Timely Private Information Retrieval

Karim Banawan[1], Ahmed Arafa[2], and Sennur Ulukus[3]

[1]Electrical Engineering Department, Faculty of Engineering, Alexandria University
[2]Electrical and Computer Engineering Department, University of North Carolina at Charlotte
[3]Department of Electrical and Computer Engineering, University of Maryland

***Abstract*—We introduce the problem of *timely* private information retrieval (PIR) from N non-colluding and replicated servers. In this problem, a user desires to retrieve a message out of M messages from the servers, whose contents are continuously updating. The retrieval process should be executed in a timely manner such that no information is leaked about the identity of the message. To assess the timeliness, we use the *age of information* (AoI) metric. Interestingly, the timely PIR problem reduces to an AoI minimization subject to PIR constraints under *asymmetric traffic*. We explicitly characterize the optimal tradeoff between the PIR rate and the AoI metric (peak AoI or average AoI) for the case of $N = 2$, $M = 3$. Further, we provide some structural insights on the general problem with arbitrary N, M.**

I. Introduction

Private information retrieval (PIR), introduced in [1], investigates the privacy of downloaded content from distributed databases. In classical PIR, a user needs to retrieve a message without disclosing the identity of the desired message to any individual server. Sun and Jafar in [2] investigate the PIR problem using information-theoretic measures. The goal of [2] is to characterize the PIR capacity, which is the supremum of PIR retrieval rates among all retrieval schemes. The PIR rate is the ratio between the desired message size and the total download from the servers. The optimal scheme in [2] is a greedy retrieval scheme that downloads *symmetric* amount of downloaded bits from each server. Several variants of the classical problem were studied afterwards, e.g., [3]–[36].

All these works, however, do not consider the *timeliness* of the retrieval process. This is crucial in real-time applications. For instance: Investors in the stock market need to retrieve information about the desired stocks without leaking information about their interests. The stock prices change rapidly over time. This calls for retrieving stock information *privately* and *timely*, motivating the *timely PIR* problem. The timeliness can be measured by the *age of information* (AoI) metric, defined as the time elapsed since the latest message has been retrieved. AoI has been originally studied in queuing networks, e.g., [37], [38], and has found its way into other numerous contexts, see, e.g., [39]–[53], and the recent survey in [54].

From a technical standpoint, the timely PIR problem investigates an interesting tension between privacy and AoI. Specifically, minimizing AoI (if privacy is ignored) necessitates downloading from a *single* server with minimal delay statistics. This however results in the worst achievable PIR rate. This extreme case poses an interesting question: **what is the optimal tradeoff between the PIR rate and the AoI?**

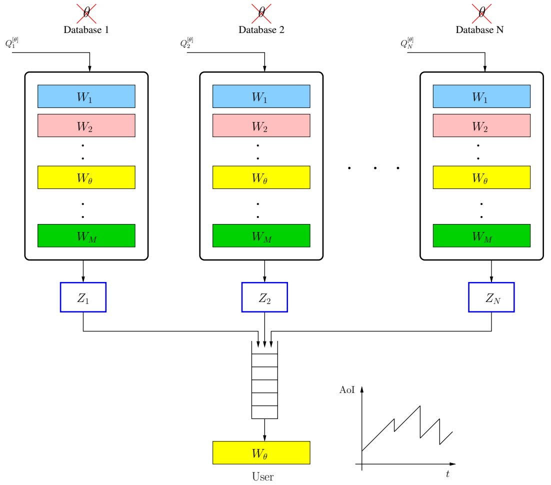

Fig. 1. The timely PIR problem.

In this paper, we introduce the timely PIR problem. A user aims at correct retrieval of a message W_θ (of size L) out of M messages from N non-colluding and replicated servers with different delay statistics. To that end, the user designs queries to minimize the AoI metric such that the queries do not leak θ to any individual server and achieve a minimum PIR rate of R_{min}. This reduces to optimizing the download size from each server. Interestingly, the asymmetry of the download sizes fundamentally links the timely PIR problem to the PIR problem with asymmetric traffic constraints [32], which results in $N!\binom{N+M-1}{M}$ constraints on the achievable PIR rate. We consider two age metrics, namely, the peak-age metric and the average-age metric. For the case of $N = 2$ servers and $M = 3$ messages, we analytically characterize the optimal tradeoff between the minimum PIR rate R_{min} and AoI. The peak AoI problem is expressed as a linear program, while an inner-outer minimization procedure is used to solve the average AoI problem. In both cases, the optimal age-metric is a non-decreasing function of R_{min}. For the general problem with an arbitrary number of servers and messages, we show that 1) for peak AoI, a user downloads more bits from servers with lower expected delay, 2) if delay statistics are identical, there is a synergy between optimizing PIR rate and AoI.

II. The Timely PIR Problem

Consider a dynamic distributed storage setting with N replicated and non-colluding servers. Each server possesses a library of M i.i.d. messages of size L bits. We denote these messages by $\{W_m\}_{m=1}^M$. A user is interested in retrieving one of these messages, W_θ, without revealing the identity of the status update θ to any individual server. Following a completed retrieval process, the contents of the servers may change to a new set of messages, in an i.i.d. manner. In order to follow this changing nature, the user aims at retrieving its intended message in a *timely* manner while preserving privacy. Thus, a *status update* is received with each successful retrieval process.

Towards retrieving W_θ, the user submits a query $Q_n^{[\theta]}$ to server $n \in [N]$. Server n then truthfully responds with an answer string $A_n^{[\theta]}$, which is a deterministic function of the query $Q_n^{[\theta]}$ and the contents of the servers $\{W_m\}_{m=1}^M$, i.e.,

$$H\left(A_n^{[\theta]}\middle|Q_n^{[\theta]}, W_{1:M}\right) = 0, \quad n \in [N]. \tag{1}$$

In this work, we assume that the user incurs a negligible delay to convey the queries to the servers.[1] The returned answer strings are time-stamped by the server at the exact time instant of receiving the query.[2] The user incurs a random delay, $Z_{n,i}$ time units, in order to the receive the ith answer from server n. This models the total delay of the server including the processing and transmission delays, in addition to the propagation delay to the user. The random variables $\{Z_{n,i}\}_{i=1}^\infty$ are i.i.d. with mean μ_n and variance σ_n^2. The statistics of the server delays are known to the user.

The user designs queries that result in downloading d_n bits from server n. The value of d_n is fixed across different status updates, i.e., the same number of bits d_n are downloaded from server n whenever a new status update is sought. The answer strings from all servers are received through a single shared queue (see Fig. 1), and therefore the jth status update requires

$$T_j \triangleq \sum_{n=1}^{N} \sum_{i=(j-1)d_n+1}^{jd_n} Z_{n,i} \tag{2}$$

time units to be received in full. We denote by an *epoch* the time elapsed in between two consecutive status updates. Since $\{Z_{n,i}\}_{i=1}^\infty$ are i.i.d. and $\{d_n\}_{n=1}^N$ are fixed, it follows that the epoch times $\{T_j\}_{j=1}^\infty$ are i.i.d. $\sim T$. We assume that the servers do not update their message contents within the same epoch.

Furthermore, the user designs the queries such that they satisfy the following PIR constraints:

Correctness. The user should be able to reconstruct the message with no error given the answer strings, i.e.,

$$H\left(W_\theta\middle|A_{1:N}^{[\theta]}, Q_{1:N}^{[\theta]}\right) = 0. \tag{3}$$

Perfect Privacy. The queries should not leak any information about the identity of the message, i.e.,

$$I\left(\theta; Q_n^{[\theta]}\right) = 0, \quad n \in [N]. \tag{4}$$

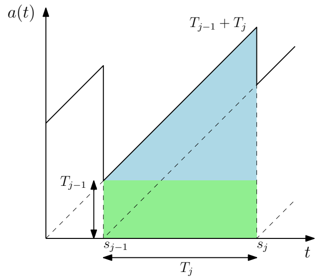

Fig. 2. An example evolution of AoI in epoch j.

Minimum Retrieval Rate. We use the retrieval rate as the performance metric of the retrieval scheme. The retrieval rate, R, is the ratio between the status update size L and the *expected*[3] total download cost D, i.e., $R = \frac{L}{\mathbb{E}[D]}$. The designed queries need to achieve a minimum retrieval rate $R_{\min}$,

$$R \geq R_{\min} \quad \Leftrightarrow \quad \mathbb{E}[D] \leq D_{\max} = \frac{L}{R_{\min}}. \tag{5}$$

We use the AoI metric to assess the timeliness of the received status updates. The AoI at time t is defined as

$$a(t) \triangleq t - u(t), \tag{6}$$

where $u(t)$ denotes the time stamp of the latest received status update before time t. We focus on minimizing two AoI metrics. The first is the long-term *peak* average AoI:

$$\texttt{pAoI} \triangleq \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{l(T)}\sum_{j=1}^{l(T)} a\left(s_j^-\right)\right], \tag{7}$$

where $l(T)$ denotes the number of status updates received by time T, and s_j^- denotes the time right before receiving the jth status update. The second is the long-term *time* average AoI:

$$\texttt{aAoI} \triangleq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T a(t)dt\right], \tag{8}$$

The expectations in (8) and (7) are taken over the distributions of the incurred delays, as manifested in (2). We focus on *zero-wait* policies in which a new status update is requested right after the previous one is received.[4]

In Fig. 2, we show an example of how the AoI may evolve during epoch j. From the figure, one can see that the epoch's peak AoI has a value of $T_{j-1} + T_j$. Since T_j's are i.i.d., one can show that (7) reduces to

$$\texttt{pAoI} = \mathbb{E}\left[T_{j-1} + T_j\right] = 2\mathbb{E}\left[T\right]. \tag{9}$$

Similarly, one can compute the area of epoch j and use the renewal-reward theorm [55] to show that (8) reduces to

$$\texttt{aAoI} = \frac{1}{\mathbb{E}[T_j]}\left[\mathbb{E}\left[T_{j-1}T_j\right] + \frac{1}{2}\mathbb{E}\left[T_j^2\right]\right] = \mathbb{E}\left[T\right] + \frac{\mathbb{E}\left[T^2\right]}{2\mathbb{E}\left[T\right]}. \tag{10}$$

[1] We focus on the case in which the sole source of the delay is in the downlink—while the servers are responding to the user. This assumption can be motivated by the fact that the upload cost in PIR problems does not scale with the message size L in contrast to the download cost [2]. Moreover, the queries can be reused if the user is interested in the same message.

[2] The servers' time stamps are the same since the upload delay is negligible.

[3] By measuring the download cost in the expected sense, we effectively allow for *mixed* retrieval strategies, i.e., we allow stochastic time sharing (across epochs) between pure retrieval strategies in this work.

[4] Zero-wait policies are generally optimal when minimizing `pAoI` but may be suboptimal when minimizing `aAoI` [40]. However, since the purpose of this work is to introduce the notion of timely PIR, we focus on zero-wait policies for simplicity of presentation.

We now formulate the **timely PIR problem** as follows:

$$\begin{aligned}\min_{\{d_n,\ Q_n^\theta\}_{n=1}^N} \quad & \texttt{pAoI} \text{ or } \texttt{aAoI} \\ \text{s.t.} \quad & H\left(A_n^{[\theta]}\middle|Q_n^{[\theta]}, W_{1:M}\right) = 0, \quad n \in [N] \\ & H\left(W_\theta\middle|A_{1:N}^{[\theta]}, Q_{1:N}^{[\theta]}\right) = 0 \\ & I\left(\theta; Q_n^{[\theta]}\right) = 0, \qquad\qquad n \in [N] \\ & R \geq R_{\min}. \end{aligned} \tag{11}$$

Observe that designing the queries $\{Q_n^\theta\}_{n=1}^N$ implicitly leads to the number of downloads from each server $\{d_n\}_{n=1}^N$ and the expected download cost $\mathbb{E}[D]$.

Problem (11) holds an interesting tension. For instance, if $R_{\min}$ allows it, the user may opt to retrieve its message from the server with the *most favorable statistics* (AoI-wise). However, this would come at the expense of *downloading more bits* to satisfy privacy. We shall see that it may be more rewarding to interact with servers with *individually* less favorable statistics so as to maintain privacy with a fewer number of downloads, and achieve an *overall* smaller AoI.

For a given server statistic vectors $\boldsymbol{\mu} \triangleq [\mu_1\, \mu_2 \cdots \mu_N]^T$ and $\boldsymbol{\sigma} \triangleq [\sigma_1^2\ \ \sigma_2^2\ \ \cdots\ \ \sigma_N^2]^T$, we aim at characterizing the optimal tradeoff between the long-term average AoI and the retrieval rate. More specifically, we aim at characterizing $\alpha^*(R_{\min})$, where $\alpha^*(\cdot)$ is the optimal value function of the optimization problem in (11) for all $\frac{1}{M} \leq R_{\min} \leq C_{\text{PIR}}$.[5]

III. Problem Re-Formulation: PIR with Asymmetric Traffic

In this section, we focus on disentangling the constraint set of problem (11). With a slight abuse of notation, we refer to the expected download cost by merely D. We will also allow for non-integer solutions of the download size vector $\mathbf{d} \triangleq [d_1\ \ \ d_2\ \ \ \cdots\ \ \ d_N]^T$, and overcome this by stochastic time sharing between well-defined PIR schemes. Time sharing will be shown optimal for the case of minimizing `pAoI`, and only slightly suboptimal for the case of minimizing `aAoI`.

We now express the AoI in terms of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\mathbf{d}$. Given the model in (2), and that epochs are i.i.d., one can write

$$\mathbb{E}[T] = \sum_{n=1}^N \sum_{i=1}^{d_n} \mathbb{E}[Z_{n,i}] = \sum_{n=1}^N \mu_n d_n = \boldsymbol{\mu}^T\mathbf{d}, \tag{12}$$

$$\mathbb{E}\left[T^2\right] = \sum_{n=1}^N \sigma_n^2 d_n + (\boldsymbol{\mu}^T\mathbf{d})^2 = \boldsymbol{\sigma}^T\mathbf{d} + (\boldsymbol{\mu}^T\mathbf{d})^2. \tag{13}$$

Based on this, we now have

$$\texttt{pAoI} = 2\boldsymbol{\mu}^T\mathbf{d}, \quad \texttt{aAoI} = \frac{3}{2}\boldsymbol{\mu}^T\mathbf{d} + \frac{1}{2}\frac{\boldsymbol{\sigma}^T\mathbf{d}}{\boldsymbol{\mu}^T\mathbf{d}}. \tag{14}$$

Observe that the AoI is solely controlled by the download size vector $\mathbf{d}$. Such vector must satisfy the constraint

$$\mathbf{1}^T\mathbf{d} = D. \tag{15}$$

Next, we focus on the PIR constraints. Since the user may download more data from one server relative to others, it becomes natural to consider PIR schemes with asymmetric traffic. The work in [32] characterizes the optimal retrieval rate $C(\boldsymbol{\tau})$ for an arbitrary server traffic ratio vector $\boldsymbol{\tau} = [\tau_1\quad \tau_2\quad \cdots\quad \tau_N]^T$, where $\tau_n = \frac{d_n}{D}$. Reference [32] provides an optimal retrieval scheme that satisfies the correctness and privacy constraints for monotonically decreasing traffic vector $\boldsymbol{\tau}$, such that $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_N$ for the cases of $M = 2, 3$. The results of [32] imply that there exist $\binom{M+N-1}{M}$ corner points corresponding to explicit achievable schemes. For any other traffic ratio vector, the optimal scheme is realized by time sharing between adjacent corner points. We focus in this work on the cases of $M = 2, 3$, for which the PIR capacity $C(\boldsymbol{\tau})$ is characterized for each $\boldsymbol{\tau}$ as follows [32]:

$$C(\boldsymbol{\tau}) = \begin{cases} \displaystyle\min_{n_0 \in [N]} \frac{1 + \frac{\sum_{n=n_0+1}^N \tau_n}{n_0}}{1 + \frac{1}{n_0}}, & M = 2 \\ \displaystyle\min_{n_0 \leq n_1 \in [N]} \frac{1 + \frac{\sum_{n=n_0+1}^N \tau_n}{n_0} + \frac{\sum_{n=n_1+1}^N \tau_n}{n_0 n_1}}{1 + \frac{1}{n_0} + \frac{1}{n_0 n_1}}, & M = 3 \end{cases} \tag{16}$$

Thus, the query design in addition to the PIR constraints in (11) reduce to choosing one of the corner points of [32] (or a convex mixture of them). Consequently, the general problem in (11) can be re-formulated by incorporating the capacity expressions for $M = 3$ in (16),

$$\begin{aligned}\min_{\mathbf{d}, D} \quad & \texttt{pAoI} \text{ or } \texttt{aAoI} \\ \text{s.t.} \quad & \mathbf{d} \geq 0, \quad \mathbf{1}^T\mathbf{d} = D, \quad D \leq D_{\max} \\ & \min_{\Pi([N])} \left\{ \sum_{n=n_0+1}^N d_{i_n} + \sum_{n=n_1+1}^N d_{i_n} \right\} \geq L\left(1 + \frac{1}{n_0} + \frac{1}{n_0 n_1}\right) - D, \\ & \qquad\qquad\qquad\qquad n_0 \leq n_1 \in [N], \end{aligned} \tag{17}$$

where $\Pi([N])$, is the set of all permutations of the index set $\{i_1, i_2, \cdots, i_N\}$ of the vector $[d_{i_1}\ d_{i_2}\ \ \cdots\ \ d_{i_N}]$. We include all the permutations as the optimal ordering is unknown in general. Note that for $M = 2$, the third constraint is replaced by $\min_{\Pi([N])} \left\{\sum_{n=n_0+1}^N d_{i_n}\right\} \geq L\left(1 + \frac{1}{n_0}\right) - D,\ \ n_0 \in [N]$. We focus on problem (17) in what follows.

IV. Timely PIR for $N = 2$ and $M = 3$

In this section, we solve problem (17) explicitly for the special case of $N = 2$ servers and $M = 3$ messages. In this case, we have 4 corner points corresponding to 4 different achievable schemes for PIR under asymmetric traffic constraints (see [32, Section V.A]). Note that the achievable schemes in [32, Section V.A] are constructed for different message sizes; specifically, L can be $1, 2, 4$, or 8 for $N = 2$ and $M = 3$. Since we formulate our problem for a fixed L,

[5] The achievable PIR rate is bounded below by $\frac{1}{M}$, which corresponds to the trivial scheme of downloading all messages from one server, and bounded above by the PIR capacity $C_{\text{PIR}} = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}}\right)^{-1}$ [2].

we choose it to be the least common multiple of all possible message sizes, i.e., we set $L = 8$ bits. This implies that the retrieval tables in [32, Section V.A] need to be repeated to match the chosen message size.

Furthermore, for $N = 2$ and $M = 3$, the third constraint in (17) can be explicitly written as

$$d_2 \geq \frac{3}{2}L - \frac{1}{2}D,\ d_2 \geq \frac{5}{2}L - D,\ D \geq \frac{7}{4}L. \tag{18}$$

Note that these constraints are written with the assumption that $d_1 \geq d_2$. For a general $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}$, this may not be optimal. Thus, we need to add two more constraints analogous to (18) after replacing d_2 by d_1, to cover the case $d_2 \geq d_1$.

A. Peak-Age Minimization Under Perfect Privacy

Problem (17) with the pAoI metric is now given by

$$\begin{aligned} \min_{\mathbf{d},D} \quad & 2\boldsymbol{\mu}^T\mathbf{d} \\ \text{s.t.} \quad & \mathbf{d} \geq 0,\ \mathbf{1}^T\mathbf{d} = D,\ (7/4)L \leq D \leq L/R_{\min},\ L = 8 \\ & d_n \geq \max\left\{\frac{3}{2}L - \frac{1}{2}D, \frac{5}{2}L - D\right\},\ n = 1, 2. \end{aligned} \tag{19}$$

Without loss of generality (WLOG), assume that $\mu_1 \leq \mu_2$. This implies that $d_1 \geq d_2$, since the constraint set is symmetric (cf. Lemma 1). The optimization problem (19) is a linear program (LP), whose solution resides at one of the corner points of the constraint set [56], which we categorize next.

1) $R_{\min} \in [\frac{1}{2}, \frac{4}{7}]$: In this case, we have 2 feasible corner points. The first is the corner point corresponding to the scheme [32, Table III] with $\tau_2 = \frac{3}{7}$, i.e., with $d_1 = 8$, and $d_2 = 6$.[6] This gives $\texttt{pAoI}(R_{\min}) \triangleq 2\boldsymbol{\mu}^T\mathbf{d} = 16\mu_1 + 12\mu_2$.

The second corner point results from the intersection of the constraints: $D \leq \frac{L}{R_{\min}}$ and $d_2 \geq \frac{5}{2}L - D$. This gives

$$d_2 = \frac{5}{2}L - \frac{L}{R_{\min}} = \left(\frac{5}{2} - \frac{1}{R_{\min}}\right)L, \tag{20}$$

$$d_1 = D - d_2 = \left(\frac{2}{R_{\min}} - \frac{5}{2}\right)L. \tag{21}$$

To achieve such d_1, d_2, we employ stochastic time sharing between schemes [32, Table III] and [32, Table IV], i.e., at the start of each epoch, the user randomly applies the scheme [32, Table III] with probability $\theta = \frac{4}{R_{\min}} - 7$ and the scheme [32, Table IV] with probability $1 - \theta$. This gives $\texttt{pAoI}(R_{\min}) = 16\left[\frac{5}{2}(\mu_2 - \mu_2) + \frac{1}{R_{\min}}(2\mu_1 - \mu_2)\right]$. Consequently, for $R_{\min} \in [\frac{1}{2}, \frac{4}{7}]$, the solution of problem (19) is

$$\begin{aligned} \alpha^*(R_{\min}) = \min\Bigg\{ & 16\mu_1 + 12\mu_2, \\ & 16\left[\frac{5}{2}(\mu_2 - \mu_2) + \frac{1}{R_{\min}}(2\mu_1 - \mu_2)\right]\Bigg\}. \end{aligned} \tag{22}$$

We note that the scheme [32, Table II] is strictly sub-optimal since it achieves the same PIR rate of [32, Table III], which is $\frac{4}{7}$, while incurring stricltly higher pAoI.

2) $R_{\min} \in [\frac{1}{3}, \frac{1}{2}]$: In this case, in addition to the corner point of the scheme [32, Table III], we have two more corner points. The first is that corresponding to the scheme [32, Table IV] with $\tau_2 = \frac{1}{4}$, i.e., with $d_1 = 12$, and $d_2 = 4$. This gives $\texttt{pAoI}(R_{\min}) = 24\mu_1 + 8\mu_2$.

The second corner point results from the intersection of the constraints: $D \leq \frac{L}{R_{\min}}$ and $d_2 \geq \frac{3}{2}L - \frac{1}{2}D$. This gives

$$d_2 = \frac{3}{2}L - \frac{1}{2}\frac{L}{R_{\min}} = \left(\frac{3}{2} - \frac{1}{2R_{\min}}\right)L, \tag{23}$$

$$d_1 = D - d_2 = \left(\frac{3}{2R_{\min}} - \frac{3}{2}\right)L. \tag{24}$$

This is achieved by stochastic time sharing between the scheme [32, Table I] with probability $\theta = \frac{1}{R_{\min}} - 2$ and the scheme [32, Table IV] with probability $(1 - \theta)$. This gives $\texttt{pAoI}(R_{\min}) = 8\left[3(\mu_2 - \mu_1) + \frac{1}{R_{\min}}(3\mu_1 - \mu_2)\right]$. Consequently, for $R_{\min} \in [\frac{1}{3}, \frac{1}{2}]$, the solution of problem (19) is

$$\begin{aligned} \alpha^*(R_{\min}) = \min\Bigg\{ & 16\mu_1 + 12\mu_2, 24\mu_1 + 8\mu_2, \\ & 8\left[3(\mu_2 - \mu_1) + \frac{1}{R_{\min}}(3\mu_1 - \mu_2)\right]\Bigg\}. \end{aligned} \tag{25}$$

B. Average-Age Minimization Under Perfect Privacy

The aAoI expression in (14) has a linear fractional term. To deal with this, we use the Charnes-Cooper transformation in [57]. More specifically, we define $t \triangleq \frac{1}{\boldsymbol{\mu}^T\mathbf{d}}$, and change the optimization variable to $\mathbf{x} \triangleq \mathbf{d} \cdot t$. Therefore, we now have

$$\texttt{aAoI} = \frac{3}{2}\frac{1}{t} + \frac{1}{2}\boldsymbol{\sigma}^T\mathbf{x} \tag{26}$$

We note that this is a convex function in $(\boldsymbol{x}, t)$ [56]. The constraint set after transformation becomes

$$\begin{aligned} \mathcal{X} \triangleq \Bigg\{ & (\mathbf{x}, t, D) : \mathbf{x} \geq 0, \mathbf{1}^T\mathbf{x} = Dt, \boldsymbol{\mu}^T\mathbf{x} = 1, x_n \geq \frac{3}{2}Lt - \frac{1}{2}Dt, \\ & x_n \geq \frac{5}{2}Lt - Dt, \frac{7}{4}L \leq D \leq \frac{L}{R_{\min}}\Bigg\}. \end{aligned} \tag{27}$$

The constraint set is convex for fixed D. This suggests using inner-outer minimization techniques, in which the inner minimization is over $(\mathbf{x}, t)$ for a fixed D, and the outer minimization is over D. The inner problem is a convex problem, while the outer problem can be solved by a one-dimensional line search over $\frac{7}{4}L \leq D \leq \frac{L}{R_{\min}}$.

For the inner minimization, we first note that since $R_{\min} \geq \frac{1}{3}$ holds in our case, the largest value that D takes is $3L$, and hence the PIR constraint $x_n \geq \frac{3}{2}Lt - \frac{1}{2}Dt$ implies $\mathbf{x} \geq 0$ and makes it redundant. As in the pAoI case, we assume $\mu_1 \leq \mu_2$. Next, we consider the two equality constraints in $\mathcal{X}$. These yield a unique solution for $\mathbf{x}$ provided that $\mu_1 \neq \mu_2$, which is given as follows:

$$x_1^* = \frac{\mu_2 Dt - 1}{\mu_2 - \mu_1}, \quad x_2^* = \frac{1 - \mu_1 Dt}{\mu_2 - \mu_1}. \tag{28}$$

Substituting the above in the objective function, taking deriva-

[6] Details of the PIR scheme tables in [32] are omitted due to space limits.

tive with respect to t and equating it to 0 gives

$$t^*(D) = \sqrt{3\frac{\mu_2-\mu_1}{\sigma_1^2\mu_2-\sigma_2^2\mu_1}\frac{1}{D}}. \tag{29}$$

However, for this approach to be valid, we must have $t^*(D) \in \mathbb{R}_{++}$ above. Assuming this is the case, it now remains to find the optimal D^* while satisfying the PIR constraints. Towards that end, we substitute $t^*(D)$ in both the objective function and the constraints, and solve the following outer minimization:

$$\begin{aligned} \min_{D} \quad & \sqrt{3\frac{\sigma_1^2\mu_2-\sigma_2^2\mu_1}{\mu_2-\mu_1}D}+\frac{1}{2}\frac{\sigma_2^2-\sigma_1^2}{\mu_2-\mu_1} \\ \text{s.t.} \quad & \min\left\{\mu_2 D-\frac{1}{t^*(D)}, \frac{1}{t^*(D)}-\mu_1 D\right\} \\ & \qquad \geq (\mu_2-\mu_1)\max\left\{\frac{3}{2}L-\frac{1}{2}D, \frac{5}{2}L-D\right\} \\ & \frac{7}{4}L \leq D \leq \frac{L}{R_{\min}}, \ L=8. \end{aligned} \tag{30}$$

The above shows that aAoI scales with $\sqrt{D}$ (note that $\sigma_1^2\mu_2 > \sigma_2^2\mu_1 > 0$; otherwise $t^*(D) \notin \mathbb{R}_{++}$). The solution of the outer minimization, D^*, is given by the least value of D satisfying the constraint set of problem (30). Therefore,

$$\alpha^*(R_{\min}) = \sqrt{3\frac{\sigma_1^2\mu_2-\sigma_2^2\mu_1}{\mu_2-\mu_1}D^*}+\frac{1}{2}\frac{\sigma_2^2-\sigma_1^2}{\mu_2-\mu_1}. \tag{31}$$

We now consider the case in which $\mu_1 = \mu_2 \triangleq \mu$. In this case, the equality constraints in $\mathcal{X}$ do not yield a unique solution for $\mathbf{x}$; they would instead impose that $t^*(D) = \frac{1}{\mu D}$. Assuming WLOG that $\sigma_1^2 \leq \sigma_2^2$, further manipulations would lead to the following outer problem (derivation details are omitted due to space limits):

$$\begin{aligned} \min_{D} \quad & \frac{3}{2}\mu D+\frac{\sigma_2^2-\sigma_1^2}{\mu}\max\left\{\frac{3}{2}\frac{L}{D}-\frac{1}{2}, \frac{5}{2}\frac{L}{D}-1\right\}+\frac{\sigma_1^2}{\mu} \\ \text{s.t.} \quad & \frac{7}{4}L \leq D \leq \frac{L}{R_{\min}}, \quad L=8. \end{aligned} \tag{32}$$

which is a convex problem in D that can be solved, e.g., by a bisection search to give $\alpha^*(R_{\min})$.

Finally, if the value of t^* in (29) does not yield a positive real solution, or if problem (30) is not feasible, then this means that the two servers cannot be simultaneously active. This would require $R_{\min} = \frac{1}{3}$, in which case $d_{n^*} = D = 3L$, where $n^* \triangleq \arg\min \frac{9}{2}\mu_n L + \frac{1}{2}\frac{\sigma_n^2}{\mu_n}$, and

$$\alpha^*(R_{\min}) = \frac{9}{2}\mu_{n^*}L + \frac{1}{2}\frac{\sigma_{n^*}^2}{\mu_{n^*}}. \tag{33}$$

V. Extending to Arbitrary Number of Servers

For the case with an arbitrary number of servers, we observe that the number of PIR constraints grows as $N!\binom{N+M-1}{M}$ to encapsulate all possible permutations of the servers. This deems an analytic resolution of the problem extremely challenging. Nevertheless, we have the following structural results (proofs are omitted due to space limits):

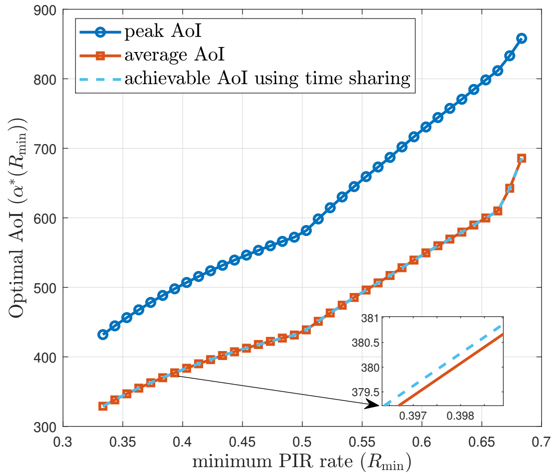

Fig. 3. Optimal tradeoff between AoI and the PIR rate for $N = 3$, $M = 3$, $L = 72$, with the servers having $\boldsymbol{\mu} = [1\ 5\ 10]$ and $\boldsymbol{\sigma} = [10\ 5\ 1]$

Lemma 1 *In problem (17) with pAoI, if the mean delays are increasing, $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_N$, then the download sizes are decreasing, $d_1 \geq d_2 \geq \cdots \geq d_N$.*

Lemma 2 *In problem (17) with pAoI or aAoI, if $\mu_n = \mu$, $\sigma_n^2 = \sigma^2$, $\forall n \in [N]$, then $d_n = d = \frac{L}{NC_{PIR}}$, $\forall n \in [N]$ and the symmetric Sun-Jafar PIR scheme in [2] is optimal.*

Lemma 2 shows that for symmetric server statistics, load balancing, which is known to maximize the PIR rate is optimal. Consequently, *symmetric statistics produces a synergy between maximizing the PIR rate and minimizing AoI.*

In Fig. 3, we show how the optimal AoI $\alpha^*(R_{\min})$ behaves with the minimum PIR rate $R_{\min}$ for the case of $N = 3$ servers, $M = 3$ messages, and $L = 72$ bits. One can observe a clear tradeoff between AoI and the PIR rate, and that time sharing for aAoI has a negligible performance loss.

VI. Conclusions

In this paper, we introduced the timely PIR problem, in which a user wishes to privately retrieve a message in timely fashion from N replicated and non-colluding servers, whose M messages are continuously updating. We showed that the query design problem in this case reduces to choosing the download sizes from each server, tying the problem to PIR under asymmetric traffic. The optimal tradeoff between the PIR rate and (peak and average) AoI has then been studied.

Several extensions can be pursued. First, reducing the number of constraints by identifying the optimal ordering of the servers and/or the trajectory of activating the servers as a function of $R_{\min}$. Second, designing waiting policies for the average AoI case. Third, considering different models for the servers including unresponsiveness, which requires employing robust PIR schemes as in [3], [4]. Fourth, proposing explicit PIR schemes to avoid the extra AoI penalty due to time sharing for the average AoI case. Finally, treating the message size L as a control variable by which the user may receive shorter messages at a lower resolution versus awaiting fixed size ones.

References

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, November 1998.

[2] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. Inf. Theory*, 63(7):4075–4088, July 2017.

[3] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. *IEEE Trans. on Info. Theory*, 64(4):2361–2370, April 2018.

[4] R. Bitar and S. El Rouayheb. Staircase-PIR: Universally robust private information retrieval. In *IEEE ITW*, pages 1–5, November 2018.

[5] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Transactions on Information Theory*, 65(1):322–329, January 2019.

[6] T. Guo, R. Zhou, and C. Tian. On the information leakage in private information retrieval systems. Available at arXiv: 1909.11605.

[7] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. on Info. Theory*, 64(10):6842–6862, October 2018.

[8] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. on Info. Theory*, 65(2):1206–1219, February 2019.

[9] R. Tandon. The capacity of cache aided private information retrieval. In *Allerton Conference*, October 2017.

[10] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *IEEE Trans. on Info. Theory*, 65(5):3215–3232, May 2019.

[11] S. Kumar, A. G. i Amat, E. Rosnes, and L. Senigagliesi. Private information retrieval from a cellular network with caching at the edge. *IEEE Trans. on Communications*, 67(7):4900–4912, July 2019.

[12] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. *IEEE Trans. on Info. Theory*, 66(4):2032–2043, April 2020.

[13] Z. Chen, Z. Wang, and S. Jafar. The capacity of T-private information retrieval with private side information. Available at arXiv:1709.03022.

[14] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. *IEEE Trans. on Info. Theory*, 65(12):8222–8231, December 2019.

[15] S. P. Shariatpanahi, M. J. Siavoshani, and M. A. Maddah-Ali. Multi-message private information retrieval with private side information. In *IEEE ITW*, pages 1–5, November 2018.

[16] S. Li and M. Gastpar. Converse for multi-server single-message PIR with side information. Available at arXiv:1809.09861.

[17] H. Sun and S. A. Jafar. The capacity of private computation. *IEEE Trans. on Info. Theory*, 65(6):3880–3897, June 2019.

[18] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. In *IWCIT*, pages 1–6, April 2018.

[19] Z. Chen, Z. Wang, and S. Jafar. The asymptotic capacity of private search. In *IEEE ISIT*, June 2018.

[20] M. A. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. Available at arXiv:1805.04104v2.

[21] C. Tian, H. Sun, and J. Chen. Capacity-achieving private information retrieval codes with optimal message size and upload cost. *IEEE Trans. on Info. Theory*, 65(11):7613–7627, Nov 2019.

[22] K. Banawan, B. Arasli, and S. Ulukus. Improved storage for efficient private information retrieval. In *IEEE ITW*, August 2019.

[23] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus. The capacity of private information retrieval from decentralized uncoded caching databases. *Information*, 10, December 2019.

[24] K. Banawan, B. Arasli, Y. P. Wei, and S. Ulukus. The capacity of private information retrieval from heterogeneous uncoded caching databases. *IEEE Trans. on Info. Theory*, 66(6):3407–3416, 2020.

[25] K. Banawan and S. Ulukus. Private information retrieval from non-replicated databases. In *IEEE ISIT*, pages 1272–1276, July 2019.

[26] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. on Info. Theory*, 66(7):4129–4149, 2020.

[27] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. *IEEE Trans. on Info. Forensics and Security*, 12(12):2920–2932, December 2017.

[28] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. *IEEE Trans. on Info. Theory*, 65(5):3198–3214, May 2019.

[29] H. Yang, W. Shin, and J. Lee. Private information retrieval for secure distributed storage systems. *IEEE Trans. on Info. Forensics and Security*, 13(12):2953–2964, December 2018.

[30] Z. Jia, H. Sun, and S. Jafar. Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval. *IEEE Trans. on Info. Theory*, 65(9):5783–5798, September 2019.

[31] R. Zhou, C. Tian, H. Sun, and T. Liu. Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size. Available at arXiv: 1903.08229.

[32] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric-traffic constraints. *IEEE Trans. Inf. Theory*, 65(11):7628–7645, November 2019.

[33] K. Banawan and S. Ulukus. Noisy private information retrieval: On separability of channel coding and information retrieval. *IEEE Trans. on Info. Theory*, 65(12):8232–8249, December 2019.

[34] R. Tajeddine, A. Wachter-Zeh, and C. Hollanti. Private information retrieval over random linear networks. Available at arXiv:1810.08941.

[35] Z. Wang, K. Banawan, and S. Ulukus. Private set intersection: A multi-message symmetric private information retrieval perspective. Available at arXiv: 1912.13501.

[36] Z. Wang, K. Banawan, and S. Ulukus. Multi-party private set intersection: An information-theoretic approach. Available at arXiv: 2008.07504.

[37] S. K. Kaul, R. D. Yates, and M. Gruteser. Real-time status: How often should one update? In *Proc. IEEE Infocom*, March 2012.

[38] C. Kam, S. Kompella, and A. Ephremides. Age of information under random updates. In *Proc. IEEE ISIT*, July 2013.

[39] Y. Hsu, E. Modiano, and L. Duan. Age of information: Design and analysis of optimal scheduling algorithms. In *Proc. IEEE ISIT*, June 2017.

[40] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff. Update or wait: How to keep your data fresh. *IEEE Trans. Inf. Theory*, 63(11):7492–7508, November 2017.

[41] X. Wu, J. Yang, and J. Wu. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Trans. Green Commun. Netw.*, 2(1):193–204, March 2018.

[42] P. Mayekar, P. Parag, and H. Tyagi. Optimal lossless source codes for timely updates. In *Proc. IEEE ISIT*, June 2018.

[43] A. Baknina, O. Ozel, J. Yang, S. Ulukus, and A. Yener. Sending information through status updates. In *Proc. IEEE ISIT*, June 2018.

[44] B. Zhou and W. Saad. Optimal sampling and updating for minimizing age of information in the internet of things. In *Proc. IEEE Globecom*, December 2018.

[45] R. D. Yates and S. K. Kaul. The age of information: Real-time status updating by multiple sources. *IEEE Trans. Inf. Theory*, 65(3):1807–1827, March 2019.

[46] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu. Optimal status updating with a finite-battery energy harvesting source. *J. Commun. Netw.*, 21(3):280–294, June 2019.

[47] M. Zhang, A. Arafa, J. Huang, and H. V. Poor. How to price fresh data. In *Proc. WiOpt*, June 2019.

[48] A. M. Bedewy, Y. Sun, and N. B. Shroff. The age of information in multihop networks. *IEEE/ACM Trans. Netw.*, 27(3):1248–1257, June 2019.

[49] R. Talak and E. Modiano. Age-delay tradeoffs in single server systems. In *Proc. IEEE ISIT*, July 2019.

[50] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor. Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies. *IEEE Trans. Inf. Theory*, 66(1):534–556, January 2020.

[51] H. H. Yang, A. Arafa, T. Q. S. Quek, and H. V. Poor. Age-based scheduling policy for federated learning in mobile edge networks. In *Proc. IEEE ICASSP*, May 2020.

[52] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor. Timely estimation using coded quantized samples. In *Proc. IEEE ISIT*, June 2020.

[53] T. Z. Ornee and Y. Sun. Sampling for remote estimation through queues: Age of information and beyond. Available Online: arXiv:1902.03552.

[54] R. D. Yates, Y. Sun, R. D. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus. Age of information: Introduction and survey. Available Online: arXiv:2007.08564.

[55] S. M. Ross. *Stochastic processes*. Wiley New York, 1996.

[56] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[57] A. Charnes and W. W. Cooper. Programming with linear fractional functionals. *Naval Research logistics quarterly*, 9(3-4):181–186, 1962.