Towards Practical (?) Private Information Retrieval

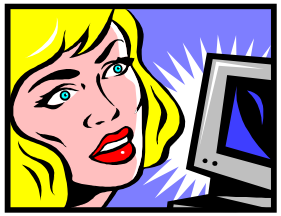
Speaker: Giovanni Di Crescenzo
Telcordia Technologies,
Piscataway, NJ, USA
E-mail: giovanni@research.telcordia.com

Venue: Second International Conference on Security and Privacy in Communication Networks (SECURECOMM 2006)

What I am really going to tell you

- What Private Information Retrieval (PIR) is
- How important is PIR anyway
  - Applications and perceptions
- What researchers have done about it
- What researchers may want to do about it
  - which "efficiency metric" ?
  - which "functionality model" ?
  - which "interaction model" ?
  - which "privacy model" ?
- Status and future

What Private Information Retrieval (PIR) is



Alice

(Data Search) Query

(Data Download) Answer

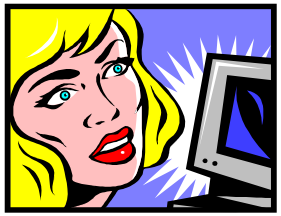
Database

- Database: string x of bits x[1],…,x[n]
- Alice's input: private input i in {1,…,n}
- Three main requirements: correctness, privacy, non-triviality
  - Correctness: Alice gets x[i]
  - Privacy: Database administrator does not obtain any partial information about index i
  - Non-triviality: communication complexity < n

How important is PIR anyway

- Applications:
  - "Direct" Applications of PIR
    - Search by keywords
    - Searching documents in file systems, records in databases
    - Private information sharing
  - Known result: given a PIR protocol, one can construct an Oblivious Transfer (OT) protocol
    - OT is also called Symmetric PIR (SPIR)
  - "Direct" Applications of OT
    - Privacy in Auctions, Stock Market operations,…
  - Other applications of OT
    - 2-party and multi-party private function evaluation protocols
- Perceptions:
  - PIR may be of large interest to some
    - clients, users, government (?),…
  - PIR may be annoying to many others
    - servers, businesses that commoditize information,…

What researchers have done about it



Alice

(Data Search) Query

(Data Download) Answer

Database

- Basic protocol idea:
  - Alice's query is a "homomorphic encryption" c of her private index i;
  - Database's answer is a manipulation c' of c, using x[1],…,x[n];
  - At the end, Alice recovers i after decrypting c'
- Most studied questions:
  - Decreasing communication complexity
  - Decreasing server's computational complexity
  - Applications to / relationships with cryptographic primitives
  - Problem variants
    - Multi-database model

What researchers may want to do about it

- Asking and answering the "right questions"
- Here are some of them:
  - What is the most appropriate "efficiency metric" ?
  - What is the most practical "functionality model" ?
  - What is the most practical "interaction model" ?
  - What is the most practical "privacy model" ?
  - Any other "requirements" for PIR ?

- My very debatable opinions follow…

Which efficiency metric ?

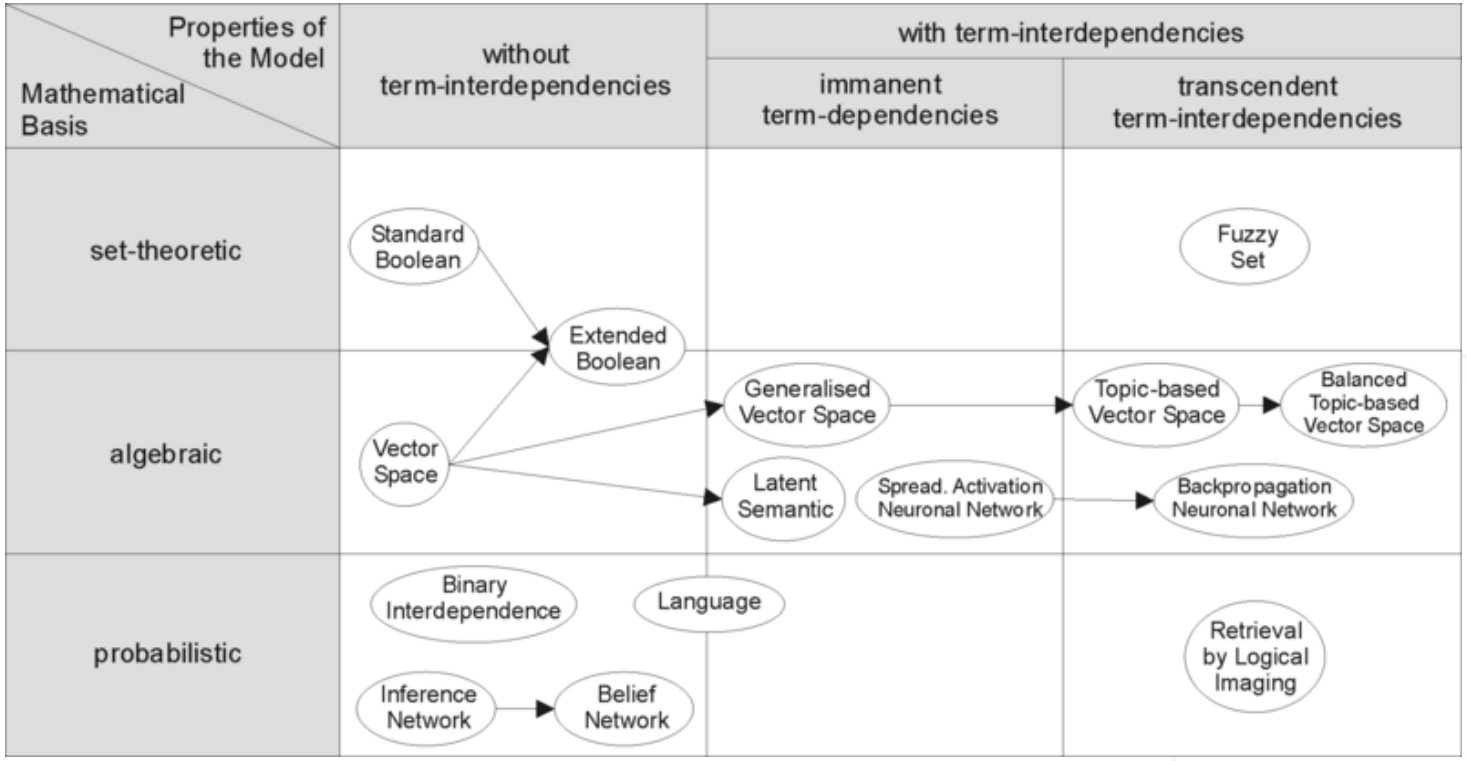
- Current metrics:
  - Communication complexity
    - Sublinear, then Polylogarythmic in the database length
  - (Server's) Computational complexity
    - Smaller than database length
    - perceived as major bottleneck towards practical PIR
- Current approaches:
  - Minimal amount of computation per database bit
  - Efficient retrieval of blocks
- Suggestion:
  - Scenario-dependent, non-asymptotic, amortized over multiple queries, combination of both metrics
- Real target:
  - Low-communication
    - In a non-asymptotic sense, as a function of all parameters
  - Computational efficiency of Hybrid-Encryption
    - Symmetric Encryption after Asymmetric Encryption of a short key

Which functionality model ? (1)

- Current PIR functionality definition:
  - retrieval of a bit from a string of n bits
    - elegant but huge over-simplification
- Other studied functionalities:
  - Efficient retrieval of a block of contiguous bits
  - Keyword search

Which functionality model ? (2)

- Actual information retrieval:



- Suggestion: Studying which other practical functionalities are efficiently achievable using PIR
  - Ultimately dictated by real-life applications

Which interaction model ?

- Current PIR interaction definitions:
  - One client retrieving data from one server
  - One client retrieving data from multiple servers that do not communicate (!)
- Suggestions for more practical interaction models:
  - Single-server PIR using help from other trusted servers (E.g., commodity-based PIR)
  - Off-line data preprocessing from a single server
  - Single-database retrievals from multiple users

Which privacy model ?

- Current PIR interaction definition:
  - Server cannot guess any partial information about client's index
- Suggestion from a theoretical point of view:
  - Universally-composable PIR
- Suggestions from a practical point of view:
  - Partial privacy is enough for most applications
  - "Repudiative" PIR
  - K-Anonymity
  - Application-dependent privacy definition

Any other requirements for PIR ?

- Current PIR requirements:
  - Correctness, privacy, non-triviality
- Suggestions:
  - How about "simplicity of implementation" ?

Status and future

- PIR software libraries are publicly available
- But (imho) PIR still seems very far from an applicable cryptographic primitive
  - Even in small-database scenarios !
- Do we understand the challenges ?
  - Mostly YES
- Can we solve them ?
  - Some seem really hard… ☹
  - On the other hand we have plenty of open problems! ☺