Recommendation Systems meet PIR

Adithya Vadapalli, Fattaneh Bayatbabolghani, and Ryan Henry

Recommendation Systems

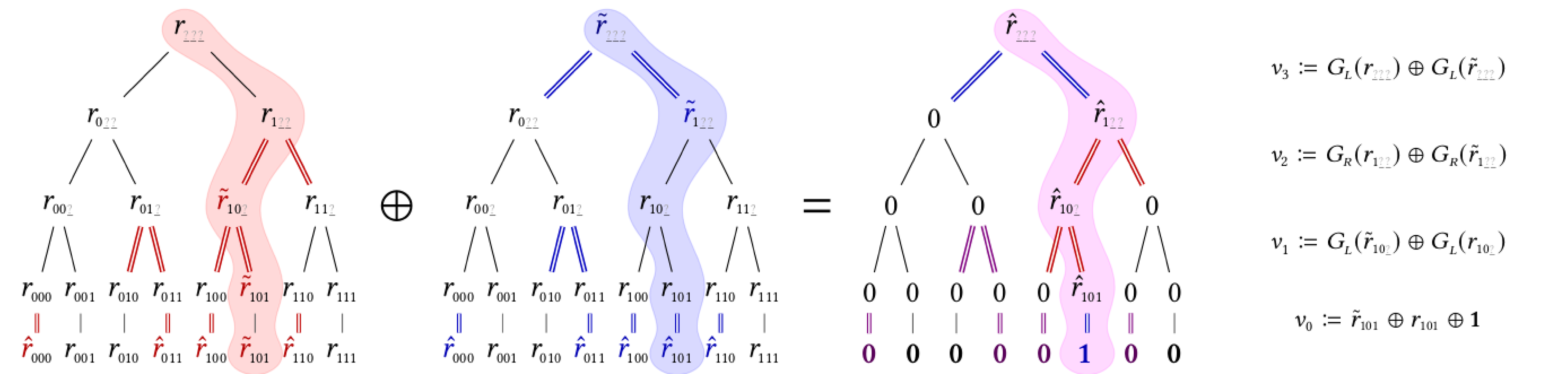
Allows business to increases their sales. User information collected, which could potentially be misused, stolen, sold.

Our Goal

Our goal is to build a recommendation system, that:

1. Provides relevant recommendations to the users.
2. Is completely oblivious to users' consumption patterns.

Distributed Point Functions

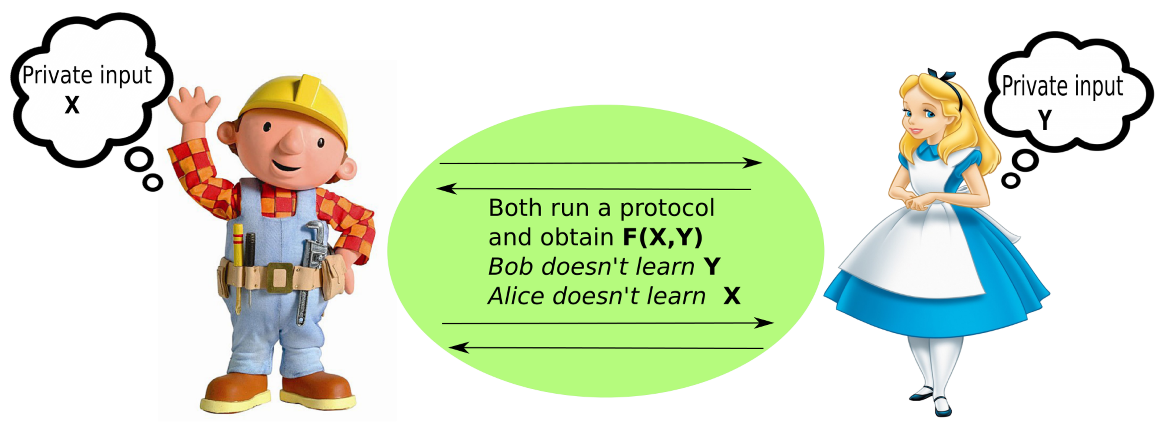


(a) First tree share (b) Second tree share (c) Reconstructed tree (d) Correction words

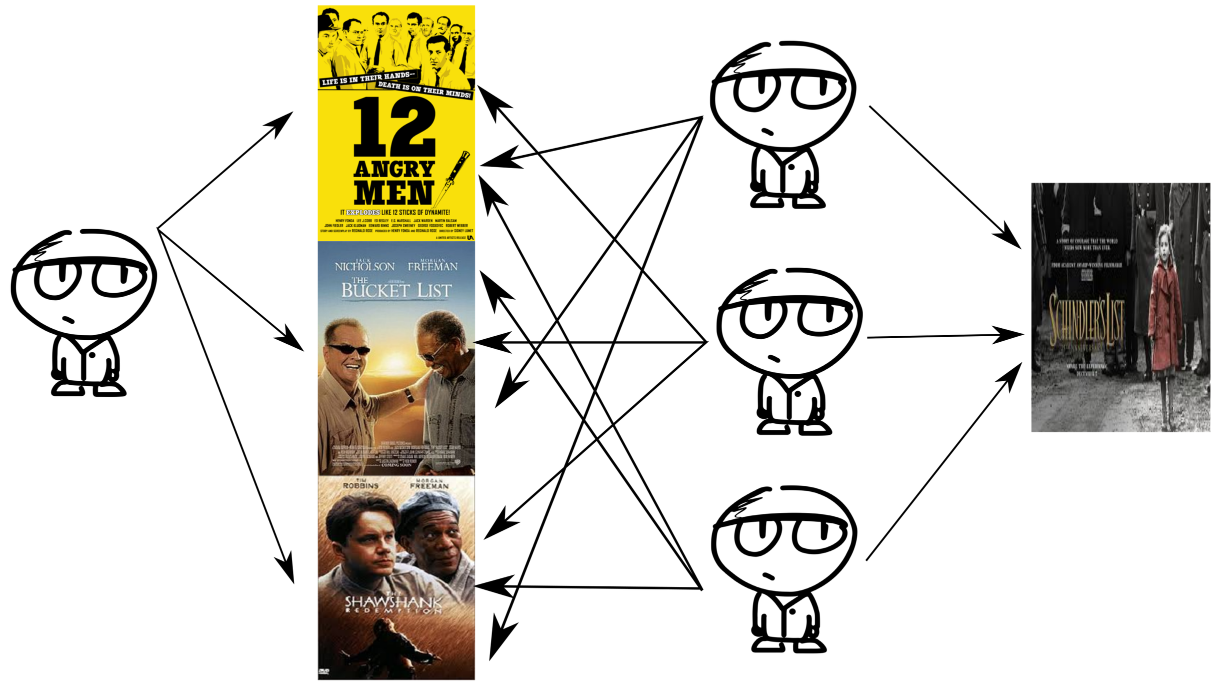
We show how to use the DPFs to realize two-party fixed-selection-wire multiplexers and demultiplexers, which serve as extremely fast and non-interactive drop-in replacements for what would otherwise be the two most expensive steps in MPC-based gradient descent.

Building Blocks

1. **Private Information Retrieval** allows us to obliviously fetch data from a database. For example, a PIR based Netflix would allow users to watch movies while Netflix is completely oblivious to the movies watched by the users.
2. $(2,1)$ **Distributed Point Functions** provide a way to distribute a point function P_i amongst 2 servers such that the servers learn nothing about i, if they don't collude. A point function P_i evaluates to 0 at every input except i.
3. **Multi-Party Computation**
   - $P_1, \cdots, P_n$, with private inputs $w_1, \cdots, w_n$ respectively.
   - Compute a function $\mathcal{F}(w_1, \cdots, w_n)$ while keeping their private inputs secret.



Collaborative Filtering

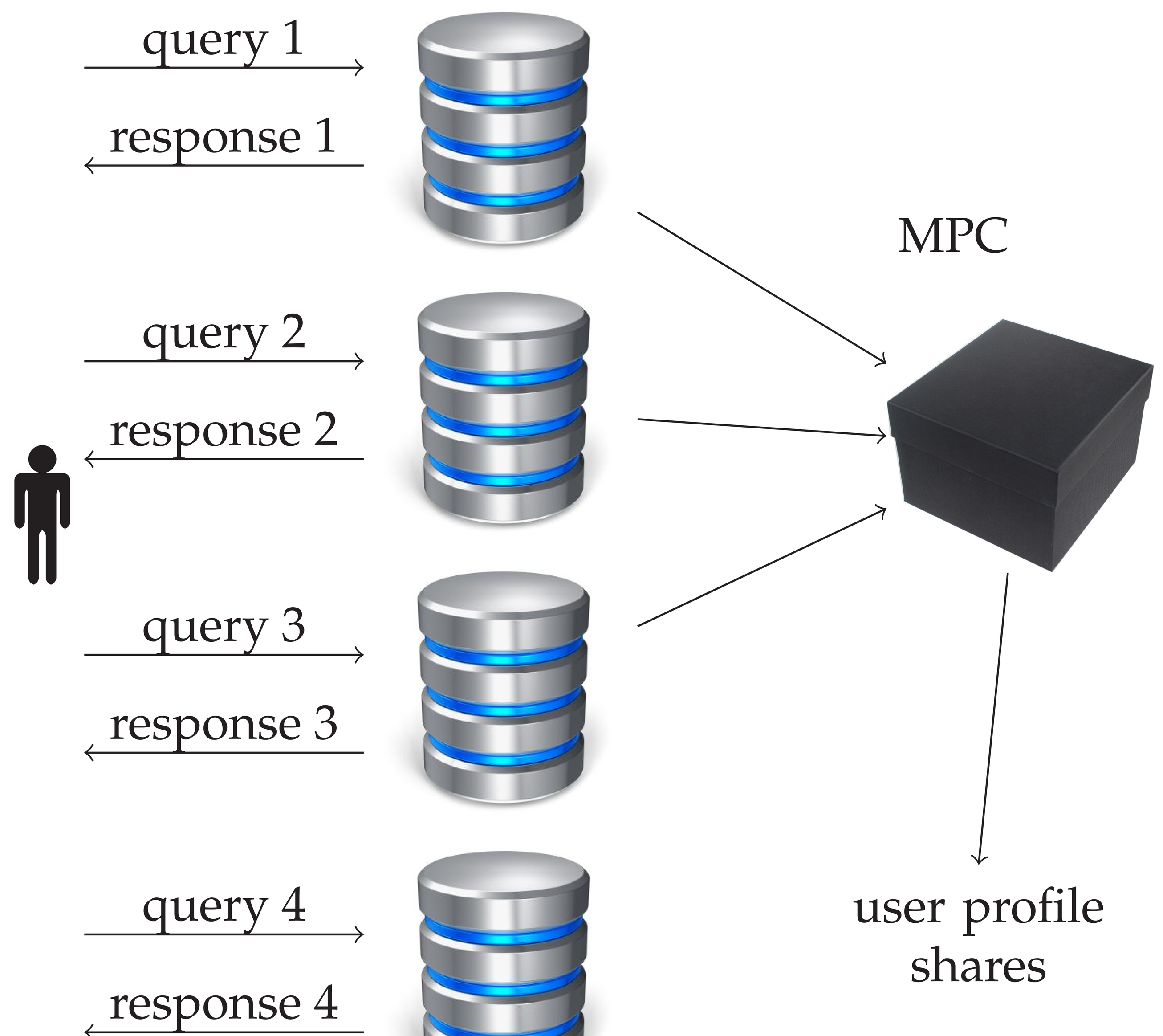


$$\underbrace{\begin{bmatrix}1 & 1 & 0\\0 & 0 & 0\\1 & 0 & 1\\1 & 1 & 0\end{bmatrix}}_{\mathbf{M}} = \underbrace{\begin{bmatrix}u_{11} & u_{12}\\u_{21} & u_{22}\\u_{31} & u_{32}\\u_{41} & u_{42}\end{bmatrix}}_{\mathbf{U}} \times \underbrace{\begin{bmatrix}v_{11} & v_{12} & v_{13}\\v_{21} & v_{22} & v_{23}\end{bmatrix}}_{\mathbf{V}}$$

1. $\mathbf{M}_{ij} = 1$ if a user i has queried for item j, otherwise $\mathbf{M}_{ij} = 0$.
2. Find $\mathbf{U}$, $\mathbf{V}$ (for some λ, μ) which minimizes: $\sum_{\mathbf{M}_{ij}=1}(\mathbf{M}_{ij} - \langle \mathbf{U}_i, \mathbf{V}_j^T \rangle)^2 + \lambda\|\mathbf{U}\|_2 + \mu\|\mathbf{V}\|_2$.
3. For (i', j') such that $\mathbf{M}_{i'j'} = 0$, use $\langle \mathbf{U}_{i'}, \mathbf{V}_{j'}^T \rangle$ as the prediction.

Gradient descent is used to solve the optimization problem.

Our System



PIR:

1. Several replicas of the database.
2. To retrieve a record, users send different query vectors to each replica and get a response.
3. Individual query vectors reveal nothing about the retrieved record's index.
4. Users combine the responses to get the desired record.

MPC:

1. Keep collecting the PIR queries until the end of every epoch.
2. 3PC Protocol on the secret-shared data.
3. 3PC outputs secret shared user profiles.
4. Users reconstruct corresponding profiles.
5. Item profiles are public.

References

[1] Syed Mahbub Hafiz and Ryan Henry. A bit more than a bit is more than a bit better: Improved constructions for faster optimal-rate multiserver PIR.

[2] Boyle, Gilboa, and Ishai. Function secret sharing: Improvements and extensions. CCS '16.

Assumptions

- We use the recent Hafiz-Henry PIR protocol which is computationally optimum and has an optimal download cost.
- Its upload is made extremely low by using DPFs to encode the queries.
- The price that is paid: The protocol requires that no two servers collude.

Contact Information

Adithya avadapal@iu.edu

Fattaneh fbayatba@berkeley.edu

Ryan ryan.henry@ucalgary.ca