Conference Paper

Lengthening and Extending Binary Private Information Retrieval Codes

Author(s):
Lin, Hsuan-Yin; Rosnes, Eirik

Publication Date:
2018-02

Permanent Link:
https://doi.org/10.3929/ethz-b-000245077 →

Lengthening and Extending Binary Private Information Retrieval Codes

Hsuan-Yin Lin and Eirik Rosnes
Simula@UiB, N–5020 Bergen, Norway
(Emails: hsuan-yin.lin@ieee.org and eirikrosnes@simula.no)

***Abstract*—It was recently shown by Fazeli *et al.* that the storage overhead of a traditional t-server private information retrieval (PIR) protocol can be significantly reduced using the concept of a *t-server PIR code*. In this work, we show that a family of t-server PIR codes (with increasing dimensions and blocklengths) can be constructed from an existing t-server PIR code through lengthening by a single information symbol and code extension by at most $\lceil t/2 \rceil$ code symbols. Furthermore, by extending a code construction notion from Steiner systems by Fazeli *et al.*, we obtain a specific family of t-server PIR codes. Based on a code construction technique that lengthens and extends a t-server PIR code simultaneously, a basic algorithm to find good (i.e., small blocklength) t-server PIR codes is proposed. For the special case of $t = 5$, we find provably optimal PIR codes for code dimensions $k \leq 6$, while for all $7 \leq k \leq 32$ we find codes of smaller blocklength than the best known codes from the literature. Furthermore, in the case of $t = 8$, we also find better codes for $k = 5, 6, 11, 12$. Numerical results show that most of the best found 5-server PIR codes can be constructed from the proposed family of codes connected to Steiner systems.**

I. Introduction

Private information retrieval (PIR) has attracted significant attention for well over a decade since its introduction by Chor *et al.* in [1]. A formal PIR protocol allows to privately retrieve a single file among the servers storing it without revealing any information about the requested file to each individual server. Traditional PIR protocols operate on a database of n bits, which is replicated among several servers to achieve PIR. Thus, the storage overhead of traditional PIR protocols is at least 2, and the overall goal is to reduce the total upload and download cost of the protocol.

PIR for distributed storage systems was first addressed in [2]. For distributed storage systems the size of the requested file is typically much larger than the number of files, and thus the upload cost is much lower than the download cost. Hence, only the download cost is considered, as opposed to traditional PIR protocols. Recent work on PIR protocols for distributed storage systems typically assumes that the storage code is given, and then the PIR protocol is designed as a second layer to the system [3], [4]. This is in contrast to the work by Fazeli *et al.* in [5], where, in order to reduce the storage overhead of traditional PIR protocols, the concept of a *t-server PIR code* was proposed. A t-server PIR code is an $[n, k]$ linear code satisfying the so-called t-PIR property, i.e., for every information symbol, there exist t mutually disjoint subsets of $\{1, 2, \ldots, n\}$ such that it can be recovered from the code symbols indexed by any of these t subsets. By employing an $[n, k]$ t-server PIR code, they have shown that all known t-server information-theoretic PIR protocols can be emulated by a coded PIR protocol with storage overhead equal to n/k.

Finding good codes that operate efficiently with a small storage overhead, i.e., designing a t-server PIR code with a small blocklength for a given dimension, is an important research challenge. In [5], an insightful series of t-server PIR code constructions based on existing code construction techniques were presented. In the recent work of [6], the authors found that the so-called *shortened projective Reed Muller (SPRM)* codes are good t-server PIR codes for $t = 2^\ell - 1$ and 2^ℓ where ℓ is a positive integer. For $t = 3, 4$, it was shown in [6] that SPRM codes are indeed optimal in the sense of achieving a lower bound on the blocklength of a t-server PIR code.

In this work, we will show that a t-server PIR code with small blocklength can be constructed by lengthening and extending an existing PIR code. Furthermore, we prove that a certain family of codes associated with Steiner systems possesses the t-PIR property. Since optimal codes for $t \leq 4$ are known (see [5], [6]), we mainly focus on the special case of $t = 5$ (or, equivalently, $t = 6$) for which we show that provably optimal PIR codes can be constructed from lengthening and extending an existing PIR code for code dimensions $k \leq 6$, while for all $7 \leq k \leq 32$ we find codes of smaller blocklength than the best known codes from the literature. Moreover, we also show that for certain values of k, SPRM codes are not optimal for $t = 8$.

II. Definitions and Preliminaries

Throughout this paper, we will focus on binary codes only. Component-wise addition of vectors from a vector space will be written as normal addition, and as is customary in coding theory, we denote row vectors by boldface italic Roman letters, e.g., $\boldsymbol{x}$. However, sometimes we will slightly abuse this notational convention by using $\boldsymbol{c}$ to refer to a column vector. Moreover, whether an all-zero vector $\mathbf{0}$ (or an all-one vector $\mathbf{1}$) is a row vector or a column vector will become clear from the context. The Hamming weight of a binary vector $\boldsymbol{x}$ is denoted by $w_{\mathrm{H}}(\boldsymbol{x})$ throughout the paper.

A. t-Server PIR Codes

Definition 1: Consider an $[n, k]$ linear code $\mathscr{C}$ and its corresponding generator matrix $\mathsf{G} \triangleq \left[\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\right]$. This $[n, k]$

This work was partially funded by the Research Council of Norway (grant 240985/F20).

code is said to be an $[n, k; t]$ PIR code if for every $i \in \mathbb{N}_k \triangleq \{1, 2, \ldots, k\}$, there exist t mutually disjoint sets $\mathcal{R}_h^{(i)}$, $h \in \mathbb{N}_t$, such that

$$\boldsymbol{e}_i \triangleq (\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0)^{\mathsf{T}} = \sum_{j \in \mathcal{R}_h^{(i)}} \boldsymbol{c}_j, \quad \forall\, h \in \mathbb{N}_t,$$

where superscript "$^{\mathsf{T}}$" denotes vector transposition. We also say that such a code $\mathscr{C}$ (or G) has the t-PIR property. Moreover, given a message symbol u_i, $i \in \mathbb{N}_k$, those mutually disjoint sets $\mathcal{R}_h^{(i)}$, $h \in \mathbb{N}_t$, are called the recovering sets for u_i.

For given values of k and t, the minimum value of n for which an $[n, k; t]$ PIR code exists is of great interest. This motivates us to look at a related parameter in conventional coding theory: the length of the shortest binary linear code with dimension k and minimum Hamming distance d. The smallest blocklength of a linear code for fixed values of (k, d) has been discussed extensively in the existing literature. Note that our notation of an $[n, k; t]$ PIR code should not be confused with the usual three parameters notation of an $[n, k, d]$ linear code, where the third parameter d denotes the minimum Hamming distance of the $[n, k]$ code. We make the following definitions.

Definition 2:

$$N_{\mathrm{P}}(k, t) \triangleq \min\{n\colon \text{an } [n, k; t] \text{ binary PIR code exists}\}.$$
$$N(k, d) \triangleq \min\{n\colon \text{an } [n, k, d] \text{ binary linear code exists}\}.$$

B. Bounds for t-Server PIR Codes

It is well-known that the minimum Hamming distance d of a t-server PIR code must be at least t [7].

Proposition 1: If an $[n, k; t]$ PIR code exists, then its minimum Hamming distance d must satisfy $d \geq t$.

Corollary 1: For given values of k and t, $N_{\mathrm{P}}(k, t)$ is lower-bounded by the smallest blocklength n such that an $[n, k, t]$ code exists, i.e., $N_{\mathrm{P}}(k, t) \geq N(k, t)$.

Proof: See the extended version [8]. ■

In [6], a lower bound on the minimum blocklength $N_{\mathrm{P}}(k, t)$ for any *systematic* $[n, k; t]$ PIR code was presented. As shown in [9], the bound from [6] also holds for any binary $[n, k; t]$ PIR code. The lower bound from [6], denoted by $L_{\mathrm{P}}(k, t)$, is

$$L_{\mathrm{P}}(k, t) \triangleq k + \left\lceil \sqrt{2k + \frac{1}{4}} + \frac{1}{2} \right\rceil + t - 3, \quad t \geq 3.$$

It can easily be verified that in general $N(k, t) \geq L_{\mathrm{P}}(k, t)$ for small values of $t > 4$. In fact, we will show in Section V that $N(k, t)$ is a tighter lower bound on $N_{\mathrm{P}}(k, t)$ than $L_{\mathrm{P}}(k, t)$ for $t = 6$.

Some useful upper and lower bounds on $N_{\mathrm{P}}(k, t)$ were provided by Fazeli *et al.* in [5]. Together with the constructions introduced therein, the authors provided an upper bound table on $N_{\mathrm{P}}(k, t)$ for all values of $k \leq 32$ and $t \leq 16$. We briefly summarize their results below.

Lemma 1 (Lemmas 13 and 14 in [5]):

(a) $N_{\mathrm{P}}(k, t + t') \leq N_{\mathrm{P}}(k, t) + N_{\mathrm{P}}(k, t')$,
(b) $N_{\mathrm{P}}(k + k', t) \leq N_{\mathrm{P}}(k, t) + N_{\mathrm{P}}(k', t)$,
(c) $N_{\mathrm{P}}(k, t) \leq N_{\mathrm{P}}(k + 1, t) - 1$,
(d) $N_{\mathrm{P}}(k, t) \leq N_{\mathrm{P}}(k, t + 1) - 1$, and
(e) if t is odd, then $N_{\mathrm{P}}(k, t + 1) = N_{\mathrm{P}}(k, t) + 1$.

III. Code Constructions

In this section, we first present a code construction by lengthening and extending a given PIR code, and then present an extension of a code construction inspired by Steiner systems proposed by Fazeli *et al.* in [5]. An earlier work constructing PIR codes (and even stronger batch codes) for $t = k$ based on Steiner systems (and more general block designs) was presented in [10].

A. Lengthening and Extending PIR Codes

In the following theorem, we will investigate an important property of a PIR code with an arbitrary positive integer t.

Theorem 1: For any given $t \in \mathbb{N} \triangleq \{1, 2, \ldots\}$, we have

$$N_{\mathrm{P}}(k + 1, t) \leq N_{\mathrm{P}}(k, t) + \left\lceil \frac{t}{2} \right\rceil.$$

Proof: See the extended version [8]. ■

Theorem 1 is an improved version of part (b) of Lemma 1 for $k' = 1$, while for $k' > 1$, it is an improved version only if $k' \lceil \frac{t}{2} \rceil < N_{\mathrm{P}}(k', t)$. This theorem suggests that for a given even value of t, a new t-server PIR code can always be generated by adding one information symbol and appending at most $t/2$ code symbols to the original t-server PIR code.

Next, we will discuss a special family of systematic codes that will help in the numerical search for good PIR codes with small blocklength, especially when k is large.

B. Construction of PIR Codes Based on Steiner Systems

In [5], a systematic code construction based on Steiner systems was proposed, in which the authors introduce a representation method of systematic codes, and give a sufficient (but not necessary) condition for constructing PIR codes.

Definition 3: Let $\mathscr{P}_k = \{\mathcal{P}_j\}_{j=1}^r$ be a collection of subsets of $\mathbb{N}_k$. A systematic $[n = k + r, k]$ code $\mathscr{C}$ can be represented by defining the codewords of $\mathscr{C}$ as $\boldsymbol{x} \triangleq (u_1, \ldots, u_k, x_{k+1}, \ldots, x_{k+r})$, where $u_1, \ldots, u_k$ are the information bits of the code and each redundancy bit x_{k+j} is defined as $x_{k+j} \triangleq \sum_{i \in \mathcal{P}_j} u_i$, $j \in \mathbb{N}_r$.

We denote the constructed code by $\mathscr{C}(\mathscr{P}_k)$. Furthermore, for the sake of notational convenience, we define $\mathcal{J}^{(i)} \triangleq \{j \in \mathbb{N}_r : i \in \mathcal{P}_j\}$ to be the set of indices $j \in \mathbb{N}_r$ such that $i \in \mathcal{P}_j$.

The systematic generator matrix G of this code can be written as $\mathsf{G} = [\mathsf{I}_k | \mathsf{P}_{k \times r}]$, where I_k is the $k \times k$ identity matrix and the $k \times r$ redundancy matrix $\mathsf{P}_{k \times r} = \{p_{ij}\}_{1 \leq i \leq k,\, 1 \leq j \leq r}$ is defined by

$$p_{ij} \triangleq \begin{cases} 1, & \text{if } i \in \mathcal{P}_j, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 2 (Lemma 7 in [5]): Suppose that a collection $\mathscr{P}_k = \{\mathcal{P}_j\}_{j=1}^r$ satisfies the following properties.

1) For all $i \in \mathbb{N}_k$, $\left|\mathcal{J}^{(i)}\right| \geq t - 1$, and
2) for all $j \neq j' \in \mathbb{N}_r$, $\left|\mathcal{P}_j \cap \mathcal{P}_{j'}\right| \leq 1$.

Then, the corresponding systematic code $\mathscr{C}(\mathscr{P}_k)$ is a t-server PIR code.

The above lemma only leads to an absorbing upper bound on the redundancy $N_{\mathrm{P}}(k,t) - k$ for fixed t and sufficiently large k, which shows that it is equal to $O(\sqrt{k})$. However, for smaller values of the parameter k, whether or not this upper bound is tight is still unknown. Moreover, in [5] a similar PIR code construction based on constant-weight codes was provided, where all rows of $\mathsf{P}_{k\times r}$ have constant weight and a given minimum Hamming distance.

It is known that the minimum Hamming distance d of a PIR code must be larger than or equal to the desired parameter t (see Proposition 1), and so are the row Hamming weights of any generator matrix G for the code. Hence, it is reasonable to change the sufficient condition of $\left|\mathcal{J}^{(i)}\right| \geq t-1$ in Lemma 2 to $\left|\mathcal{J}^{(i)}\right| = t-1$, $\forall\, i \in \mathbb{N}_k$.

Motivated by Steiner systems, we define a more elaborate systematic code family as follows.

Definition 4: For any integer $t \in \mathbb{N}$ and a given collection $\mathscr{P}_k = \{\mathcal{P}_j\}_{j=1}^{r}$ of subsets of $\mathbb{N}_k$, we say that a systematic code $\mathscr{C}(\mathscr{P}_k)$ (or its corresponding generator matrix) has property S_t if all of the following conditions are satisfied.

1) $\mathcal{P}_r = \mathbb{N}_k$,
2) $\left|\mathcal{J}^{(i)}\right| = t-1$ for all $i \in \mathbb{N}_k$,
3) $\left|\mathcal{P}_j \cap \mathcal{P}_{j'}\right| \leq 1$ for all $j \neq j' \in \mathbb{N}_{r-1}$, and
4) for any given $m \in \mathbb{N}_k$, there exists a subset $\mathcal{I}(m) \subseteq \mathbb{N}_k$ with $\mathcal{I}(m) \cap \bigl(\bigcup_{j\in\mathcal{J}^{(m)}\setminus\{r\}} \mathcal{P}_j\bigr) = \emptyset$ and a subset $\mathcal{V}^{(m)} \subseteq \mathbb{N}_{r-1}$ with $\mathcal{V}^{(m)} \cap \mathcal{J}^{(m)} = \emptyset$ such that
$$u_m + \sum_{i\in\mathcal{I}(m)} u_i + \sum_{j\in\mathcal{V}^{(m)}} \sum_{i\in\mathcal{P}_j} u_i = \sum_{i=1}^{k} u_i.$$

Similarly to Lemma 2, a systematic code with property S_t turns out to be an $[n,k;t]$ PIR code.

Lemma 3: If a systematic code $\mathscr{C}(\mathscr{P}_k)$ has property S_t, then it is an $[n = k+r, k; t]$ PIR code.

Proof: See the full version [8]. ■

The following example illustrates the code design of Lemma 3.

Example 1: For an $[n,k] = [17,8]$ systematic code, we describe it in terms of $\mathscr{P}_8$ as follows:
$$\begin{aligned}\mathscr{P}_8 \triangleq \bigl\{&\mathcal{P}_1 \triangleq \{1,2,3\}, \mathcal{P}_2 \triangleq \{1,4,6\}, \mathcal{P}_3 \triangleq \{1,5,7\},\\ &\mathcal{P}_4 \triangleq \{2,4,8\}, \mathcal{P}_5 \triangleq \{2,5,6\}, \mathcal{P}_6 \triangleq \{3,4,7\},\\ &\mathcal{P}_7 \triangleq \{3,5,8\}, \mathcal{P}_8 \triangleq \{6,7,8\}, \mathcal{P}_9 \triangleq \mathbb{N}_8\bigr\}.\end{aligned}$$

One can see that $r = 9$ and that the systematic code $\mathscr{C}(\mathscr{P}_8)$ has property S_5. Here, condition 4) can be verified by the following observations (e.g., take $m = 1, 8$):
$$\begin{aligned}&\mathcal{J}^{(1)} = \{1,2,3,9\},\ \mathcal{J}^{(8)} = \{4,7,8,9\},\\ &\mathcal{I}(1) = \{8\},\ \mathcal{I}(8) = \{1\},\ \mathcal{V}^{(1)} = \mathcal{V}^{(8)} = \{5,6\},\\ &\quad \mathbb{N}_8 = \{1\} \cup \mathcal{P}_5 \cup \mathcal{P}_6 \cup \{8\}.\end{aligned}$$

Then, we can conclude that this code is a 5-server $[17,8]$ PIR code. For example, the recovering sets for the first information bit are determined by $\mathcal{R}_1^{(1)} = \{1\}$,
$$\begin{aligned}\mathcal{R}_2^{(1)} &= \{m \in \mathcal{P}_1 \colon m \neq 1\} \cup \{k+1\} = \{2,3,9\},\\ \mathcal{R}_3^{(1)} &= \{m \in \mathcal{P}_2 \colon m \neq 1\} \cup \{k+2\} = \{4,6,10\},\\ \mathcal{R}_4^{(1)} &= \{m \in \mathcal{P}_3 \colon m \neq 1\} \cup \{k+3\} = \{5,7,11\},\\ \mathcal{R}_5^{(1)} &= \{8, k+5, k+6, k+r\} = \{8,13,14,17\}.\end{aligned}$$

In fact, the idea behind Lemma 3 is to try to combine the properties of Steiner systems and part (e) of Lemma 1, in such a way that we can construct an $[n+1,k;t+1]$ PIR code from an $[n,k;t]$ PIR code when t is even.

We also remark that a systematic $[n,k;t]$ PIR code with property S_t usually has different cardinalities of its recovering sets (the so-called *non-uniform information-symbol locality* property). For instance, for the code of Example 1, each information symbol has 1 recovering set of cardinality 1, 3 recovering sets of cardinality 3, and 1 recovering set of cardinality 4. This is also in alignment with [6], where the presented PIR codes in general have recovering sets of different cardinalities. In Section V, we will show that codes having property S_5 are good 5-server PIR codes with small blocklength.

IV. Searching for Optimal PIR Codes

In this section, we present an algorithm to search for good (i.e., small blocklength) PIR codes. Since optimal codes for $t \leq 4$ are already known for all code dimensions k, we concentrate on $t = 5$. Because Theorem 1 implies that we can construct a t-server PIR code by lengthening and extension, hence, combined with the idea of lexicographic code construction [11], Algorithm 1 is proposed to find a sequence of good systematic PIR codes for $t = 5$.[1]

Initially, we choose the best known $[n,k;5]$ code with a systematic generator matrix in which all rows have weight 5. Note that for small values of n and k, such a code is not too difficult to find. As an example, the generator matrix G of a systematic $[8,2;5]$ code in which all rows have weight 5 is
$$\mathsf{G} = \left[\begin{array}{cc|ccccc|c} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1\\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1\end{array}\right]. \tag{1}$$

The outer while loop of Algorithm 1 increases a counter (denoted by i) from 1 to $\binom{r+w}{4}$ (the counter runs over all possible length-$(r+w)$ binary vectors of weight 4). The function `LengtheningExtending`$(\mathsf{G}_{\text{best}}, \boldsymbol{z})$ in Line 6 of Algorithm 1 is defined by
$$\tilde{\mathsf{G}} \triangleq \left[\begin{array}{c|c|c} \mathsf{I}_{k_{\text{best}}} & \mathbf{0} & \mathsf{P}_{k_{\text{best}}\times(r+w)}\\ \hline \mathbf{0} & 1 & \boldsymbol{z}\end{array}\right], \qquad \underbrace{\qquad\quad}_{k_{\text{best}}+1}\ \underbrace{\qquad\quad}_{r+w}$$

where $\mathsf{G}_{\text{best}} = \bigl[\mathsf{I}_{k_{\text{best}}} | \mathsf{P}_{k_{\text{best}}\times(r+w)}\bigr]$ and $w_{\mathrm{H}}(\boldsymbol{z}) = 4$.[2] Note that if $w = 2$, it follows from the proof of Theorem 1 in [8] that $k_{\text{best}} \geq k+1$; explaining why we choose $1 \leq w \leq 2$ from the beginning. Furthermore, notice that for $w = 1$, sometimes

[1] In general, this algorithm can be applied for any t. The main reason why we focus on small values of t is that when t is increasing, the complexity to determine whether a code has the t-PIR property is also increasing.

[2] Note that the definition of $\tilde{\mathsf{G}}$ guarantees that G_{best} is always in systematic form in each iteration.

Algorithm 1: Searching for optimal 5-server PIR codes

```
Input  : A systematic constant row-weight-5 generator
         matrix $\mathsf{G} = [\mathsf{I}_k|\mathsf{P}_{k\times r}]$ for an $[n, k; 5]$ code, and a
         given $w \in \mathbb{N}_2$.
Output: A systematic constant row-weight-5 generator
         matrix $\mathsf{G}_{\text{best}}$ for an $[n_{\text{best}}, k_{\text{best}}; 5]$ code, where
         $k_{\text{best}} \geq k$ is the largest possible code dimension
         found and $n_{\text{best}} = k_{\text{best}} + r + w$.
1  $\mathsf{G}_{\text{best}} \leftarrow [\mathsf{I}_k|\mathsf{P}_{k\times r}|\mathsf{O}_{k\times w}]$, $k_{\text{best}} \leftarrow k$
2  /* $\mathsf{O}_{k\times w}$ is a $k \times w$ all-zero matrix */
3  $i \leftarrow 1$
4  $\boldsymbol{z} \leftarrow$ the row vector $(1, 1, 1, 1, 0, \ldots, 0)$ of length $r + w$
5  while $i \leq \binom{r+w}{4}$ do
6      $\tilde{\mathsf{G}} \leftarrow$ LengtheningExtending($\mathsf{G}_{\text{best}}, \boldsymbol{z}$)
7      $\tilde{d} \leftarrow$ minimum Hamming distance of $\tilde{\mathsf{G}}$
8      /* we simply say a code $\tilde{\mathsf{G}}$ is the set of all rows of $\tilde{\mathsf{G}}$ */
9      if $\tilde{d} \geq 6$ then
10         if $\tilde{\mathsf{G}}$ has the 5-PIR property then
11             $\mathsf{G}_{\text{best}} \leftarrow \tilde{\mathsf{G}}$, $k_{\text{best}} \leftarrow k_{\text{best}} + 1$
12         else
13             return $(\mathsf{G}_{\text{best}}, k_{\text{best}})$
14         end
15     end
16     $i \leftarrow i + 1, \boldsymbol{z} \leftarrow$ Lexical($\boldsymbol{z}$)
17 end
18 if $k_{\text{best}} = k$ then
19     $\mathsf{G}_{\text{best}} \leftarrow \mathsf{G}$
20 end
21 return $(\mathsf{G}_{\text{best}}, k_{\text{best}})$
```

the algorithm only results in the original input code. We also verify whether $\tilde{d} \geq 6$ or not in Line 9 of Algorithm 1. This is to ensure that the resulting code generated by $\tilde{\mathsf{G}}$ can potentially satisfy Proposition 1.[3] Finally, given a vector $\boldsymbol{z}$, Lexical($\boldsymbol{z}$) generates the next lexicographical constant-weight $\boldsymbol{z}$ of length $r + w$, e.g., Lexical($\boldsymbol{z}$) $= (1, 1, 1, 0, 1, 0, \ldots, 0)$ for $\boldsymbol{z} = (1, 1, 1, 1, 0, \ldots, 0)$.

We also remark that the resulting k_{best} from Algorithm 1 strongly depends on the selected $\mathsf{G} = [\mathsf{I}_k|\mathsf{P}_{k\times r}]$ and the given w in the input. It is difficult to predict whether the corresponding blocklength n_{best} is good or not. For example, given the systematic $[n = k + r, k; t] = [8, 2; 5]$ code defined in (1) and $w = 1$, the output from Algorithm 1 is an $[n_{\text{best}}, k_{\text{best}}; 5] = [11, 4; 5]$ code without property S_5, while for $w = 2$, Algorithm 1 results in an $[n_{\text{best}}, k_{\text{best}}; 5] = [13, 5; 5]$ code with property S_5 (see Section V that follows). Now, for code dimension $k = 4$, the $[11, 4; 5]$ code is better than the $[12, 4; 5]$ code obtained by shortening the optimal $[13, 5; 5]$ code. Hence, for a fixed code dimension k, to find a good 5-server PIR code with small blocklength, we have to compare all the resulting $[n, k; 5]$ codes found by Algorithm 1.

In general, the complexity of exhaustively examining the t-PIR property for a given code becomes infeasible for large n and k, even for $t = 5$. However, according to our numerical results, for small code dimensions k, an optimal 5-server PIR code often has property S_5. Therefore, we investigate a sequence of good PIR codes with respect to property S_5. In fact, a sequence of good codes with small blocklength can always be generated by lengthening by one information symbol and extending at most 2 coordinates from a smaller-sized code with property S_5, as shown in the theorem below.

Theorem 2: For any given values of n and k, if a systematic $[n, k]$ code has property S_5, then there must exist a systematic $[n + 3, k + 1]$ code that also has property S_5.

Proof: See the details in the extended version [8]. ■

Based on Theorem 2, we can slightly modify Algorithm 1 to investigate 5-server PIR codes with property S_5. First, we replace the input generator matrix by a generator matrix $\mathsf{G} = [\mathsf{I}_k|\mathsf{P}_{k\times(r-1)}|\mathbf{1}]$ with property S_5, and modify the starting G_{best} to $[\mathsf{I}_k|\mathsf{P}_{k\times(r-1)}|\mathsf{O}_{k\times w}|\mathbf{1}]$ in Line 1 of Algorithm 1. The function LengtheningExtending($\mathsf{G}_{\text{best}}, \boldsymbol{z}$) for $\mathsf{G}_{\text{best}} = \bigl[\mathsf{I}_{k_{\text{best}}}|\mathsf{P}_{k_{\text{best}}\times(r+w-1)}|\mathbf{1}\bigr]$ in Line 6 of Algorithm 1 is accordingly re-defined as

$$\tilde{\mathsf{G}} \triangleq \left[\underbrace{\begin{array}{c|c} \mathsf{I}_{k_{\text{best}}} & \mathbf{0} \\ \hline \mathbf{0} & 1 \end{array}}_{k_{\text{best}}+1} \middle| \underbrace{\begin{array}{c} \mathsf{P}_{k_{\text{best}}\times(r+w-1)} \\ \hline \boldsymbol{z} \end{array}}_{r+w-1} \middle| \begin{array}{c} \mathbf{1} \\ \hline 1 \end{array} \right],$$

where $w_{\text{H}}(\boldsymbol{z}) = 5 - 2 = 3$. Notice that the outer while loop counter now should increase from 1 to $\binom{r+w-1}{3}$, and the initial $\boldsymbol{z}$ in Line 4 should be replaced by the length-$(r + w - 1)$ vector $\boldsymbol{z} = (1, 1, 1, 0, \ldots, 0)$. In fact, there is no need to modify Line 9 of Algorithm 1, since the resulting $\tilde{\mathsf{G}}$ will again satisfy conditions 1)–3) of Definition 4.[4] As a result, after the modifications to Algorithm 1 outlined above, and if Line 10 of Algorithm 1 is replaced by the verification of property S_5 for $\tilde{\mathsf{G}}$, we are able to find good 5-server PIR codes with property S_5 for large code dimensions $k \geq 16$ (see Section V below). From Theorem 2 it follows that if $w = 2$, $k_{\text{best}} \geq k + 1$.

V. Numerical Results

In this section, upper bounds on $N_{\text{P}}(k, t)$ for $1 \leq k \leq 32$ and $t = 4, 6, 8$ are summarized in Table I. In particular, for $t = 6$, we also present the numerical results obtained using the search algorithm from Section IV. Entries for which strictly better codes are found than in the current literature are marked in bold. In comparison with the obtained improved upper bound, a lower bound on $N_{\text{P}}(k, 6)$ is also given. For $t = 4$, the SPRM codes provided in [6] are optimal. More specifically, the blocklength is equal to the lower bound $L_{\text{P}}(k, 4)$.

In order to show how good our constructed 6-server PIR codes are, we also list the best (smallest) known blocklength

[3] Since the construction guarantees that all rows have equal Hamming weights, the Hamming distance between any pair of rows is even, i.e., the necessary condition $\tilde{d} \geq 5$ is equivalent to $\tilde{d} \geq 6$.

[4] Note that the construction of $\tilde{\mathsf{G}}$ will make all the row-weights of $\tilde{\mathsf{G}}$ equal to 5 and the last column equal to the all-one vector (i.e., conditions 1) and 2) of Definition 4 are satisfied). In order to satisfy condition 3) of Definition 4, the minimum Hamming distance of $\tilde{\mathsf{G}}$ must be larger than or equal to $2\cdot(5-2) = 6$, since any two row vectors in $\tilde{\mathsf{G}}$ must have a common 1 in at most two coordinates.

TABLE I

BEST KNOWN BOUNDS ON $N_{\mathrm{P}}(k,t)$ FOR SMALL VALUES OF k AND EVEN $t = 4, 6, 8$. IN THE CASE OF $t = 6$, n_{B} DENOTES THE BEST FOUND BLOCKLENGTH BASED ON OUR PROPOSED SEARCH ALGORITHM, AND n_{U} IS DEFINED IN (2). STARRED VALUES (OR COLUMNS) CAN BE PROVED TO BE OPTIMAL, WHILE BOLD ENTRIES ARE NEW RESULTS.

$k\backslash t$	4^{*} [6]	6			8 [6]
		$N(k,t)$ [12]	n_{B}	n_{U}	
1	4	–	6^{*}	–	8^{*}
2	6	–	9^{*}	–	12^{*}
3	7	–	11^{*}	–	14^{*}
4	9	–	$12^{*\diamond}$	–	15^{*}
5	10	14	14^{*}	$13^{!}$	19
6	11	15	$15^{*\diamond}$	$14^{!}$	21
7	13	16	**17**	$15^{!}$	22
8	14	17	**18**	20	24
9	15	18	**20**	23	25
10	16	20	**21**	24	26
11	18	21	**22**	25	30
12	19	22	**23**	26	32
13	20	23	$\mathbf{25}^{\diamond}$	27	33
14	21	24	$\mathbf{27}^{\diamond}$	29	35
15	22	26	$\mathbf{28}^{\diamond}$	34	36
16	24	27	**31**	35	37
17	25	28	**32**	37	39
18	26	29	**33**	38	40
19	27	30	**35**	39	41
20	28	31	**36**	40	42
21	29	32	**37**	42	46
22	31	33	**39**	46	48
23	32	34	**40**	47	49
24	33	36	**41**	49	51
25	34	37	**42**	50	52
26	35	38	**43**	51	53
27	36	39	**44**	53	55
28	37	40	**46**	54	56
29	39	41	**47**	55	57
30	40	42	**48**	56	58
31	41	43	**50**	58	60
32	42	44	**52**	59	61

for $t = 8$ (the smallest blocklength of the SPRM codes from [6]). They will result in an improved upper bound for $t = 6$, since by part (d) of Lemma 1, $N_{\mathrm{P}}(k,6) \leq N_{\mathrm{P}}(k,8)-2$. Hence,

$$n_{\mathrm{U}} \triangleq \min\{n_1, n_2 - 2\} \tag{2}$$

is the best known upper bound for $t = 6$, where n_1 denotes the best known blocklength provided in [5], and n_2 is the smallest blocklength of SPRM codes for $t = 8$ provided in [6].

Note again that, according to part (e) of Lemma 1 and in order to compare our findings with [5, Table III] and [6, Table II], only even values of t are interesting. Here, for $t = 6$ the blocklengths n_{B} of Table I are obtained by adding one to the blocklengths of our best found 5-server PIR codes. We make the following remarks to Table I.

1) The superscript "$*$" indicates that the corresponding blocklength can be shown to be optimal. We use the lower bound $N(k,t)$, whose value can be obtained from [12], since $L_{\mathrm{P}}(k,6) = L_{\mathrm{P}}(k,4) + 2 \leq N(k,6)$ and no tighter lower bound for $t = 6$ is known.
2) The superscript "$\diamond$" indicates that the best found systematic $[n,k;5]$ code has a constant-weight generator matrix of row-weight 5 and without property S_5.
3) The superscript "!" indicates that the corresponding blocklength is impossible, since it is smaller than $N(k,t)$ (a contradiction to Corollary 1). We believe that the value of $n_{\mathrm{U}} = 15$ for $(k,t) = (7,6)$ in [5, Table III] was obtained from [5, Thm. 9] and should have corresponded to $(k,t) = (6,6)$ due to a misprint in [13, p. 289] in the redundancy of *type*-1 *doubly transitive invariant codes*. We believe this explains the contradictions.
4) The superscript "[$\cdot$]" indicates the reference number.

We also remark that for $t = 8$, using our algorithm we are able to find better PIR codes for certain values of k: we have obtained $n_{\mathrm{B}} = 18, 20, 29, 31$ for $k = 5, 6, 11, 12$, respectively. This indicates that the SPRM codes are not optimal for $t = 8$.

VI. CONCLUSION

In this paper, we presented a construction of a t-server PIR code by lengthening and extension of an existing PIR code. We also presented an extension of a code construction inspired by Steiner systems proposed by Fazeli *et al.*, which was used in the proposed algorithm to search for good (i.e., small blocklength) 5-server PIR codes. For code dimensions $k \leq 6$, provably optimal PIR codes were found, while for all $7 \leq k \leq 32$, codes of smaller blocklength than the best known codes from the literature were found and presented. Moreover, better 8-server PIR codes were also found for $k = 5, 6, 11, 12$.

REFERENCES

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proc. 36th IEEE Symp. Found. Comp. Sci.*, Milwaukee, WI, USA, Oct. 1995, pp. 41–50.

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 856–860.

[3] R. Tajeddine and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 1411–1415.

[4] S. Kumar, E. Rosnes, and A. Graell i Amat, "Private information retrieval in distributed storage systems using an arbitrary linear code," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 1421–1425.

[5] A. Fazeli, A. Vardy, and E. Yaakobi, "PIR with low storage overhead: Coding instead of replication," May 2015, arXiv:1505.06241v1 [cs.IT]. [Online]. Available: http://arxiv.org/abs/1505.06241

[6] M. Vajha, V. Ramkumar, and P. V. Kumar, "Binary, shortened projective Reed Muller codes for coded private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 2648–2652.

[7] V. Skachek, "Batch and PIR codes and their connections to locally repairable codes," Jun. 2017, arXiv:1611.09914v3 [cs.IT]. [Online]. Available: https://arxiv.org/abs/1611.09914

[8] H.-Y. Lin and E. Rosnes, "Lengthening and extending binary private information retrieval codes," Jan. 2018, arXiv:1707.03495v3 [cs.IT]. [Online]. Available: https://arxiv.org/abs/1707.03495

[9] S. Rao and A. Vardy, "Lower bound on the redundancy of PIR codes," Feb. 2017, arXiv:1605.01869v2 [cs.IT]. [Online]. Available: https://arxiv.org/abs/1605.01869

[10] Z. Wang, O. Shaked, Y. Cassuto, and J. Bruck, "Codes for network switches," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1057–1061.

[11] V. I. Levenstein, "A class of systematic codes," *Sov. Math.-Dokl.*, vol. 1, no. 1, pp. 368–371, 1960.

[12] M. Grassl, "Bounds on the minimum distance of linear codes and quantum codes," accessed on 2017-03-31. [Online]. Available: http://www.codetables.de

[13] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2004.