Upper Bound on the Communication Complexity of Private Information Retrieval*

Andris Ambainis

Institute of Mathematics and Computer Science, University of Latvia, Raina bulv. 29, Riga, Latvia, e-mail: ambainis@csu.lv. From August 1997 at Computer Science Division, University of California at Berkeley.

Abstract. We construct a scheme for private information retrieval with $k$ databases and communication complexity $O(n^{1/(2k-1)})$.

1 Introduction

Much attention has been given to the problem of protecting a database from a user that tries to retrieve an information that he is not allowed to access[2, 8, 12].

In some scenarios, an opposite problem can appear: a user wishes to retrieve some information from a database without revealing to the database what information he needs. For example[7], an investor wishes to receive an information about a certain stock but he does not wish others (even the database) to know in which particular stock he is interested.

However, there is only one way to reach a complete privacy: the user should ask for the copy of the entire database. Otherwise, the database will get some information what the user wishes to know. This is not a good solution because it requires much time and much communication from the database to the user.

If there are several identical copies of the database, another scenario is possible[7]:

The user asks a query to each database and combines the results of the queries, obtaining the desired information. Each query alone gives no information what the user is interested in.

Chor, Goldreich, Kushilevitz, Sudan[7] introduced this model and constructed several schemes for a private retrieval of one bit from a database:

1. A scheme for 2 databases with $O(n^{1/3})$ communication. ($n$ is the size of the database)
2. A scheme for $k$ databases with $O(n^{1/k})$ communication.
3. A scheme for $O(\log n)$ databases with $O(\log^2 n \log \log n)$ communication.

In this paper, we improve their result, constructing a protocol for $k$ databases with $O(n^{1/(2k-1)})$ communication.

* The author was supported by Latvia Science Council Grant 96.0282 and scholarship "SWH Izglītībai, Zinātnei un Kultūrai" from Latvia Education Foundation
Related work. Protocols for private information retrieval in [7] and this paper have used ideas from several related problems (instance hiding and multiparty communication complexity).

Instance hiding[1, 5, 6] is the problem of obtaining the $i$th bit from the oracle so that $i$ remains secret. There are some similarities and some substantial differences between instance hiding and private information retrieval (see [7] for more detailed discussion).

Techniques from instance hiding were relevant to protocols for private information retrieval in [7]. However, they are not used in this paper.

Multiparty communication complexity is also related to private information retrieval. Pudlak, Rödl, Sgall[11] and Ambainis[3] have considered the problem of computing $x_{(i+j)\mod n}$ where $x$ is a string of $n$ bits and $i, j$ are integers in the following model:

Player 1 knows $x, i$, Player 2 knows $x, j$. Each of them sends one message to Player 3. Player 3 computes the result, using only the messages received from Players 1 and 2.

Any protocol for the above problem can be easily transformed into protocol for private information retrieval. Thus, we can obtain nontrivial protocols for private information retrieval with $o(n)$ communication.

Another communication complexity problem was studied by Babai, Kimmel and Lokam[4]. It also can be applied to private information retrieval.

However, all these protocols are less efficient than the protocols for private information retrieval designed in [7]. Still, the ideas from [3, 4, 11] (not explicit protocols) can be useful in the study of private information retrieval. In particular, this paper is based on the idea of combining two protocols which appeared in the setting of multiparty communication complexity[3, 11].

2 Model

Formally, we view the database as a string $x$ consisting of $n$ bits. $k$ denotes the number of identical databases. We assume that the user wishes to retrieve a single bit $x_i$ from the database.

We require that, for every database, indices $i, j$ and any message from the user, the probability of the database receiving this message is equal when the user retrieves the $i$th bit and when the user retrieves the $j$th bit. This means that database does not get any information about $i$.

There are several extensions of this model. [7] considered schemes which allow to retrieve blocks of information and give a higher degree of privacy (knowing $k - 1$ of $k$ queries gives no information about the bit that the user retrieves). Ostrovsky and Shoup[9] have extended the results of [7] and designed schemes for private information storage. Using their schemes, the user can both read and write to the database without revealing which bit is accessed. They have shown that any protocol for private information retrieval can be transformed to the protocol for private information storage with a slight increase in the number of databases and communication.