Timely Private Information Retrieval

Karim Banawan\textsuperscript{1}, Ahmed Arafa\textsuperscript{2}, and Sennur Ulukus\textsuperscript{3}

\textsuperscript{1}Electrical Engineering Department, Faculty of Engineering, Alexandria University
\textsuperscript{2}Electrical and Computer Engineering Department, University of North Carolina at Charlotte
\textsuperscript{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?

In this paper, we introduce the timely PIR problem. A user aims at correct retrieval of a message $W_\theta$ (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 $\theta$ to any individual server and achieve a minimum PIR rate of $R_{\text{min}}$. 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 $R_{\text{min}}$ 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 $R_{\text{min}}$. 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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{The timely PIR problem.}
\end{figure}
messages by \( \{W_m\}_{m=1}^M \). A user is interested in retrieving one of these messages, \( W_\theta \), without revealing the identity of the status update \( \theta \) 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 \( W_\theta \), the user submits a query \( Q_\theta^{[n]} \) to server \( n \in \{N\} \). Server \( n \) then truthfully responds with an answer string \( A_\theta^{[n]} \), which is a deterministic function of the query \( Q_\theta^{[n]} \) and the contents of the servers \( \{W_m\}_{m=1}^M \), i.e.,

\[
H\left(A_\theta^{[n]}|Q_\theta^{[n]}, W_{1:M}\right) = 0, \quad n \in \{N\}.
\]

(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 receive the \( i \)th 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 \( \mu_n \) and variance \( \sigma_n^2 \). 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 \( j \)th status update requires

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

(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. ~ \( T \). 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.,

\[
H\left(W_\theta|A_1^{[\cdot]}, Q_\theta^{[\cdot]} \right) = 0.
\]

(3)

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

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

(4)

\(^1\)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 \( 2L \). 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.

**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 total download cost \( D \), i.e., \( R = \frac{L}{\mathbb{E}[D]} \). The designed queries need to achieve a minimum retrieval rate \( R_{\text{min}} \),

\[
R \geq R_{\text{min}} \iff \mathbb{E}[D] \leq D_{\text{max}} = \frac{L}{R_{\text{min}}}.
\]

(5)

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

\[
a(t) \triangleq t - u(t),
\]

(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:

\[
p\text{AoI} \triangleq \lim_{T \rightarrow \infty} \mathbb{E} \left[ \frac{1}{l(T)} \sum_{j=1}^{l(T)} a(s_j^-) \right],
\]

(7)

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

\[
a\text{AoI} \triangleq \lim_{T \rightarrow \infty} \mathbb{E} \left[ \frac{1}{T} \int_0^T a(t)dt \right],
\]

(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

\[
p\text{AoI} = \mathbb{E}[T_{j-1} + T_j] = 2\mathbb{E}[T].
\]

(9)

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

\[
a\text{AoI} = \frac{1}{\mathbb{E}[T]} \left[ \mathbb{E}[T_{j-1} T_j] + \frac{1}{2} \mathbb{E}[T_j^2] \right] = \mathbb{E}[T] + \frac{\mathbb{E}[T^2]}{2\mathbb{E}[T]}.
\]

(10)

\(^3\)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 \( p\text{AoI} \) but may be suboptimal when minimizing \( a\text{AoI} \) [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:

\[
\min_{\{d_n, Q_n^{|}\}_{n=1}^{N}} \ p_{\text{AoI}} \text{ or } a_{\text{AoI}} \\
\text{s.t. } H\left( A^{[\theta]}_n, Q_n^{|} \right) = 0, \ n \in [N] \\
H\left( W_0, A^{[\theta]}_1, Q^{|}_1 \right) = 0 \\
I\left( \theta; Q_n^{|} \right) = 0, \ n \in [N] \\
R \geq R_{\text{min}}. \tag{11}
\]

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

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

III. PROBLEM RE-FORMULATION: PIR WITHASYMMETRIC 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 \( d \triangleq [d_1, d_2, \cdots, d_N]^T \), and overcome this by stochastic time sharing between well-defined PIR schemes. Time sharing will be shown optimal for the case of minimizing \( p_{\text{AoI}} \), and only slightly suboptimal for the case of minimizing \( a_{\text{AoI}} \).

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

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

\[
E[T^2] = \frac{\sum_{n=1}^{N} \sigma_n^2 d_n + (\mu^T d)^2}{\sum_{n=1}^{N} \mu^2 d_n} = \sigma^T d + (\mu^T d)^2. \tag{13}
\]

Based on this, we now have

\[
p_{\text{AoI}} = 2\mu^T d, \quad a_{\text{AoI}} = \frac{3}{2} \mu^T d + \frac{1}{2} \sigma^T d. \tag{14}
\]

Observe that the AoI is solely controlled by the download size vector \( d \). Such vector must satisfy the constraint

\[
1^T d = D. \tag{15}
\]

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 \( C(\tau) \) for an arbitrary server traffic ratio vector \( \tau = [\tau_1, \tau_2, \cdots, \tau_N]^T \), 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 \( \tau \), 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 \( \left( M + N - 1 \right) \) 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 \( C(\tau) \) is characterized for each \( \tau \) as follows [32]:

\[
C(\tau) = \begin{cases} 
1 + \sum_{n=0}^{N-1} \frac{\tau_n}{n}, & M = 2 \\
1 + \frac{1}{n_0} + \frac{1}{n_0 n_1}, & M = 3 
\end{cases} 
\]

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),

\[
\min_{d,D} \ p_{\text{AoI}} \text{ or } a_{\text{AoI}} \\
\text{s.t. } d \geq 0, \quad 1^T d = D, \quad D \leq D_{\text{max}} \\
\min_{\Pi([N])} \left\{ \sum_{n=n_0+1}^{N} \frac{d_n}{n}, \sum_{n=n_1+1}^{N} \frac{d_n}{n} \right\} \geq L \left( 1 + \frac{1}{n_0} + \frac{1}{n_0 n_1} \right) - D, \quad n_0 \leq n_1 \in [N], \tag{17}
\]

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 \( \min_{\Pi([N])} \left\{ \sum_{n=n_0+1}^{N} d_n \right\} \geq L \left( 1 + \frac{1}{n_0} \right) - D, \quad n_0 \in [N] \). 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 \( L \),
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

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

Note that these constraints are written with the assumption that \( d_1 \geq d_2 \). For a general \( \mu \), \( \sigma \), 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 \( \text{pAoI} \) metric is now given by

\[
\min_{d, D} 2 \mu^T d
\]

s.t. \( d \geq 0, \quad 1^T d = D, \quad (7/4) L \leq D \leq L / R_{\text{min}}, \quad L \geq 8 \)

\[
d_n \geq \max \left\{ \frac{3}{2} L - \frac{1}{2} D, \frac{5}{2} L - D \right\}, \quad n = 1, 2. \tag{19}
\]

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

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

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

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

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 \( \theta = \frac{1}{R_{\text{min}}} - 7 \) and the scheme [32, Table IV] with probability \( 1 - \theta \). This gives \( \text{pAoI}(R_{\text{min}}) = 16 \left( \frac{5}{2} \mu_2 - \mu_2 \right) + \frac{1}{R_{\text{min}}} (2 \mu_1 - \mu_2) \). Consequently, for \( R_{\text{min}} \in \left[ \frac{1}{2}, \frac{3}{2} \right] \), the solution of problem (19) is

\[
\alpha^*(R_{\text{min}}) = \min \left\{ 16 \mu_1 + 12 \mu_2, \right. \]

\[
16 \left( \frac{5}{2} \mu_2 - \mu_2 \right) + \frac{1}{R_{\text{min}}} (2 \mu_1 - \mu_2) \} \right. \tag{22}
\]

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 \( \frac{1}{4} \), while incurring strictly higher \( \text{pAoI} \).

2) \( R_{\text{min}} \in \left[ \frac{1}{2}, \frac{3}{4} \right] \): 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 \( \text{pAoI}(R_{\text{min}}) = 24 \mu_1 + 8 \mu_2 \).

The second corner point results from the intersection of the constraints: \( D \leq \frac{L}{R_{\text{min}}} \) and \( d_2 \geq \frac{5}{2} L - \frac{1}{2} D \). This gives

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

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

This is achieved by stochastic time sharing between the scheme [32, Table I] with probability \( \theta = \frac{1}{R_{\text{min}}} - 2 \) and the scheme [32, Table IV] with probability \( (1 - \theta) \). This gives

\[
\text{pAoI}(R_{\text{min}}) = 8 \left( \frac{3 \mu_2 - \mu_1}{} + \frac{1}{R_{\text{min}}} (3 \mu_1 - \mu_2) \right). \tag{25}
\]

**B. Average-Age Minimization Under Perfect Privacy**

The \( \text{aAoI} \) expression in (14) has a linear fractional term. To deal with this, we use the Charnes-Cooper transformation in [57]. More specifically, we define \( t = \frac{1}{\mu^2 - d^2} \), and change the optimization variable to \( x = d \cdot t \). Therefore, we now have

\[
\text{aAoI} = \frac{31}{2} + \frac{1}{2} \sigma^T x \tag{26}
\]

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

\[
X = \left\{ (x, t, D): x \geq 0, \quad 1^T x = D t, \quad \mu^T x = 1, x_n \geq \frac{3}{4} L t - \frac{1}{2} D t, \right.
\]

\[
\left. x_n \geq \frac{5}{4} L t - D t, \quad \frac{7}{4} L \leq D \leq \frac{L}{R_{\text{min}}} \right\}. \tag{27}
\]

The constraint set is convex for fixed \( D \). This suggests using inner-outer minimization techniques, in which the inner minimization is over \( (x, t) \) 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 \( \frac{1}{4} L \leq D \leq \frac{L}{R_{\text{min}}} \).

For the inner minimization, we first note that since \( R_{\text{min}} \geq \frac{1}{2} \) holds in our case, the largest value that \( D \) takes is \( 3L \), and hence the PIR constraint \( x_n \geq \frac{3}{4} L t - \frac{1}{2} D t \) implies \( x \geq 0 \) and makes it redundant. As in the \( \text{pAoI} \) case, we assume \( \mu_1 \leq \mu_2 \). Next, we consider the two equality constraints in \( X \). These yield a unique solution for \( x \) provided that \( \mu_1 \neq \mu_2 \), which is given as follows:

\[
x^*_1 = \frac{\mu_2 D t - \frac{1}{\mu_2 - \mu_1}, \quad x^*_2 = \frac{1}{\mu_2} - \mu_1 D t. \tag{28}
\]

Substituting the above in the objective function, taking derivat-
tive with respect to $t$ and equating it to 0 gives

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

However, for this approach to be valid, we must have $t^*(D) \in \mathbb{R}_+$. 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:

$$\min_{D} \left\{ \mu_2 D - \frac{1}{t^*(D)} \right\} \text{s.t.} \left\{ \mu_2 D - \frac{1}{t^*(D)} D - \mu_1 D \right\} \geq (\mu_2 - \mu_1) \max \left\{ \frac{3}{2} L - \frac{1}{2} D, \frac{5}{2} L - D \right\}$$

$$\frac{7}{4} L \leq D \leq \frac{L}{R_{\min}}, \quad L = 8.$$ (30)

The above shows that $\sigma^2$ scales with $\sqrt{D}$ (note that $\sigma_2^2 \mu_2 > \sigma_2^2 \mu_1 > 0$; 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,

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

We now consider the case in which $\mu_1 = \mu_2 \neq \mu$. In this case, the equality constraints in $\mathcal{X}$ do not yield a unique solution for $x$; 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):

$$\min_{D} \frac{3}{2} \mu D + \frac{\sigma_2^2 - \sigma_1^2}{\mu} \max \left\{ \frac{3}{2} L - \frac{1}{2} D, \frac{5}{2} L - 1 \right\} + \frac{\sigma_2^2}{\mu} \mu,$$

$$\text{s.t.} \quad \frac{7}{4} L \leq D \leq \frac{L}{R_{\min}}, \quad L = 8.$$ (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 $R_{\min} = \frac{1}{5}$, in which case $d_{n^*} = D = 3L$, where $n^* = \arg\min \frac{9}{2\mu_{n^*} L} + \frac{1}{2 \mu_{n^*}}$, and

$$\alpha^*(R_{\min}) = \frac{9}{2 \mu_{n^*} L} + \frac{1}{2 \mu_{n^*}}.$$ (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 $\mathcal{O}(N^2 M)$ 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 $N = 3$ and $M = 3$, $L = 72$, with the servers having $\mu = [1 5 10]$ and $\sigma = [10 5 1]$. The solid line shows how the optimal $\alpha^*(R_{\min})$ 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 $\alpha_{\text{AoI}}$ 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.
