ye and C.-C. Chiang, "A parallel apriori algorithm for frequent itemsets mining," in *Proc. 4th IEEE Int. Conf. Softw. Eng. Res., Manage. Appl.*, Seattle, WA, USA, Aug. 2006, pp. 87–94.
[25] W. B. Norton, *The 2014 Internet Peering Playbook: Connecting to the Core of the Internet*. Polo Alto, CA, USA: DrPeering Press, 2014.
[26] F. Kocak, G. Kesidis, T.-M. Pham, and S. Fdida, "The effect of caching on a model of content and access provider revenues in information-centric networks," in *Proc. IEEE Int. Conf. Soc. Comput. (SocialCom)*, Washington, DC, USA, Sep. 2013, pp. 45–50.
[27] M. Arifuzzaman, K. Yu, and T. Sato, "Content distribution in information centric network: Economic incentive analysis in game theoretic approach," in *Proc. Living Converged World-Impossible Standards, ITU Kaleidoscope Academic Conf.*, St. Petersburg, Russia, Jun. 2014, pp. 215–220.
[28] 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.
[29] S. Oueslati, J. Roberts, and N. Sbihi, "Flow-aware traffic control for a content-centric network," in *Proc. IEEE Int. Conf. Comput. Commun.*, Orlando, FL, USA, Mar. 2012, pp. 2417–2425.
[30] A. Ndikumana *et al.*, "Network-assisted congestion control for information centric networking," in *Proc. 17th IEEE Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Busan, South Korea, Aug. 2015, pp. 464–467.
[31] E. Aubry, T. Silverston, and I. Chrisment, "SRSC: SDN-based routing scheme for CCN," in *Proc. 1st IEEE Conf. Netw. Softwarization (NetSoft)*, London, U.K., Apr. 2015, pp. 1–5.
[32] M. Chiang, *Networked Life: 20 Questions and Answers*. Cambridge, U.K.: Cambridge Univ. Press, Sep. 2012.
[33] A. Ndikumana, S. Ullah, K. Thar, N. H. Tran, B. J. Park, and C. S. Hong, "Novel cooperative and fully-distributed congestion control mechanism for content centric networking," *IEEE Access*, vol. 5, pp. 27691–27706, 2017.
[34] E. Kreuzer, "System and method for content distribution and reselling," U.S. Patent 10 219 958, Aug. 16, 2002.
[35] A. Ndikumana, S. Ullah, and C. S. Hong, "Scalable aggregation-based packet forwarding in content centric networking," in *Proc. 18th IEEE Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Kanazawa, Japan, Oct. 2016, pp. 1–4.
[36] L. Blumrosen and N. Nisan, "Combinatorial auctions," School Comput. Sci. Eng., Hebrew Univ. Jerusalem, Jerusalem, Israel, May 2015.

[37] V. Gurvich and L. Khachiyan, "On generating the irredundant conjunctive and disjunctive normal forms of monotone Boolean functions," *Discrete Appl. Math.*, vols. 96–97, pp. 363–373, Oct. 1999.

[38] D. Lee *et al.*, "LRFU: A spectrum of policies that subsumes the least recently used and least frequently used policies," *IEEE Trans. Comput.*, vol. 50, no. 12, pp. 1352–1361, Dec. 2001.

[39] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Rev.*, vol. 59, no. 1, pp. 65–98, 2017.

[40] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," Univ. California, Los Angeles, Los Angeles, CA, USA, Tech. Rep. NDN-0005, 2012.

[41] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2.0: A new version of the NDN simulator for NS-3," Univ. California, Los Angeles, Los Angeles, CA, USA, Tech. Rep. NDN-0028, 2015.

[42] G. Solver. *Gurobi Optimization*. Accessed: Apr. 9, 2018. [Online]. Available: http://www.gurobi.com/

[43] D. Lehmann, R. Müller, and T. Sandholm, "The winner determination problem," in *Combinatorial Auctions*. Pittsburgh, PA, USA: Carnegie Mellon Univ., 2006, pp. 297–318. [Online]. Available: http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/sandholm/winner-determination-final.pdf

[44] M. Garetto, E. Leonardi, and V. Martina, "A unified approach to the performance analysis of caching systems," *ACM Trans. Model. Perform. Eval. Comput. Syst.*, vol. 1, no. 3, p. 12, 2016.

[45] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP topologies with Rocketfuel," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 4, pp. 133–145, 2002.

[46] Anselme. *The Settings for Rocketfuel Topology*. Accessed: Apr. 9, 2018. [Online]. Available: https://github.com/anselme333/PaidContentCaching/blob/master/examples/topologies/Simulation_Topology.txt

[47] K. Suksomboon *et al.*, "PopCache: Cache more or less based on content popularity for information-centric networking," in *Proc. 38th IEEE Conf. Local Comput. Netw. (LCN)*, Sydney, NSW, Australia, Oct. 2013, pp. 236–243.

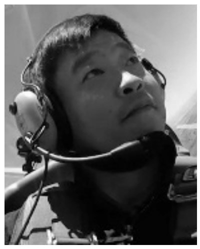
[48] A. Ndikumana *et al.* (Mar. 2018). "Joint communication, computation, caching, and control in big data multi-access edge computing." [Online]. Available: https://arxiv.org/abs/1803.11512

ANSELME NDIKUMANA received the B.S. degree in computer science from the National University of Rwanda in 2007. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Kyung Hee University, South Korea. His professional experience includes a Chief Information Officer, a System Analyst, and a Database Administrator with the Rwanda Utilities Regulatory Authority from 2008 to 2014. His research interests include congestion control, multi-access edge computing, and in-Network caching.

NGUYEN H. TRAN (S'10–M'11) received the B.S. degree from the Ho Chi Minh City University of Technology in 2005 and the Ph.D. degree in electrical and computer engineering from Kyung Hee University in 2011. He was an Assistant Professor with the Department of Computer Science and Engineering, Kyung Hee University, from 2012 to 2017. Since 2018, he has been with the School of Information Technologies, The University of Sydney, where he is currently a Senior Lecturer. His research interest is applying analytic techniques of optimization, game theory, and machine learning to cutting-edge applications, such as cloud and mobile-edge computing, datacenters, resource allocation for 5G networks, and Internet of Things. He received the best KHU Thesis Award in Engineering in 2011 and several best paper awards, including IEEE ICC 2016, APNOMS 2016, and IEEE ICCS 2016. He receives the Korea NRF Funding for Basic Science and Research from 2016 to 2023. He has been an Editor of the IEEE Transactions on Green Communications and Networking since 2016.

TAI MANH HO received the B.Eng. and M.S. degrees from the Hanoi University of Science and Technology, Vietnam, in 2006 and 2008, respectively, and the Ph.D. degree from Kyung Hee University, South Korea, in 2018, all in computer engineering. His research interests include radio resource management for wireless communication systems with special emphasis on heterogeneous networks.

DUSIT NIYATO (M'09–SM'15–F'17) received the B.Eng. degree from the King Mongkut's Institute of Technology Ladkrabang, Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include energy harvesting for wireless communication, Internet of Things, and sensor networks.



ZHU HAN (S'01–M'04–SM'09–F'14) received the B.S. degree in electronic engineering from Tsinghua University in 1997 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland at College Park, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was a Research and Development Engineer with JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department and with the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He received the NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the Best Paper Award for the *EURASIP Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of communications systems (the Best Paper Award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. He is an IEEE Communications Society Distinguished Lecturer. He is a 1% Highly Cited Researcher 2017 according to the Web of Science.

CHOONG SEON HONG received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Japan, in 1997. In 1988, he joined KT Corporation, where he was involved in broadband networks, as a Member of Technical Staff. Since 1993, he has been with Keio University. He was with the Telecommunications Network Laboratory, KT Corporation, as a Senior Member of Technical Staff and as the Director of the Networking Research Team until 1999. Since 1999, he has been a Professor with the Department of Computer Engineering, Kyung Hee University. His research interests include future Internet, ad hoc networks, network management, and network security. He is a member of the ACM, IEICE, IPSJ, KIISE, KICS, KIPS, and OSIA. He has served as a General Chair, a TPC Chair/Member, and an Organizing Committee Member for international conferences, such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, TINA, SAINT, and ICOIN. He is currently an Associate Editor of the IEEE Transactions on Network and Service Management, the *International Journal of Network Management*, and the IEEE Journal of Communications and Networks, and an Associate Technical Editor of *IEEE Communications Magazine*.