Private Information Retrieval, Distributed Storage and Network Coding

8 July 2016

Royal Holloway University of London

All talks will be in the Moore Building. Coffee and lunch will be provided to all registered participants.

10:00-10:30 Coffee

10:30-11:30 P. Vijay Kumar Coding for Distributed Storage: An Overview

11:45-12:30 Alex Vardy Private Information Retrieval without Storage Overhead: Coding Instead of Replication (Part 1)

12:30-13:30 Lunch

13:30-14:15 Alex Vardy Private Information Retrieval without Storage Overhead: Coding Instead of Replication (Part 2)

14:15-15:15 Salim El Rouayheb Private Information Retrieval from Coded Data

15:15-15:45 Tea

15:45-16:45 Tuvi Etzion Combinatorial framework for network coding, distributed storage, and PIR codes

Abstracts:

P. Vijay Kumar (Indian Institute of Science, Bangalore) Coding for Distributed Storage: An Overview

In this talk, an accessible overview of recent developments on the topic of coding for distributed storage will be provided. The talk will cover both regenerating codes and codes with locality. Performance bounds and code constructions will be discussed.

Alex Vardy (UC San Diego) Private Information Retrieval without Storage Overhead: Coding Instead of Replication

Private information retrieval protocols allow a user to retrieve a data item from a database without revealing any information about the identity of the item being retrieved. Specifically, in information-theoretic *k*-server PIR, the database is replicated among *k* non-communicating servers, and each server learns nothing about the item retrieved by the user. The effectiveness of PIR protocols is usually measured in terms of their communication complexity, which is the total number of bits exchanged between the user and the servers. However, another important cost parameter is the *storage overhead*, which is the ratio between the total number of bits stored on all the servers and the number of bits in the database. Since single-server information-theoretic PIR is impossible, the storage overhead of all existing PIR protocols is at least 2 (or *k*, in the case of *k*-server PIR).

In this work, we show that information-theoretic PIR can be achieved with storage overhead arbitrarily close to the optimal value of 1, without sacrificing the communication complexity. Specifically, we prove that *all* known *k*-server PIR protocols can be efficiently emulated, while preserving both privacy and communication complexity but significantly reducing the storage overhead. To this end, we distribute the *n* bits of the database among *s+r* servers, each storing *n/s* coded bits (rather than replicas). Notably, our coding scheme remains the same, regardless of the specific *k*-server PIR protocol being emulated. For every fixed *k*, the resulting storage overhead *(s+r)/s* approaches

1 as s grows; explicitly we have $r \leq k\, s^{1/2}$ (1 + o(1)). Moreover, in the special case $k = 2$, the storage overhead is only $1 + 1/s$.

In order to achieve these results, we introduce and study a new kind of binary linear code, called here *k-server PIR codes*. We then show how such codes can be constructed from Steiner systems, from one-step majority-logic decodable codes, from constant-weight codes, and from certain locally recoverable codes. We also establish several bounds on the parameters of *k*-server PIR codes, and tabulate the results for all $s \leq 32$ and $k \leq 16$.

Salim El Rouayheb (Illinois Institute of Technology) Private Information Retrieval from Coded Data

An information theoretic Private Information Retrieval (PIR) scheme ensures that a user can retrieve records in a database or files in a distributed storage system (DSS) while revealing no information on which record or file is being retrieved. A user can achieve PIR by downloading all the data in the DSS. However, this is not a feasible solution due to its high communication cost. I will present constructions of PIR schemes with low download communication cost when the data is stored on the DSS using Maximum Distance Separable (MDS) codes.I will also discuss open questions in this area.

Tuvi Etzion (Technion): Combinatorial framework for network coding, distributed storage, and PIR codes

For the last fifteen years, the area of network coding has been rapidly expanding. Five years ago, a novel connection between network coding and distributed storage was shown. In the last two years private information retrieval (PIR) protocols have gained a considerable amount of attention from a (distributed) storage point of view. Various techniques, from many areas of mathematics, have been applied to these three research areas. We will present first the basic concepts of network coding and consider the smallest alphabet size for which a network has a solution. Finally, we outline new results we obtained lately on PIR array codes.

Contact: Simon Blackburn, Department of Mathematics, Royal Holloway University of London, Egham, Surrey TW20 0EX. s.blackburn@rhul.ac.uk

Supported by EPSRC Grant EP/N022114/1 ('Private Information Retrieval with limited server storage: combinatorics and coding theory').