This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2021.3069763, IEEE Transactions on Wireless Communications

Data-Driven Caching with Users' Content Preference Privacy in Information-Centric Networks

Xinyue Zhang, *Student Member, IEEE*, Hongning Li, *Member, IEEE*, Jingyi Wang, *Member, IEEE*, Yuanxiong Guo, *Senior Member, IEEE*, Qingqi Pei, *Senior Member, IEEE*, Pan Li, *Senior Member, IEEE*, and Miao Pan, *Senior Member, IEEE*

***Abstract*—Information-centric networking (ICN) as an emerging networking paradigm has recently gained significant attention, due to the improvement of content delivery efficiency. The built-in network storage for caching is a key component in ICN to provide low latency service and reduce high backhaul traffic by caching popular content. However, users' content preference contains individual sensitive characteristics which is distinguishable from others. Therefore, in this work, we propose a data-driven caching revenue maximization problem with the considerations of users' local differential privacy. Specifically, we employ dBitFlip, a local differential privacy (LDP) mechanism, to locally add differential private noise to the users' preference content information. We leverage data-driven approach to predict the content popularity based on the reference distribution constructed by the reported noisy preference content data from users, mathematically present the distance between the noisy reference distribution and the true distribution by the tolerance level, and prove the relationship among the tolerance level, differential privacy budget and the confidence level. We provide feasible solutions to the proposed revenue maximization problem, and conduct simulations to show the effectiveness of the proposed scheme.**

***Index Terms*—Caching, local differential privacy, information-centric networks, data-driven optimization**

I. Introduction

As the rapid increasing of content demands in the Internet, new information-centric networking (ICN) design is motivated to be developed in the future Internet for improved delivery efficiency, content scalability and availability [1]–[3]. In addition, ICN architectures are based on named content, which is radically different from the traditional host-centric paradigm based on named hosts [1]. In this new ICN architecture, with the deployment of in-network storage for caching in the access points (AP), it is efficient to offload the tremendous increasing amount of content. In 2017, Cisco highlights that the video traffic has already reached 73 percent of all the Internet traffic in 2016, and it is estimated to be increased to 82 percent by 2021 [4]. Inspired by the fact of the speedy growth of the demand for video content, build-in caching features are supposed to be applied widely in the ICN.

Since the content provider (CP) aims to provide high quality of service (QoS) to the users, the storage for caching in APs plays an important role in reducing the network congestion and backhaul load. As the cache-enabled APs such as base stations are required to cooperate with the CP, it is necessary to find an approach to offering the economic incentives for the contributions and efficiently allocating the resources. As a result, the CP is able to cache the popular data objects with the cooperation of the cache-enabled APs by offering appreciable economic incentives [5]. For example, in [6], the authors exploit the auction theory to design the optimal allocation with jointly leasing the cache storage and bandwidth of APs. In [7], the proposed scheme focuses on optimal virtual resource allocation with integrating device-to-device communication in the ICN. However, in these works, they all use the Zipf discrete distribution [8] to represent the content popularity in Internet. In addition, content popularity may also vary over time [9]. In most cases, the true distribution is actually unknown and we can only access to a set of historical data [10]. Moreover, although the Zipf law can fit the frequency distribution, there is always distance between the the Zipfian distribution and real distribution [11]. Therefore, in our work, we employ data-driven methodology to predict the content popularity from the collected data of local CP users without premise on the content popularity distribution. Such data-driven prediction can facilitate advanced content caching schemes, such as [12], and provide risk-averse decision making under uncertainty.

When deciding the caching strategy, CP will utilize the users' content preferences to provide high QoS, it may compromise the users' privacy. To predict the content popularity, the CP aggregates the preferred content information from certain users. However, this aggregation process may elevate risks of privacy leakage [13]. As the user's content preferences may include some sensitive information, these kind of

X. Zhang and M. Pan are with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204. Email: xzhang67@uh.edu, mpan2@uh.edu. H. Li is with School of Cyber Engineering, Xidian University, Xi'an, China 710071. Email: lhn314@163.com. J. Wang is with the Department of computer science, San Francisco State University, San Francisco, CA 94132. Email: jingyiwang@sfsu.edu. Y. Guo is with the Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249. Email: yuanxiong.guo@utsa.edu. Q. Pei is with the School of Telecommunications Engineering, Xidian University, Xi'an, China 710071. Email: qqpei@mail.xidian.edu.cn. P. Li is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106. Email: lipan@case.edu.

The work of X. Zhang and M. Pan is partly supported by the National Science Foundation CNS-2029569 and CNS-1801925. This work of H. Li and Q. Pei is supported by the National Key Research and Development Program of China under Grant 2020YFB1807500, the Key Research and Development Programs of Shaanxi under Grants 2019ZDLGY13-07 and 2019ZDLGY13-04.

sensitive personal information could be sold as a commodity for commercial uses [14], [15]. For example, because of the disclosure of the private content preference data, users may receive a plenty of spam or fraud emails or phone calls. Therefore, it is necessary to pay attention on protecting on users' private content preferences. For instance, in [16], the authors propose a tag forgery based privacy-enhancing technology to protect the users' interests and preferences in social-tagging systems. In [17], the authors proposed to perturb user's content preference with local differential privacy. In [18], the authors design a tag suppression scheme based on data perturbation to protect end-user privacy in collaborative tagging services.

In our work, in order to address those issues above, we propose a scheme in which the CP offloads popular content into several storage for caching of APs according to the noisy content preference data from users. Therefore, the users' privacy is preserved and the problem of high backhaul load is resolved. Briefly, on the users' side, they first add local differential private noise to their content preferences to preserve the privacy. In other words, the users do not need to trust any third party including the CP. On the CP's side, the CP exploits the data-driven methodology to predict the content popularity distribution according to the collected noisy content preference data from the users and stimulates the APs with economic incentives to lease their storage for caching popular content. Consequently, we formulate a revenue maximization problem for the CP based on the description above and demonstrate that the revenue can be effectively optimized, while preserving the customers' local differential privacy in the ICN. Our salient contributions are summarized as follows.

- In our work, we preserve the privacy of users' preference information by locally adding differential noises on the users' side. Moreover, data-driven methodology is employed to forecast the content popularity for optimizing the revenue maximization problem on the CP side. Therefore, true preference content information of each individual user is not able to be obtained by CP or attackers like eavesdropper.
- With the assumption that the CP is semi-honest, in order to protect each individual user's content preference information, a local differential private mechanism called dBitFlip is exploited. Therefore, the CP is able to estimate the frequency distribution of different content from the users' noisy content preference information, meanwhile the true individual user's content preference information will not be leaked out.
- In the ICN, the Zipfs law is widely applied as a probabilistic model to characterize the content popularity. In our work, we employ data-driven approach to predicting the content popularity of a group of users without assumption of the distribution. We assume the CP constructs the noisy reference content popularity probability $\mathbb{P}_0$ according to the noisy users' content preferences, stimulates the cache-enabled APs to cooperate in the ICN to store popular content, and formulates the revenue maximization problem with the constraint of characteristic of uncertainty of content popularity with distance between the ambiguous distribution $\mathbb{P}$ and the noisy reference content popularity probability $\mathbb{P}_0$. We present and prove the relationship among the tolerance level θ, confidence level η and the local differential privacy level ϵ.

TABLE I
NOTATION LIST

Symbol	Definition
$\mathcal{U}$	Set of users
$\mathcal{F}$	Set of content
$\mathcal{M}$	Set of cache-enabled access points
c_m	Storage capacity of each access point
k_m	Fixed price to lease each access point
ξ^f	Possible realizations of each content
$\mathbb{P}_0$	Reference distribution of content popularity after injecting differentially private noise
$\mathbb{P}$	Ambiguous true distribution of content popularity
ϵ	DP privacy budget
$\mathcal{D}$	Confidence set
η	Confidence level
d_ζ	Distribution distance under ζ-structure probability metric
θ	Tolerance level
Ω	The sample space of ξ
$\varnothing$	The dimension of Ω
y_m	Binary value to indicate whether an access point is leased
β_f	Fraction of content f cached in the access points

- The formulated revenue maximization problem can be represented into a risk-averse two-stage stochastic problem (RA-SP). The Benders' decomposition algorithm is applied to solve the proposed problem with three different distribution distance metrics for robustness. We also conduct simulations to verify the effectiveness of the proposed scheme and discuss the impact of the of several key parameters in the proposed revenue maximization problem.

The rest of paper is organized as follows. In Section III, we describe the overview of our system, discuss the threat model and present differential privacy preliminaries. We review the related work on differential privacy and ICN in Section II. In Section IV, we introduce the dBitFlip mechanism to protect users' private content preference, give the formulation of the CP revenue maximization problem, and give a feasible solution to the problem. In Section V, we evaluate the performance of our proposed scheme. Finally, we draw conclusions in Section VI.

II. Related Work

In the ICN, in-network caching is necessary for transforming the traditional host-centric paradigm to a decentralized content-centric architecture. The ICN is being developed to provide resilient and robust network infrastructure services [19]. In [20], the authors proposed an in-network caching scheme called ProbCache that can estimate the content delivery path capacity according to the path length in order to efficiently allocate the resources and reduce the traffic redundancy. The authors in [21] integrated the ICN with wireless network virtualization architecture over the fifth-generation (5G) mobile wireless networks. They formulated a joint optimization problem with the gain of virtualization and caching in ICN in order to efficiently allocation virtual resource and at the same time reduce the backhaul traffic. In [7], the authors integrated the ICN virtualization with device-to-device (D2D) communications based on the software-defined networking (SDN) technology, which can dynamically allocate the virtual resources efficiently. Then, they formulated the utility maximization problem among all the mobile virtual network operators (MVNOs) with the consideration of imperfect channel estimation and measurement. Compared to this work, in our paper, we construct the content popularity distribution with historical data instead of assuming that the distribution follows the Zipf distribution. We leverage the data-driven optimization [22], [23] and give a feasible solution to our formulated revenue maximization problem under the characterization of uncertainty, which is brought by differentially private noise and the limited number of historical data.

Differential privacy was first introduced in [24] and has emerged as strong standard privacy guarantee to measure privacy disclosure. Basically, it is used to protect data providers' privacy when the statistical information of a database is publishing. However, the data providers will suffer from privacy leakage if the database is dishonest. Therefore, the local privacy model is utilized in differential privacy to provide local privacy guarantee to the data providers before the database aggregates the private data. In recent years, LDP has received a number of attentions [25]. For instance, in [26], the authors proposed RAPPOR to protect individual user's privacy, while it is able to estimate the occurring frequencies of a candidate set. The authors further extend this work in [27] that they proposed a mechanism to infer the association between multiple locally differentially private variables without the knowledge of a candidate set. However, when the dimension is higher, the error will increase and the computation complexity will exponentially grow. In [28], the authors proposed two optimized LDP protocol which can provide better utility with optimal variance value. Bassily al. [29] developed two locally differentially private heavy hitter mechanisms which could reach optimal or almost optimal error, and significantly reduce the cost and complexity of users' side when injecting the local noise.

III. Network Model and Preliminaries

A. System Description

In our system, as shown in Figure 1, we assume the content provider (CP), in the information-centric network (ICN), collects users' content preferences information with local differential noise, forecasts the content popularity by data-driven methodology, leases several storage for caching of the access points (APs) and offloads the popular content in advance into the cache. Additionally, the users apply the local differential privacy (LDP) protocols to add noise individually on their content preferences without a trusted third party and return the modified information to the CP.

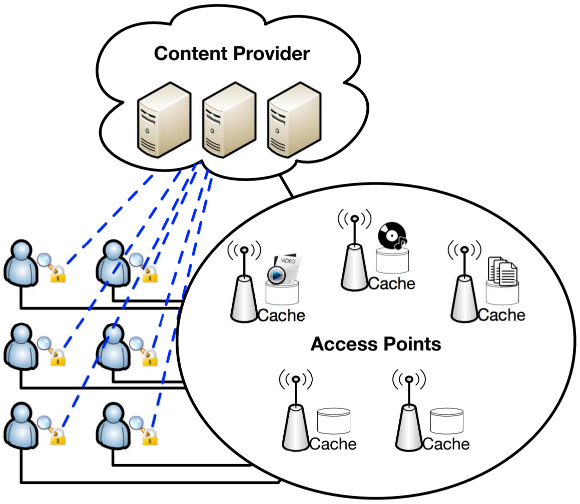


Fig. 1. System description.

In our scheme, we assume the set of users is $\mathcal{U} = \{1, \cdots, u, \cdots, U\}$, the content is represented as f with size s_f and the real content preference of each user is r_u that is selected from the set $\mathcal{F} = \{1, \cdots, f, \cdots, F\}$. There are several cache-enabled APs from a set $\mathcal{M} = \{1, \cdots, m, \cdots, M\}$ cooperating with the CP to provide high QoS. The available storage of each AP is c_m unit, which can be leased by the CP for caching popular content. The price to lease the available storage of each AP is k_m. With the LDP mechanism, the users add noise locally to their true content preference r_u, which is shown in IV-A in detail. The CP constructs the noisy reference content popularity probability $\mathbb{P}_0$ based on noisy content preference results and predicts the true popularity distribution by data-driven approach. With this distribution, the CP can decide to cache the entire content or a portion of each content and determine how many APs to be leased. Because of the uncertainty of noisy reference distribution, the revenue maximization problem is formulated, which is illustrated in IV-C. Moreover, the Benders' decomposition is deployed to solve the proposed maximization problem.

B. Threat Model

During the ICN resource allocation, the CP is trying to lease the storage of AP to cache the popular content according to the users' content preference in order to relieve congestion problem and reduce the backhaul load. Our target is to protect users' private content preference during the data aggregation since it may contain users' sensitive information which could be sold by adversary for commercial uses. Nonetheless, the privacy leakage after the aggregation process is out of our scope. We assume the attackers want to learn users' private content preference information and either the ICN server, a participatory user or the third party identity can be considered as attackers. However, data pollution attacks, that malicious users would modify their preference and try to affect the overall content popularity results, are beyond the scope of this paper. We suppose that the attackers are able to obtain side information or arbitrary background knowledge of users. Our objective is to hide the users' true preferred content despite the prior knowledge of adversaries.

C. Local Differential Privacy Preliminaries

Differential privacy [24] is used to obtain the statistical information of databases without disclosure of the data providers' privacy. Intuitively, two databases, which have only one element different from each other, are called neighbor databases. With a randomization algorithm $\mathcal{A}$ and the two neighbor databases as input, the outputs of the algorithm $\mathcal{A}$ are not distinguishable. The formal differential privacy requirements for the algorithm $\mathcal{A}$ are shown as follows.

***Definition** 1:* A randomized algorithm $\mathcal{A}$ satisfies ϵ-differential privacy (ϵ-DP), when given two neighbor databases D and D' and a privacy budget $\epsilon \geq 0$, in the following condition:

$$\frac{Pr[\mathcal{A}(D) = z]}{Pr[\mathcal{A}(D') = z]} \leq e^{\epsilon},$$

where z is the query output.

However, there must exists a trustworthy database or data aggregator when applying the centralized differential privacy. In our work, we assume the service database is *honest-but-curious*, which means the privacy leakage possibility increases. Therefore, local privacy setting is suitable in the situation that the data providers trust no one except themselves. The Warner's random response model [30] is one of the oldest local privacy model applied in survey sampling. In the Warner's model, if there are two answers of one question, the data provider will reply truly with probability of p and falsely with probability of $1 - p$. Combining local privacy and differential privacy, the definition of *local differential privacy* is shown as follows, which is similar to the centralized differential privacy.

***Definition** 2:* With a privacy confidence parameter $\epsilon \geq 0$, a randomized algorithm $\mathcal{A}$ satisfies ϵ-local differential privacy (ϵ-LDP), when given two inputs x and x' [31]:

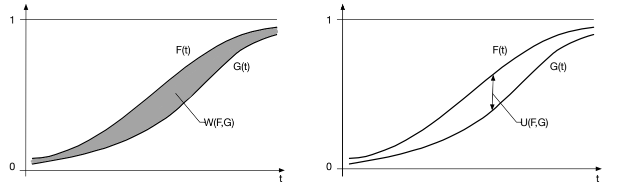
$$\frac{Pr[\mathcal{A}(x) = z]}{Pr[\mathcal{A}(x') = z]} \leq e^{\epsilon},$$

where z is the secure view of the input.
Therefore, with a specific output z from the randomized algorithm $\mathcal{A}$, it is not able to determine or can infer with negligible probability whether the input is x or x', since the data providers only return the obfuscated data $\mathcal{A}(x)$ to the data aggregator. Additionally, the privacy confidence parameter ϵ controls the privacy preservation level, which means there is more possibility to distinguish the outputs of the randomized algorithm $\mathcal{A}$ with two different inputs with a higher value of ϵ. In other words, smaller ϵ means higher privacy preservation level.

IV. Data-Driven Caching Revenue Maximization with Users' Local Differential Privacy

A. Protecting Private Content Preference with Local Differential Privacy

In information-centric networking (ICN), the content provider (CP) targets to offer satisfactory service to users. Therefore, the CP leases storage of several cache-enabled access points (APs) on the edge to cache popular content. Consequently, the CP first aggregates the users' content preference r_u from each user u and predict the content popularity. However, the users' content preference may contain private sensitive information as discussed in Section I. Thus, in this work, we employ the dBitFlip mechanism introduced in [32] to protect users' content preference.



(a) Wasserstein metrics (one-dimensional case).

(b) Uniform metric (one-dimensional case).

Fig. 2. Comparison of two metrics.

In dBitFlip mechanism, each user u first randomly selects d files without replacement from the file set $\mathcal{F}$, denoted as $\{j_1, j_2, \cdots, j_d\}$. Each user has his/her own content preference $r_u \in \mathcal{F}$. When the CP aggregates the content preference from users, each user will send a vector $b_u = [(b_{u,j_1}, j_1), (b_{u,j_2}, j_2), \cdots, (b_{u,j_d}, j_d)]$ back to the CP, where $b_{u,j_{1,2,\cdots,d}}$ are binary numbers. In order to protect users' privacy, the vector is supposed to be constructed according to the following equations,

$$\forall_{a \in [d]} Pr[b_{u,j_a} = 1] = \begin{cases} \frac{e^{\epsilon/2}}{e^{\epsilon/2}+1}, \text{when } r_u = j_a, \\ \\ \frac{1}{e^{\epsilon/2}+1}, \text{when } r_u \neq j_a. \end{cases} \quad (1)$$

According to the public coins model introduced in [33], the obfuscated output vector of each user is compressed into d bits (i.e., $b_u = [b_{u,j_1}, b_{u,j_2}, \cdots, b_{u,j_d}]$) and the index j_a can be generated by public coins. The reason is that the randomness of the index j is independent from the input and the privacy of the dBitFlip randomization holds differential privacy despite the index j. The randomness to choose the index j can be sent by the CP with public coins model and the CP receives j via other channels. Then, the obfuscated vector of each user can be represented into d bits.

According to the aggregated vectors b_u from users, the CP can estimate the content popularity histogram for all the files in the file set $\mathcal{F}$, which is shown as follows,

$$\hat{h}(v) = \frac{F}{Ud} \sum_{b_{u,v} \text{is received}} \frac{b_{u,v} \cdot (e^{\epsilon/2} + 1) - 1}{e^{\epsilon/2} - 1}. \quad (2)$$

For each file $v \in \mathcal{F}$, the CP first counts how many reported obfuscated vectors from users contain the element $b_{u,v}$ equal to 1. The correction of the influence of the randomization is implemented. Intuitively, for each file v, there are approximately Ud/F users reported $b_{u,v}$ in the obfuscated content preference vector.

***Lemma** 1:* The dBitFlip mechanism satisfies ϵ-LDP. After the CP aggregates the d bits obfuscated content preference vector b_u from each user u, the CP can estimate the content popularity $\hat{h}_t$. With probability at least $1-\delta$, the following condition holds [32],

$$\max_{v\in\mathcal{F}} |h(v)-\hat{h}(v)| \le \sqrt{\frac{5F}{Ud}}\cdot\sqrt{\log(\frac{6F}{\delta})}\cdot\frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}. \quad (3)$$

With the dBitFlip mechanism, we can get the error bound of the content popularity histogram as (3).

B. Data-Driven Analysis of Content Popularity

Most works in the ICN assume that the distribution of content popularity is known as Zipf distribution. However, practically, it may not accurately capture the statistical features in various geographical locations in ICN since the true content popularity distribution is unknown. Instead, only historical data or real content preferences of users can be obtained by the CP to construct the reference distribution of content popularity. Therefore, in our work, we employ data-driven risk-averse stochastic optimization approach (RA-SP) to making a decision to lease cache-enabled APs under the uncertainty of predicting the content popularity.

1) Reference Distribution Construction: We assume that the CP has U users in total and each user has his/her own content preference. In order to protect users' content preference, each user applies dBitFlip mechanism to obfuscate their true content preference. The overall private content popularity distribution can be represented by $\mathbb{P}_d$. Moreover, since the dBitFlip mechanism brings uncertainty, the content preference distribution $\mathbb{P}_d$ is not the true ambiguous content preference distribution. Therefore, we denote the true ambiguous content preference distribution as $\mathbb{P}$. According to the description in Section IV-A, the relationship between $\mathbb{P}$ and $\mathbb{P}_d$ can be calculated based on the knowledge of (3). The content popularity reference distribution $\mathbb{P}_0$ is constructed based on the content preference of selected Q users. Then, the selected users apply dBitFlip mechanism to obfuscate their true content preference and return the obfuscated content prefernce vector to the CP. Hence, the CP can aggregate the obfuscated content preference vectors and construct the reference distribution $\mathbb{P}_0$.

2) Convergence Rate Analysis: After constructing the reference distribution $\mathbb{P}_0$, we are going to find the relationship between the reference distributions $\mathbb{P}_0$, overall private content popularity distribution $\mathbb{P}_d$ and the true distribution $\mathbb{P}$. We use distance between the distributions to quantify such relationship. We apply a distance measurement proposed in [34] to express the distance between two distributions. We also define the distance between $\mathbb{P}_0$ and $\mathbb{P}$ is $d(\mathbb{P}_0,\mathbb{P}_d)$ constructed on the confident set $\mathcal{D}$. We represent this distance and the confident set $\mathcal{D}$ as follows:

$$\mathcal{D}=\{\mathbb{P}: d_\zeta(\mathbb{P}_0,\mathbb{P}_d)\le\theta\}, \quad (4)$$

$$d_\zeta(\mathbb{P}_0,\mathbb{P}_d)=\sup_{h\in\mathcal{H}}\left|\int_\Omega h d\mathbb{P}_0-\int_\Omega h d\mathbb{P}_d\right|, \quad (5)$$

where the distance under ζ-structure probability metric is denoted by $d_\zeta(\cdot,\cdot)$, the tolerance is denoted by θ, and $\mathcal{H}$ expresses the bounded measurable functions on Ω, which is the sample space of a random variable.

In this paper, three ζ-probability metrics are employed to calculate the distribution distance, which are derived as follows.

- **Kantorovich metric (K-metric)**: For K-metric $d_K(\mathbb{P}_0,\mathbb{P}_d)$, we have $\mathcal{H}=\{h: ||h||_L\le 1\}$, where $||h||_L := \sup\{h(x)-h(y)/\rho(x,y): x\neq y \text{ in } \Omega\}$, where $\rho(x,y)$ is the distance between two variables x and y. According to the Kantorovich-Rubinstein theorem, the K-metric and the Wasserstein metric are equivalent. Especially, with $\Omega=R$, let d_w denote the Wasserstein metric, then
$$d_w(\mathbb{P}_0,\mathbb{P}_d)=\int_{-\infty}^{+\infty}|F(x)-G(x)|dx, \quad (6)$$
where F and G are the cumulative distribution function of $\mathbb{P}_0$ and $\mathbb{P}_d$ respectively, which is demonstrated in Figure 2(a).
- **Fortet-Mourier metric (FM-metric)**: For FM-metric $d_{FM}(\mathbb{P}_0,\mathbb{P}_d)$, we have $\mathcal{H}=\{h: ||h||_C\le 1\}$, where $||h||_C := \sup\{h(x)-h(y)/c(x,y): x\neq y \text{ in } \Omega\}$ and $c(x,y)=\rho(x,y)\max\{1,\rho(x,a)^{p-1},\rho(y,a)^{p-1}\}$ for some $p\ge 1$ and any $a\in\Omega$. Given $p=1$, the FM-metric and the K-metric are equivalent.
- **Uniform metric (U-metric)**: For U-metric $d_U(\mathbb{P}_0,\mathbb{P}_d)$, we have $\mathcal{H}=\{I_{(-\infty,t]}, t\in R^n\}$. The U-metric can be represented as $d_U(\mathbb{P}_0,\mathbb{P}_d)=\sup_t|\mathbb{P}_0(x\le t),\mathbb{P}_d(x\le t)|$, which is shown in Figure 2(b). Similarly, F and G are the cumulative distribution function of $\mathbb{P}_0$ and $\mathbb{P}_d$ respectively.

The relationship among metrics is represented as for any two probability distributions $\mathbb{Q}$ and $\mathbb{R}$, $d_{FM}(\mathbb{R},\mathbb{Q})\le\Lambda\cdot d_K(\mathbb{R},\mathbb{Q})$, where $\Lambda=\max\{1,\varnothing^{p-1}\}$ for $p\ge 1$ and $\varnothing$ is the diameter of Ω [35]. In general, several properties hold in the ζ-structure metrics discussed above: 1) $d_\zeta(\mathbb{R},\mathbb{Q})=0$ if and only if $\mathbb{R}=\mathbb{Q}$; 2) ζ-structure metric satisfies the symmetric property that $d_\zeta(\mathbb{R},\mathbb{Q})=d_\zeta(\mathbb{Q},\mathbb{R})$; 3) ζ-structure metric satisfies the triangle inequality that $d_\zeta(\mathbb{R},\mathbb{Q})\le d_\zeta(\mathbb{R},\mathbb{O})+d_\zeta(\mathbb{O},\mathbb{Q})$ for any probability distribution $\mathbb{O}$.

***Proposition** 1:* For a general dimension case (i.e., $n\ge 1$),

$$Pr(d_K(\mathbb{P}_0,\mathbb{P})\le\theta)\ge 1-\exp(-\frac{(\theta-\alpha(\varnothing-1))^2}{2\varnothing^2}Q), \quad (7)$$

where $\alpha=\sqrt{\frac{5F}{Ud}}\cdot\sqrt{\log(\frac{6F}{\delta})}\cdot\frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}$.

Proof: First, we define a set $\mathcal{B}$ as follows,

$$\mathcal{B}:=\{\mu\in\mathcal{P}(\Omega): d_K(\mu,\mathbb{P})\ge\theta\}, \quad (8)$$

where $\mathcal{P}(\Omega)$ is the set of all probability measures defined on Ω. Let $\mathcal{C}(\Omega)$ be the set of bounded continuous function $\phi:\Omega\to R$. Recall that the number of selected users is Q, the ambiguous distribution of users' content preferences distribution is $\mathbb{P}$ and the constructed reference distribution of the content preferences is $\mathbb{P}_0$. Therefore, following the definitions, for each $\phi \in \mathcal{C}(\Omega)$, we have

$$
\begin{aligned}
&Pr(d_K(\mathbb{P}_0,\mathbb{P}) \geq \theta) = Pr(\mathbb{P}_0 \in \mathcal{B}) && (9)\\
&\leq Pr(\int_\Omega \phi d\mathbb{P}_0 \geq \inf_{\mu\in\mathcal{B}} \int_\Omega \phi d\mu) && (10)\\
&\leq \exp{(-Q \inf_{\mu\in\mathcal{B}} \int_\Omega \phi d\mu)} E(e^{Q\int_\Omega \phi d\mathbb{P}_0}) && (11)\\
&= \exp{(-Q \inf_{\mu\in\mathcal{B}} \{\int_\Omega \phi d\mu - \frac{1}{Q}\log E(e^{Q\int_\Omega \phi d\mathbb{P}_0})\})} && (12)\\
&= \exp{(-Q \inf_{\mu\in\mathcal{B}} \{\int_\Omega \phi d\mu - \log \int_\Omega e^{\phi} d\mathbb{P}_d\})}. && (13)
\end{aligned}
$$

In the above derivation, because of (8), we can get (9). The inequality (10) holds because $\mathbb{P}_0 \in \mathcal{B}$, $\mu \in \mathcal{B}$ and μ satisfies the minimum of $\int_\Omega \phi d\mu$. According to the Chebyshev's exponential inequality that $Pr(X \geq a) = \frac{\mathbb{E}[X]}{a}$, we can get the inequality (11). Based on the property of exponentiation, the equation (12) holds. As the historical data samples are independent from each other and drawn from the ambiguous distribution of overall private content preference distribution $\mathbb{P}_d$, we can obtain the equality (13).

Next, we define $\Delta(\mu) := \sup_{\phi\in\mathcal{C}(\Omega)} \int_\Omega \phi d\mu - \log \int_\Omega e^{\phi} d\mathbb{P}_d$. There should exist a series ϕ_n such that $\lim_{n\to\infty} \int_\Omega \phi d\mu - \log \int_\Omega e^{\phi} d\mathbb{P}_d = \Delta(\mu)$, because of the definition of $\mathcal{C}(\Omega)$. Hence, given a small positive number $\theta' > 0$, a constant number n_0 is supposed to exist so that $\Delta(\mu) - (\int_\Omega \phi_n d\mu - \log \int_\Omega e^{\phi}_n d\mathbb{P}_d) \leq \theta'$ for any $n \geq n_0$. Now, we can substitute ϕ_n for ϕ in equation (13) as follows,

$$
\begin{aligned}
&Pr(\mathbb{P}_0 \in \mathcal{B})\\
&\leq \exp\left(-U \inf_{\mu\in\mathcal{B}} \left\{\int_\Omega \phi_n d\mu - \log \int_\Omega e^{\phi}_n d\mathbb{P}_d\right\}\right) && (14)\\
&\leq \exp\left(-U \inf_{\mu\in\mathcal{B}} \{\Delta(\mu) - \theta'\}\right) && (15)
\end{aligned}
$$

As proved in [36], we have

$$\Delta(\mu) = d_{KL}\left(\mu, \mathbb{P}_d\right). \tag{16}$$

Based on the defination of Kantorovich metric and Kullback-Leibler divergence, we have

$$d_K\left(\mu, \mathbb{P}_d\right) \leq \varnothing\sqrt{2d_{KL}\left(\mu, \mathbb{P}_d\right)}, \forall \mu \in \mathcal{P}(\Omega), \tag{17}$$

where $\varnothing$ is the diameter of Ω. Recall the definition of $\mathcal{B}$ and $\mu \in \mathcal{B}$, we have $d_K\left(\mu, \mathbb{P}\right) \geq \theta$. According to the triangle inequality property of the ζ-structure metric that we introduced before, we have

$$\theta \leq d_K(\mu, \mathbb{P}) \leq d_K(\mu, \mathbb{P}_d) + d_K(\mathbb{P}_d, \mathbb{P}). \tag{18}$$

Consequently, we can obtain

$$d_K(\mu, \mathbb{P}_d) \geq \theta - d_K(\mathbb{P}, \mathbb{P}_d). \tag{19}$$

Then, we can combine (17) and (19),

$$d_{KL}\left(\mu, \mathbb{P}\right) \geq \frac{(\theta - d_K(\mathbb{P}, \mathbb{P}_d))^2}{2\varnothing^2}. \tag{20}$$

In Subsection IV-A, we've introduced the dBitFlip mechanism and given an error bound of the histogram as (3). Now, we set the right term of the inequality (3) as α (i.e., $\alpha = \sqrt{\frac{5F}{Ud} \cdot \log(\frac{6F}{\delta}) \cdot \frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}}$). Then, we can derive that $d_K(\mathbb{P}_d, \mathbb{P}_0) \leq \alpha(\varnothing - 1)$. Here, F equals to the diameter of sample space $\varnothing$.

Combining (15), (16), (20), we have

$$Pr(\mathbb{P}_0 \in \mathcal{B}) \leq \exp\left(-Q\left(\frac{(\theta - \alpha(\varnothing - 1))^2}{2\varnothing^2} - \theta'\right)\right). \tag{21}$$

We can define $\theta' = \lambda/Q$, where λ is an arbitrary small positive number. Then, we can obtain

$$
\begin{aligned}
&Pr\left(d_K\left(\mathbb{P}_0, \mathbb{P}\right) \geq \theta\right) = Pr\left(\mathbb{P}_0 \in \mathcal{B}\right)\\
&\leq \exp\left(-Q\frac{(\theta - \alpha(\varnothing - 1))^2}{2\varnothing^2} + \lambda\right). && (22)
\end{aligned}
$$

Because λ is arbitrarily small that can be ignored, we have

$$Pr(d_K(\mathbb{P}_0, \mathbb{P} \leq \theta)) \geq 1 - \exp\left(-\frac{(\theta - \alpha(\varnothing - 1))^2}{2\varnothing^2}Q\right). \tag{23}$$

■

From the definition of metrics and relationships between metrics under ζ-structure, we can derive the convergence property and convergence rate for the other metrics accordingly. For the uniform metric, the convergence rate can be derived from the Dvoretzky-Kiefer-Wolfowitz inequality [37].

***Proposition** 2:* The convergence rate of the uniform metric for a single dimension case is (i.e., $n = 1$),

$$P(d_U(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - 2e^{-2Q(\theta - \alpha)^2}, \tag{24}$$

From the relation between the Fortet-Mourier metric and Kantorovich metric, we can derive the convergence rate of Fortet-Mourier metric as follows.

***Proposition** 3:* For a general dimension (i.e., $n \geq 1$), we have

$$Pr(d_{FM}(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - \exp\Big(-\frac{(\theta - \alpha(\varnothing - 1))^2 Q}{2\varnothing^2\Lambda^2}\Big), \tag{25}$$

where $\Lambda = \max\{1, \varnothing^{p-1}\}$.

We assume the confidence level is denoted as η. Based on the derived inequalities (7), (24) and (25), we can easily calculate the relationships between the tolerance level θ and the confidence level η. For instance, in the Kantorovich metric, we can set the confidence level $\eta = 1 - \exp(-\frac{(\theta - \alpha(\varnothing - 1))^2}{2\varnothing^2}Q)$ and we can further derive

$$\theta = \varnothing\sqrt{\frac{2log(1/(1-\eta))}{Q}} + \alpha(\varnothing - 1), \tag{26}$$

where the first term affected by the amount of historical data and the confidence level, and the second term affected by the differential privacy budget ϵ. It is obvious that with higher amount of historical data, θ is smaller and with higher differential privacy level ϵ, α is smaller and θ is smaller.

C. Caching Revenue Maximization Problem with Local Privacy Preservation

As we describe before, the CP collects user's noisy content preferences with LDP and aggregates the frequency estimation of each content. Consequently, the CP is able to get the noisy content preference of a number of selected users Q. In our work, according to the noisy content popularity, the CP can construct the reference distribution $\mathbb{P}_0$ based on content preference of the selected number of users. We assume there are F candidate files in the set $\mathcal{F} = \{1, \cdots, f, \cdots, F\}$ that are selected to store in the cache-enabled APs. The CP can decide whether to store a portion β_f of each file f in the cache, where $0 \leq \beta_f \leq 1$. We assume there are M APs available and each AP has the available capacity c_m. The binary parameter y_m is used to represent if an AP is leased by the CP with the price k_m. We denote the size of each file as s_f. We represent the profit per unit size of a file stored in AP as $\phi^{(a)}$ and the profit per unit size of a file not stored in AP as $\phi^{(c)}$. We assume that the random variable ξ has F possible realizations $\{\xi_1, \xi_2, \ldots, \xi_F\}$ with the probability p_f for each realization. Here, p_f is the content popularity of the file f. The $s(\xi_f)$ can be represented the file size of each realization. In addition, there are a total number of U users served by the CP. In order to maximize revenue, the CP employs data-driven method to predict the real content popularity, decides how much of a content file stored in APs, and selects cache-enabled APs from a given group. Because of contribution of the APs, the bachhual load is reduced. Therefore, the revenue maximization problem for CP can be formulated as follows:

$$
\begin{aligned}
&\max_{y} \sum_{m=1}^{M} -k_m y_m + \\
&\min_{\mathbb{P}} \mathbb{E}_{\mathbb{P}} \max_{\beta_f} [\beta_f s(\xi_f)\phi^{(a)}U + (1-\beta_f)s(\xi_f)\phi^{(c)}U], \qquad (27)\\
&\text{s.t.:} \quad \sum_{f=1}^{F} \beta_f s_f \leq \sum_{m=1}^{M} c_m y_m, \qquad (27\text{a})\\
&\qquad y_m \in \{0,1\}, \forall m, \qquad (27\text{b})\\
&\qquad 0 \leq \beta_f \leq 1, \forall f, \qquad (27\text{c})\\
&\qquad \mathbb{P} \in \mathcal{D}. \qquad (27\text{d})
\end{aligned}
$$

In the formulation, y_m and β_f represent the first-stage and second-stage decision variables. The constraint (27a) indicates that the total amount of cached files should not exceed the capacity c_m of all the leased cache, (27b) indicates whether the cache m is leased by the CP, and (27c) shows the portion of each file cached in APs. In (27), we would like to maximize the overall revenue. Since we add noise in the processed energy profile, the distribution of real demand is ambiguous. Therefore, we construct the confident set $\mathcal{D}$, and let $\mathbb{P} \in \mathcal{D}$ so as to maximize the total revenue under the worst-case distribution realization in $\mathcal{D}$. In order to consider the worst-case scenario that can happen during the caching, we minimize the revenue based on the distribution $\mathbb{P}$. Therefore, the proposed problem is a risk-averse two-stage problem.

D. Solution to Caching Optimization under Distribution Uncertainty

We utilize the Benders' decomposition algorithms [38] to solve the proposed optimization problem. The sample space of the random variable ξ is $\Omega = \{\xi_1, \xi_2, \cdots, \xi_F\}$. For each scenario ξ_f, the second-stage maximization problem is independent with ξ_f. The optimization problem (27) can be reformulated as:

$$
\begin{aligned}
&\max_{y} \sum_{m=1}^{M} -k_m y_m + \\
&\min_{p} \max_{\alpha} \sum_{f=1}^{F} p_f [\beta_f s(\xi_f)\phi^{(a)}U + (1-\beta_f)s(\xi_f)\phi^{(c)}U], \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (28)\\
&\text{s.t.} \quad (27a) - (27d),
\end{aligned}
$$

We need to first dualize the second-stage maximization problem in order to solve (28). We replace $(1 - \beta_f)$ with the variable ρ_f and add another two constraints $0 \leq \rho_f \leq 1, \rho_f = 1 - \beta_f$ to the original second-stage problem. Hence, the reformulated second-stage maximization problem is shown as,

$$
\begin{aligned}
&\max_{\alpha} \sum_{f=1}^{F} p_f [\beta_f s(\xi_f)\phi^{(a)}U + \rho_f s(\xi_f)\phi^{(c)}U], \qquad (29)\\
&\text{s.t.} \quad (27a), (27c),\\
&\qquad 0 \leq \rho_f \leq 1, \qquad (29\text{a})\\
&\qquad \rho_f = 1 - \beta_f. \qquad (29\text{b})
\end{aligned}
$$

Then, the dual problem is shown as follows,

$$
\begin{aligned}
&\min_{\forall w} \sum_{m=1}^{M} c_m y_m w_1 + \big(\sum_{f=1}^{F} -w_2^f + w_3^f + w_4^f + w_5^f\big), \qquad (30)\\
&\text{s.t.} \quad s_f w_1 - w_2^f + w_3^f + w_4^f \geq p_f s(\xi_f)\phi^{(a)}U, \forall f, \quad (30\text{a})\\
&\qquad - w_2^f + w_3^f + w_5^f \geq p_f s(\xi_f)\phi^{(c)}U, \forall f, \qquad (30\text{b})\\
&\qquad w_1, w_2^f, w_3^f, w_4^f, w_5^f \geq 0, \forall f, \qquad (30\text{c})
\end{aligned}
$$

where all w are dual variables for all constraints of the second-stage maximization problem. Therefore, we can combine two minimization problem and obtain the subproblem as follows,

$$
\begin{aligned}
&\varphi(y) = \min_{p, \forall w} \sum_{m=1}^{M} c_m y_m w_1 + \big(\sum_{f=1}^{F} -w_2^f + w_3^f + w_4^f + w_5^f\big), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (31)\\
&\text{s.t.} \quad (30\text{a}), (30\text{b}), (30\text{c}),\\
&\qquad \sum_{f=1}^{F} p_f = 1, \qquad (31\text{a})\\
&\qquad \mathbb{P} \in \mathcal{D}. \qquad (31\text{b})
\end{aligned}
$$

According to the discussions in Section IV-B, the constraint (31b) can be reformulated as,

$$
\max_{h_f} \sum_{f=1}^{F} h_f p_f^0 - \sum_{f=1}^{F} h_f p_f \leq \theta, \forall h_f : ||h||_\zeta \leq 1, \qquad (32)
$$

where $|h||_\zeta$ is defined according to different metrics, p_f^0 is the probability based on the reference distribution $\mathbb{P}_0$, and θ is the tolerance level. For the Kantorovich metric, $|h_i - h_j| \leq \rho(\xi^i, \xi^j)$. For the Fortet-Mourier metric, $|h_i - h_j| \leq$

$\rho(i,j)\max\{1,\rho(\xi^i,a)^{p-1},\rho(\xi^j,a)^{p-1}\}$. The constraint (32) can be summarized as $\sum_{i=1}^{F} a_{ij}h_i \leq b_j, j=1,\cdots,F$. We can reformulate the constraint as the following problem:

$$\begin{aligned} \max_{h_i} \quad & \sum_{i=1}^{F} h_i p_i^0 - \sum_{i=1}^{F} h_i p_i, && (33)\\ \text{s.t.} \quad & \sum_{i=1}^{F} a_{ij}h_i \leq b_j, j=1,\cdots,F. \end{aligned}$$

The dual problem of (33) can be formulated as:

$$\begin{aligned} \min_{u} \quad & \sum_{j=1}^{F} b_j u_j, && (34)\\ \text{s.t.} \quad & \sum_{j=1}^{F} a_{ij}u_j \geq p_i^0 - p_i, i=1,\cdots,F, \end{aligned}$$

where u is the dual variable. Therefore, for the Kantorovich metric and Fortet-Mourier metric, the constraint (31b) can be replaced with $\sum_{j=1}^{F} b_j u_j \leq \theta$ and $\sum_{j=1}^{F} a_{ij}u_j \geq p_i^0 - p_i, i = 1,\cdots,F$. For the Uniform metric, the constraint (31b) can be reformulated as $\left|\sum_{f=1}^{j}\left(p_f^0 - p_f\right)\right| \leq \theta, j=1,\cdots,F$.

We denote σ as the second-stage total revenue and formulate the master problem. Then, it can be solved by iteratively generating feasibility cut and optimality cut. The master problem can be represented as follows,

$$\begin{aligned} \max_{y} & \sum_{m=1}^{M} -k_m y_m + \sigma && (35)\\ & \text{s.t. Feasibility cut, Optimality cut} \end{aligned}$$

1) Feasibility Cuts: The L-shaped method is employed to generate the feasibility cut. We need to check whether the value of the first-stage variable y is feasible for the constraint (27a), (27c), (29a), and (29b). Hence, we can formulate the feasibility check problem as follows,

$$\begin{aligned} \min_{\kappa,\beta,\rho} & \kappa_1 + \sum_{f=1}^{F}\sum_{i=2}^{5}\kappa_i^f, && (36)\\ \text{s.t.} \ & -\sum_{f=1}^{F}\beta_f s_f + \kappa_1 \geq -\sum_{m=1}^{M} c_m y_m,\\ & \beta_f + \rho_f + \kappa_2^f \geq 1,\\ & -\beta_f - \rho_f + \kappa_3^f \geq -1\\ & -\beta_f + \kappa_4^f \geq -1,\\ & -\rho_f + \kappa_5^f \geq -1,\\ & \rho_f, \beta_f \geq 0, \kappa_1, \kappa_i^f \geq 0, i=2,\ldots,5, \forall f. \end{aligned}$$

The dual problem of (36) can be represented as follows,

$$\upsilon(y) = \max_{\forall\hat{w}} -\sum_{m=1}^{M} c_m y_m \hat{w}_1 + \big(\sum_{f=1}^{F} \hat{w}_2^f - \hat{w}_3^f - \hat{w}_4^f - \hat{w}_5^f\big), \quad (37)$$

$$\begin{aligned} \text{s.t.} \quad & -s_f\hat{w}_1 + \hat{w}_2^f - \hat{w}_3^f - \hat{w}_4^f \leq 0, \forall f,\\ & \hat{w}_2^f - \hat{w}_3^f - \hat{w}_5^f \leq 0, \forall f,\\ & \hat{w}_1, \hat{w}_2^f, \hat{w}_3^f, \hat{w}_4^f, \hat{w}_5^f \in [0,1], \forall f, \end{aligned}$$

where $\hat{w}_1, \hat{w}_2^f, \hat{w}_3^f, \hat{w}_4^f$,and $\hat{w}_5^f$ are the dual variables for the constraints of the minimization problem (36). After solving this dual problem (37), if $\upsilon(y)=0$, we can confirm that the first-stage solution is feasible. If $\upsilon(y) > 0$, we need to add the following feasibility cut to the master problem,

$$-\sum_{m=1}^{M} c_m y_m \hat{w}_1 + \big(\sum_{f=1}^{F} \hat{w}_2^f - \hat{w}_3^f - \hat{w}_4^f - \hat{w}_5^f\big) \leq 0. \quad (38)$$

2) Optimality Cuts: After verifying the first-stage variable y is feasible, we need to check if the variables y and σ of the master problem are optimal. Therefore, we apply the value of y to the subproblem and solve $\varphi(y)$. If $\varphi(y)-\sigma \geq 0$, we can claim that the value of y is the optimal solution. Otherwise, we need to add an optimality cut to the master problem as follow,

$$\sum_{m=1}^{M} c_m y_m \hat{w}_1 + \big(\sum_{f=1}^{F} -\hat{w}_2^f + \hat{w}_3^f + \hat{w}_4^f + \hat{w}_5^f\big) \geq \sigma. \quad (39)$$

The detailed algorithm for the solution of our propose optimization problem is summarized in Algorithm 1.

Algorithm 1 **Algorithm for Solution to Caching Revenue Maximization Problem**

```
 1: Input: The number of selected users' data $Q$, the reference distribution of content popularity $\mathbb{P}_0$, the confidence level $\eta$, the number of candidate content files $F$, the differentially private level $\epsilon$, the value of $d$ in dBitFlip mechanism
 2: Output: Objective value of the problem (27).
 3: Calculate the error bound $\alpha$ of the dBitFlip mechanism and the tolerance level $\theta$ based on three different $\zeta$-structure metrics
 4: Reformulate the problem (27) as a master problem (35) and a subproblem (31)
 5: Solve the master problem and get the value of $y$ and $\sigma$
 6: Feasibility check by solving $\upsilon(y)$
 7: if $\upsilon(y) > 0$ then
 8:     Generate feasibility cut (38)
 9:     Go to line 5
10: end if
11: Check optimality by solving $\varphi(y)$
12: if $\varphi(y) < \sigma$ then
13:     Generate optimality cut (39)
14:     Go to line 5
15: end if
16: Output solution
```

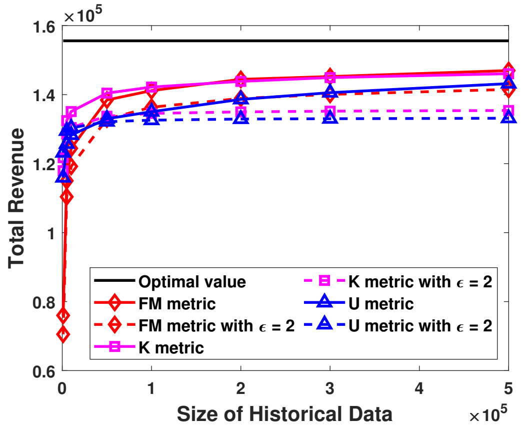


Fig. 3. Expected revenue without users' privacy preservation.

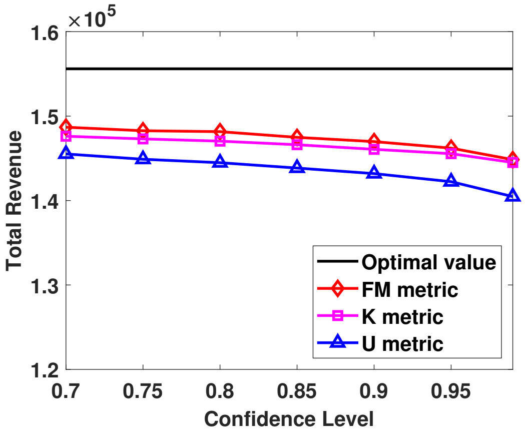


Fig. 4. Expected revenue with different confidence levels.

V. Performance Evaluation

A. Simulation Setup

In our simulations, we use the "Statistics and Social Network of YouTube Videos" dataset [39], which crawled video information from the Youtube API. We randomly select 100 videos to evaluate the performance of proposed scheme. We construct the reference distributions with the number of video views and assume that the CP provides service to 2,000,000 users and take a survey on the content preference from the selected users. The users implement dBitFlip mechanism with $d = 4$ to protect their privacy, and then send the noisy preference results to the CP. To be specific, the sample users choose interested videos from 100 video candidates, add noise and send back to the CP. We assume there are 5 cache-enabled APs with the capacity $[160, 160, 320, 320, 640]$ and the leasing cost $[2000, 2400, 4800, 4000, 7200]$. We set the profit per unit size of a file stored in AP as $\phi^{(a)} = 0.001$ and the profit per unit size of a file not stored in AP as $\phi^{(c)} = 0.0003$. We provide random caching and caching all popular content as baselines for comparisons. In random caching, videos are randomly selected to cache in the APs with the constraint that the total amount of cached videos cannot exceed the capacity of the APs. In caching all popular content, based on the reference content popularity distribution $\mathbb{P}_0$, the most popular videos are selected to be cached into the APs with the constraint of the limited capacity.

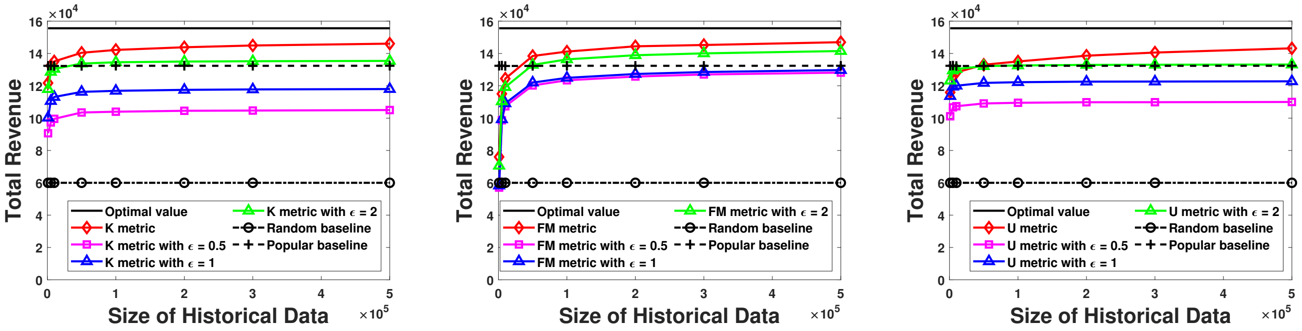
We investigate the impacts of several key parameters when solving the revenue maximization problem. The number of sampled users Q (i.e., the size of historical data) is one of the key parameters. In the LDP mechanism dBitFlip, the privacy level ϵ is the significant parameter. To solve the proposed revenue maximization problem, we apply three different metrics, Kantorovich metric, Fortet-Mourier metric and uniform metric to quantify the uncertainty of the content demand. The confidence level η is an important parameter to calculate the convergence rate as discussed in Section IV-B.

B. Privacy and performance analysis

1) Effects of the number of sampled users: We first set the confidence level η to 0.9 and study the impact of the number of aggregated users' data on the performance. We set the number of aggregated user's data from 1000 to 500,000, which is sampled based on the pre-defined true content popularity distribution. We conduct 10 independent runs of the proposed scheme and two baselines, and show the average revenue values in Figure 3, which demonstrates the performance of the expected revenues under the three introduced metrics, Kantorovich metric (K metric), Fortet-Mourier metric (FM metric) and uniform metric (U metric). It is shown that under all of the three metrics, the expected revenues are higher with a larger number of data. As discussed in Section IV-B, with a larger number of aggregated data Q, the tolerance level θ is smaller. When applying the dBitFlip mechanism, the total revenue is smaller than that without privacy preserving mechanism. The reason is that the tolerance level with privacy guarantee is larger than that without privacy preservation. In other words, with more data and without privacy preservation, the constructed reference distribution is more accurate and closer to the ambiguous distribution, which leads to a higher total revenue.

2) Effects of confidence level: Figure 4 shows the comparison under different confidence levels with the three metrics and the dBitFlip mechanism is not applied. Here, we set the number of sampled users as 500,000, and test different confidence levels between 0.7 and 0.99, respectively. We can notice that, with higher confidence level, the expected revenue of the CP is lower. A higher confidence level means it is guaranteed that the distance between the reference distribution and the ambiguous true distribution is smaller than the tolerance level θ with a very high probability. Since the tolerance level θ increases with a higher confidence level, the distance between the reference distribution and the ambiguous true distribution becomes larger. Therefore, the CP's total revenue degrades with a higher confidence level.

3) Effects of differential privacy budget: In Figure 5, we study the impact of the differential privacy levels ϵ under the three metrics. Here, we set the confidence level as 0.9, and select three different ϵ values, i.e., 2, 1, 0.5, respectively. A higher ϵ value means a lower privacy preservation level. As we conduct 10 independent runs on each simulation, the curves in Figure 5 show the mean total revenue values. For



(a) Kantorovich metric.

(b) Fortet-Mourier metric.

(c) Uniform metric.

Fig. 5. Performance comparison under different local differential privacy levels and different probability metrics.

the random baseline, the content is randomly cached in the APs with the constraint of limited capacity. Since this baseline is independent with any historical data, the performance is worst. In the caching all popular content baseline, the content is cached in the APs based on the content popularity, which is constructed with different number of historical data. We can find that without dBitFlip mechanism, the performances of all metrics are better than the popular baseline. With a larger ϵ value, which means the privacy level is lower, the total revenue of the proposed private scheme is larger than the random baseline. Moreover, we can observe that under the FM metric, the influence of privacy preservation mechanism on the total revenue value is smallest, because the privacy term in the equation of tolerance level has minor effects on the value of θ. Under all metrics, the total revenue of CP degrades with lower ϵ value, which means higher privacy level. Based on the relationship between θ, η and ϵ, we can find that with a higher ϵ value, the tolerance level θ decreases. Recall that with a lower θ level, the confidence set becomes smaller. Hence, it will lead to higher total revenue of the CP.

VI. Conclusion

In our work, we propose a scheme to predict the content popularity based on selected users' locally differentially private content preference data in information-centric networks, and formulate a revenue maximization problem for the CP. Because of the uncertainty of the content popularity distribution brought by the limited historical data and the LDP mechanism, data-driven methodology is employed to characterize such uncertainty based on the collected noisy content preference data. In addition, we develop an algorithm to feasibly solve the proposed problem. We conduct simulations to show the effectiveness of the proposed scheme and illustrate the trade-off between privacy and utility.

References

[1] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Communications Magazine*, vol. 50, no. 7, July 2012.

[2] S. Zhang, J. Li, H. Luo, J. Gao, L. Zhao, and X. S. Shen, "Low-latency and fresh content provision in information-centric vehicular networks," *IEEE Transactions on Mobile Computing (early access)*, September 2020.

[3] A. Ndikumana, K. Thar, T. M. Ho, N. H. Tran, P. L. Vo, D. Niyato, and C. S. Hong, "In-network caching for paid contents in content centric networking," in *2017 IEEE Global Communications Conference*, Singapore, December 2017, pp. 1–6.

[4] "Cisco visual networking index: Forecast and methodology, 2016–2021," Cisco, 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[5] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong, "Joint communication, computation, caching, and control in big data multi-access edge computing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1359–1374, June 2020.

[6] M. Mangili, F. Martignon, S. Paris, and A. Capone, "Bandwidth and cache leasing in wireless information-centric networks: A game-theoretic study," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 679–695, January 2017.

[7] K. Wang, H. Li, F. R. Yu, and W. Wei, "Virtual resource allocation in software-defined information-centric cellular networks with device-to-device communications and imperfect csi," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10 011–10 021, December 2016.

[8] L. A. Adamic and B. A. Huberman, "Zipf's law and the internet." *Glottometrics*, vol. 3, no. 1, pp. 143–150, 2002.

[9] J. Gao, L. Zhao, and X. Shen, "The study of dynamic caching via state transition field—the case of time-invariant popularity," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5924–5937, December 2019.

[10] C. Zhao and Y. Guan, "Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics," *Available on Optimization Online*, 2015.

[11] L. Aitchison, N. Corradi, and P. E. Latham, "Zipf's law arises naturally when there are underlying, unobserved variables," *PLoS computational biology*, vol. 12, no. 12, p. e1005110, December 2016.

[12] J. Gao, S. Zhang, L. Zhao, and X. S. Shen, "The design of dynamic probabilistic caching with time-varying content popularity," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1672–1684, April 2021.

[13] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, April-June 2018.

[14] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, July/August 2016.

[15] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial iots," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, May 2020.

[16] S. Puglisi, J. Parra-Arnau, J. Forné, and D. Rebollo-Monedero, "On content-based recommendation and user privacy in social-tagging systems," *Computer Standards & Interfaces*, vol. 41, pp. 17–27, September 2015.

[17] X. Zhang, J. Wang, H. Li, Y. Guo, Q. Pei, P. Li, and M. Pan, "Data-driven caching with users' local differential privacy in information-centric networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab, December 2018.

[18] J. Parra-Arnau, A. Perego, E. Ferrari, J. Forne, and D. Rebollo-Monedero, "Privacy-preserving enhanced collaborative tagging," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 180–193, January 2014.

[19] A. V. Vasilakos, Z. Li, G. Simon, and W. You, "Information centric network: Research challenges and opportunities," *Journal of network and computer applications*, vol. 52, pp. 1–10, June 2015.

[20] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *Proceedings of the second edition of the ICN workshop on Information-centric networking*, Helsinki, Finland, August 2012.

[21] C. Liang, F. R. Yu, and X. Zhang, "Information-centric network function virtualization over 5g mobile wireless networks," *IEEE network*, vol. 29, no. 3, pp. 68–74, June 2015.

[22] M. Pan, C. Zhang, P. Li, and Y. Fang, "Joint routing and link scheduling for cognitive radio networks under uncertain spectrum supply," in *2011 Proceedings IEEE INFOCOM*, Shanghai, China, April 2011, pp. 2237–2245.

[23] J. Wang, X. Zhang, H. Zhang, H. Lin, H. Tode, M. Pan, and Z. Han, "Data-driven optimization for utility providers with differential privacy of users' energy profile," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab, December 2018.

[24] C. Dwork, "Differential privacy: A survey of results," in *International Conference on Theory and Applications of Models of Computation*, Xi'an, China, April 2008.

[25] T. Wang, N. Li, and S. Jha, "Locally differentially private frequent itemset mining," in *2018 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, July 2018.

[26] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (CCS'14)*, Scottsdale, AZ, November 2014.

[27] G. Fanti, V. Pihur, and Ú. Erlingsson, "Building a rappor with the unknown: Privacy-preserving learning of associations and data dictionaries," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 3, pp. 41–61, May 2016.

[28] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally differentially private protocols for frequency estimation," in *Proceedings of the 26th USENIX Security Symposium*, Vancouver, BC, Canada, August 2017.

[29] R. Bassily, K. Nissim, U. Stemmer, and A. G. Thakurta, "Practical locally private heavy hitters," in *Advances in Neural Information Processing Systems*, Long Beach, CA, December 2017.

[30] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[31] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS)*, Berkeley, CA, October 2013.

[32] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," in *Advances in Neural Information Processing Systems (NIPS)*, Long Beach, CA, December 2017.

[33] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, Portland, OR, June 2015.

[34] G. C. Calafiore, "Ambiguous risk measures and optimal robust portfolios," *SIAM Journal on Optimization*, vol. 18, no. 3, pp. 853–877, October 2007.

[35] J. Wang, X. Zhang, Q. Zhang, M. Li, Y. Guo, Z. Feng, and M. Pan, "Data-driven spectrum trading with secondary users' differential privacy preservation," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 1, pp. 438–447, Jan.-Feb. 2021.

[36] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications - 2010*. Springer, 2010.

[37] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

[38] A. M. Geoffrion, "Generalized benders decomposition," *Journal of optimization theory and applications*, vol. 10, no. 4, pp. 237–260, October 1972.

[39] X. Cheng, C. Dale, and J. Liu, "Dataset for "statistics and social network of youtube videos"," http://netsg.cs.sfu.ca/youtubedata/, 2008.