Curs 10
Privacy Enhancing Architectures

Homomorphic Encryption (HE)

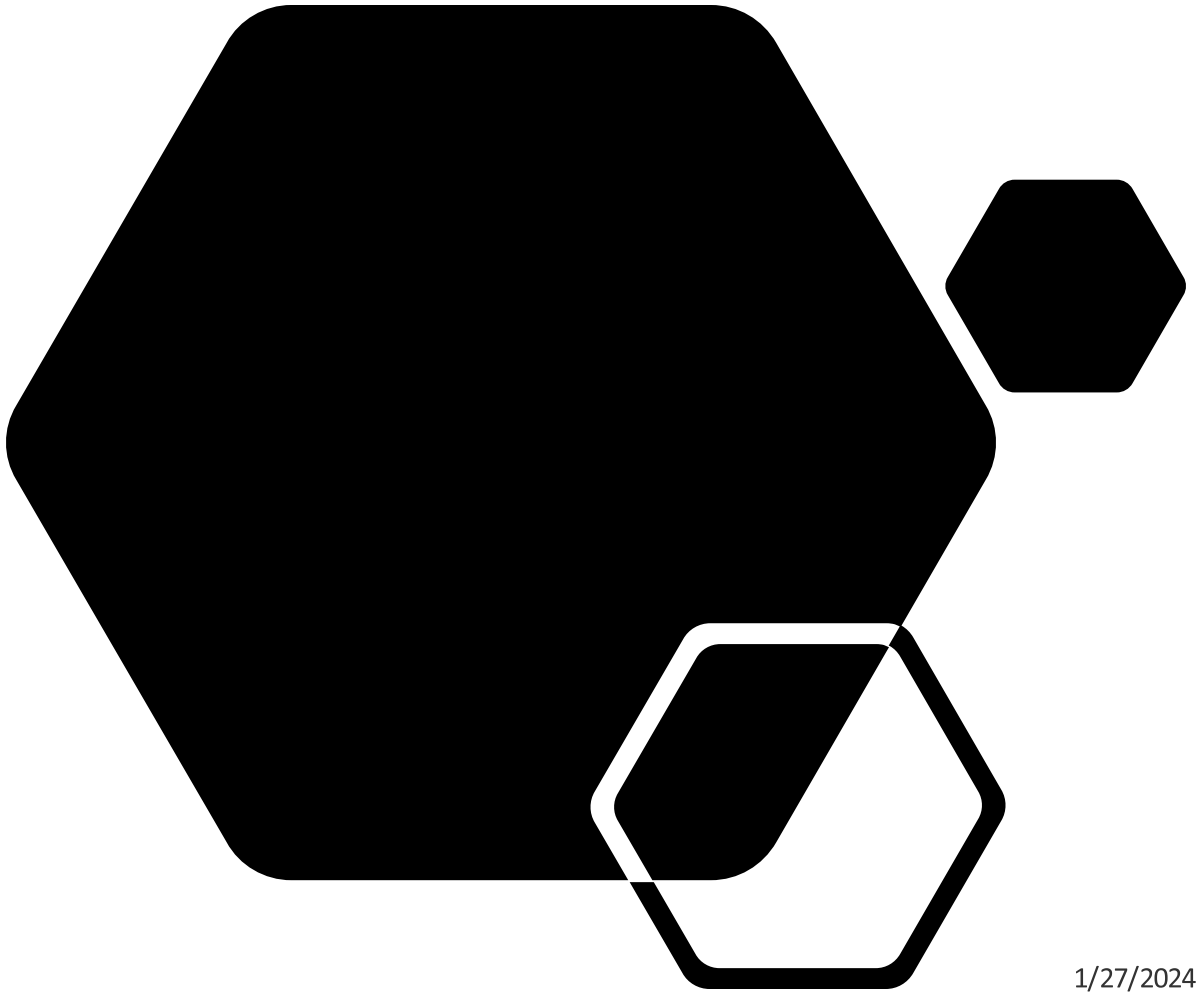
Private Information Retrieval (PIR)

1/27/2024

Course schedule

1. Why?
2. Cauzalitate
3. Măsurare
4. Modelare și eșantionare
5. Tehnici de analiză
   - Analiza factorială
   - Analiza cluster
   - Analiza de regresie
   - Analiza de rețea
   - Serii de timp
6. Predicție
7. Programare și ML
8. Why Privacy?
9. Privacy Enhancing Techniques
10. Homomorphic Encryption. PIR
11. Differential Privacy
12. Membership Inference Attacks
13. Federated Architecture. Multi-party computation
14. Explainable AI
15. Zero knowledge proof. Blockchain architecture

Homomorphic Encryption



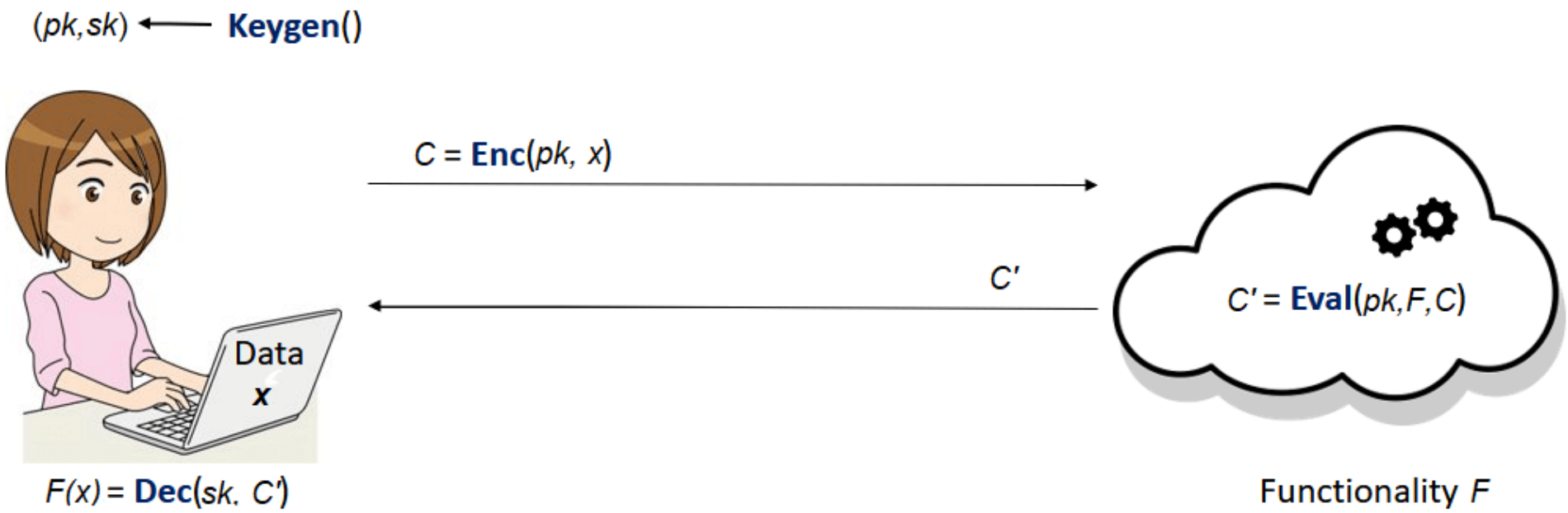
Context

- Cloud data processing
  - Data needs to be decrypted
- Decryption of data
  - Insecure
  - Costly
  - Time-consuming

Homomorphic Encryption (HE)

- Works on encrypted data
- Data can be processed without decryption
- Allows complex mathematical operations on the ciphertexts
- Safe even if a security breach occurs
- Based on the idea that all operations can be reduced to a series of additions and multiplications

Homomorphic Encryption (2)

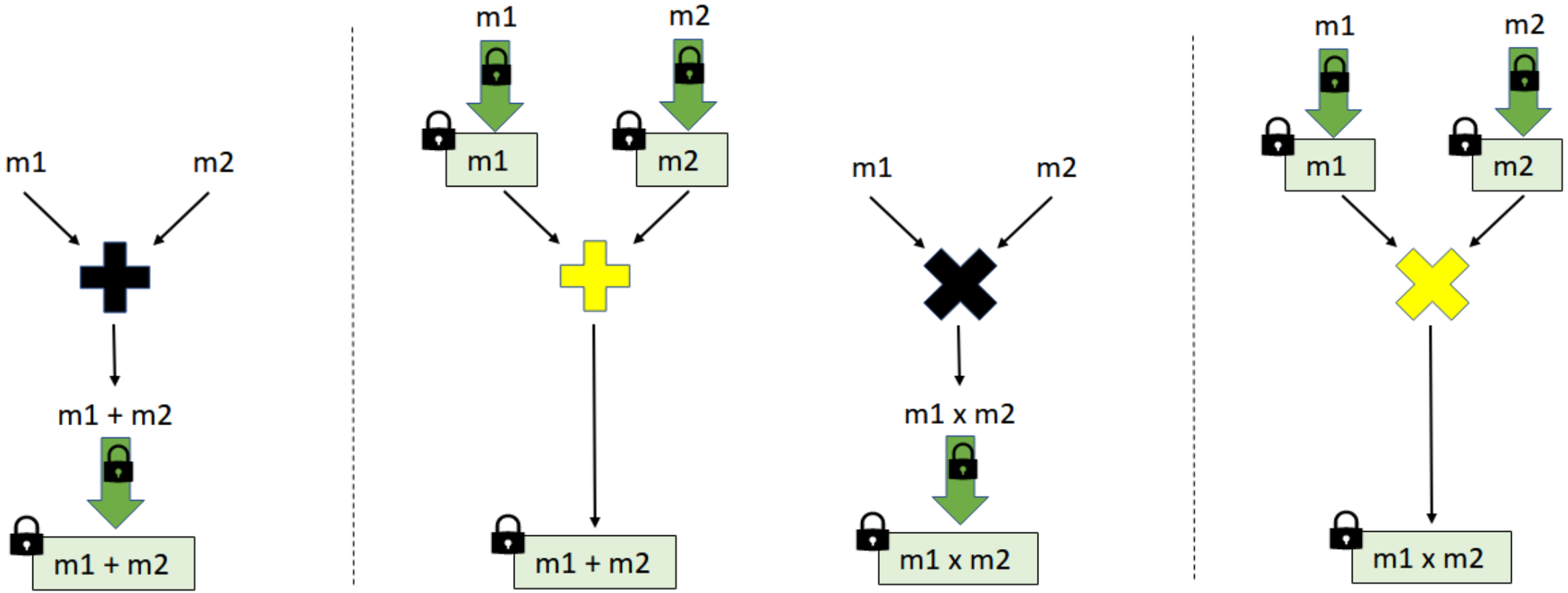


Source: https://bit-ml.github.io/blog/post/homomorphic-encryption-toy-implementation-in-python//

Closer Look to the Eval Algorithm

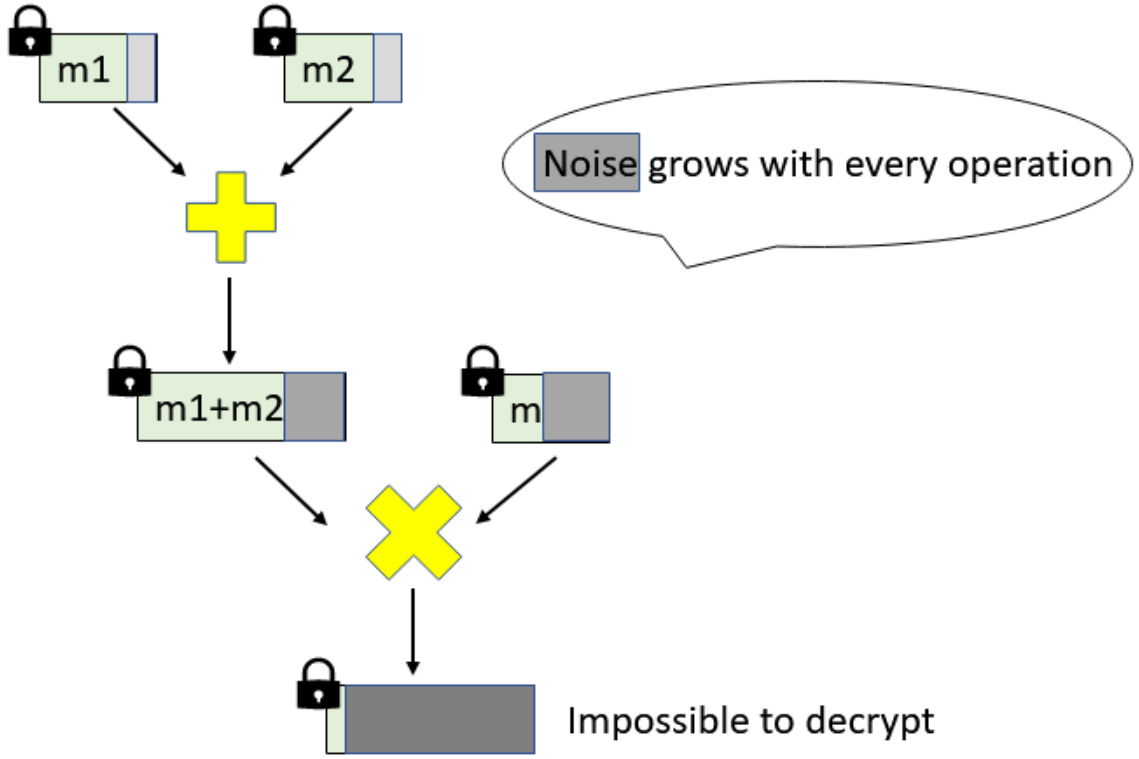
- All HE schemes are homomorphic with respect to two operations:
  - Some kind of addition (+ over integers, XOR over binary numbers)
  - Some kind of multiplication (* over integers, AND over binary numbers)

Closer Look to the Eval Algorithm (2)



Source: https://bit-ml.github.io/blog/post/homomorphic-encryption-toy-implementation-in-python//

Noise added during Eval



Source: https://bit-ml.github.io/blog/post/homomorphic-encryption-toy-implementation-in-python//

Types of HE

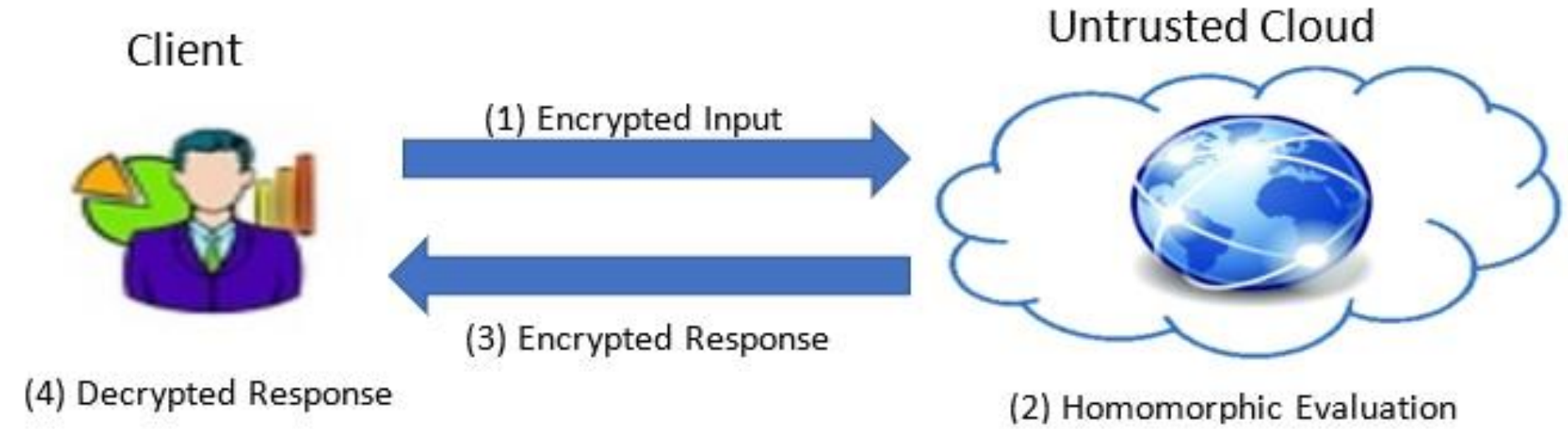
Partially HE (PHE)	• Actions: One (Addition or Multiplication) • Number of Operations: Unlimited
Somewhat HE (SHE)	• Actions: Two (Addition and Multiplication) • Number of Operations: Limited
Fully HE (FHE)	• Actions: Two (Addition and Multiplication) • Number of Operations: Unlimited

Use cases of HE

- Outsourced computing

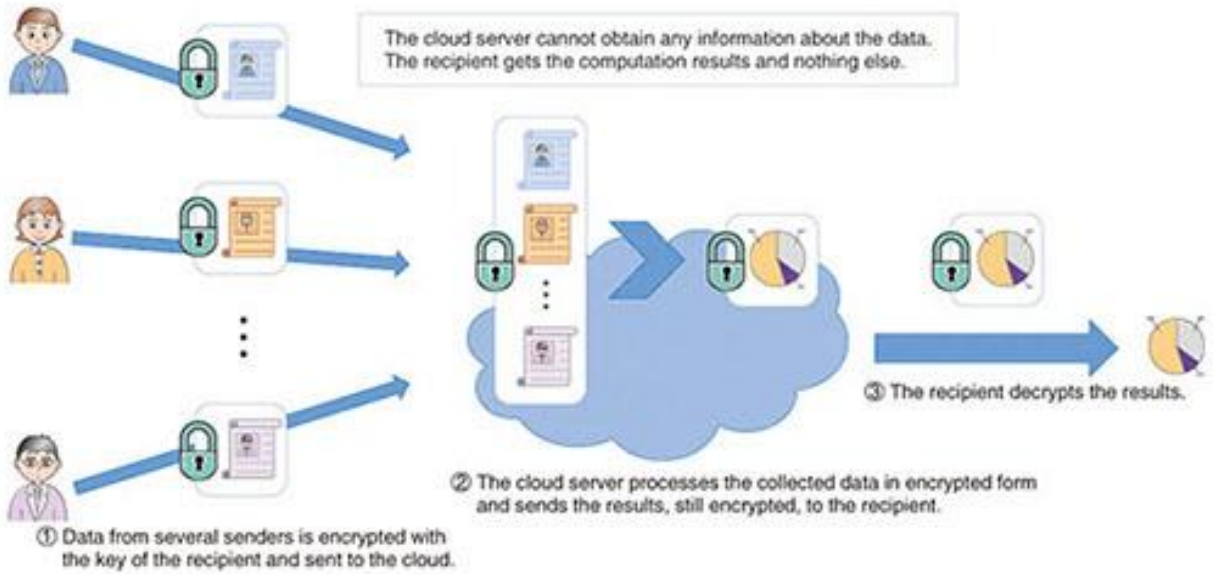
- Anonymous data processing

HE – Outsourced computing



Source: https://www.intel.com/content/www/us/en/developer/articles/technical/homomorphic-encryption/accelerating-homomorphic-encryption-for-fpga.html.html

HE – Anonymous data processing



Source: https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201407fa5_s.html/

Case Study – Hidden Vector Encryption (HVE)

An Efficient Privacy-Preserving System for Monitoring Mobile Users: Making Searchable Encryption Practical

Gabriel Ghinita
University of Massachusetts, Boston
gabriel.ghinita@umb.edu

Razvan Rughinis
Politehnica University, Bucharest
razvan.rughinis@cs.pub.ro

ABSTRACT

Monitoring location updates from mobile users has important applications in several areas, ranging from public safety and national security to social networks and advertising. However, sensitive information can be derived from movement patterns, so protecting the privacy of mobile users is a major concern. Users may only be willing to disclose their locations when some condition is met, for instance in proximity of a disaster area, or when an event of interest occurs nearby. Currently, such functionality is achieved using *searchable encryption*. Such cryptographic primitives provide provable guarantees for privacy, and allow decryption only when

mobile users wish to be immediately notified when their current location satisfies some condition, expressed as a spatial *search predicate*. For instance, in a public safety scenario, users want to be notified when they are getting close to a dangerous accident area. Alternatively, in the commercial domain, a user may want to be alerted when a retail store sales event is underway nearby.

The typical architecture of such a system uses a server that collects updates from the users and checks whether the alert condition is met. Such a service is provided by a commercial entity that is not fully trusted. Collection of user trajectories at a commercial site introduces serious privacy concerns, as sensitive personal in-

Idea

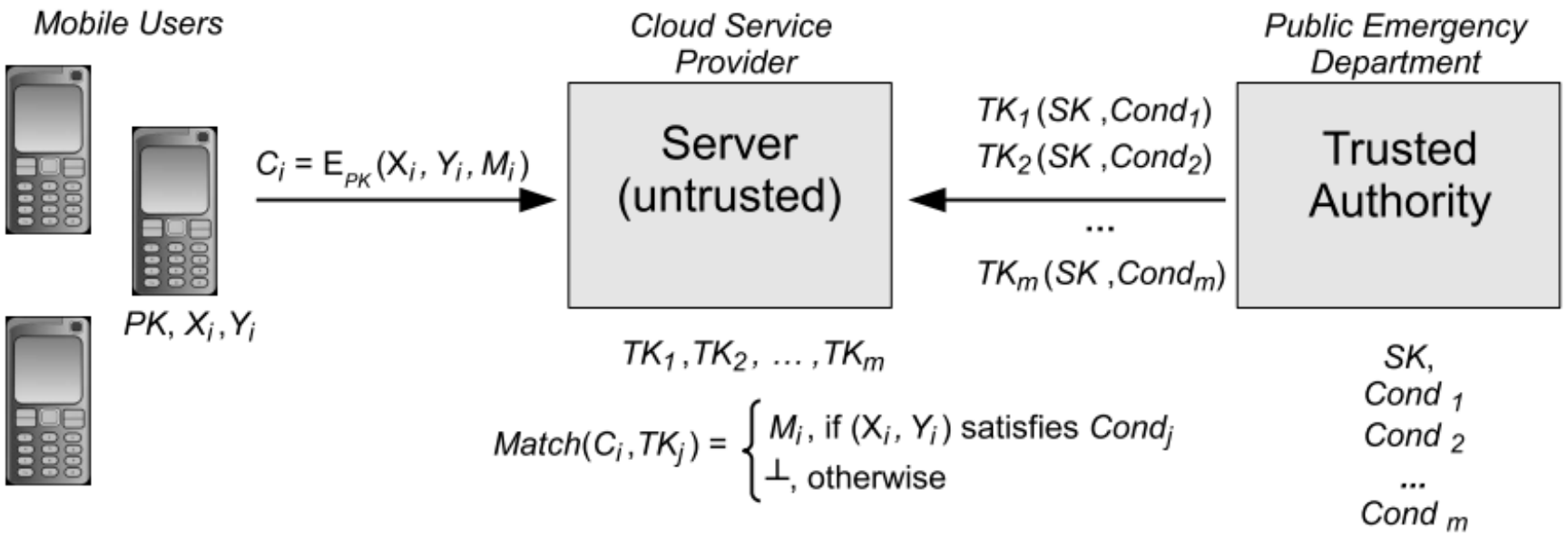
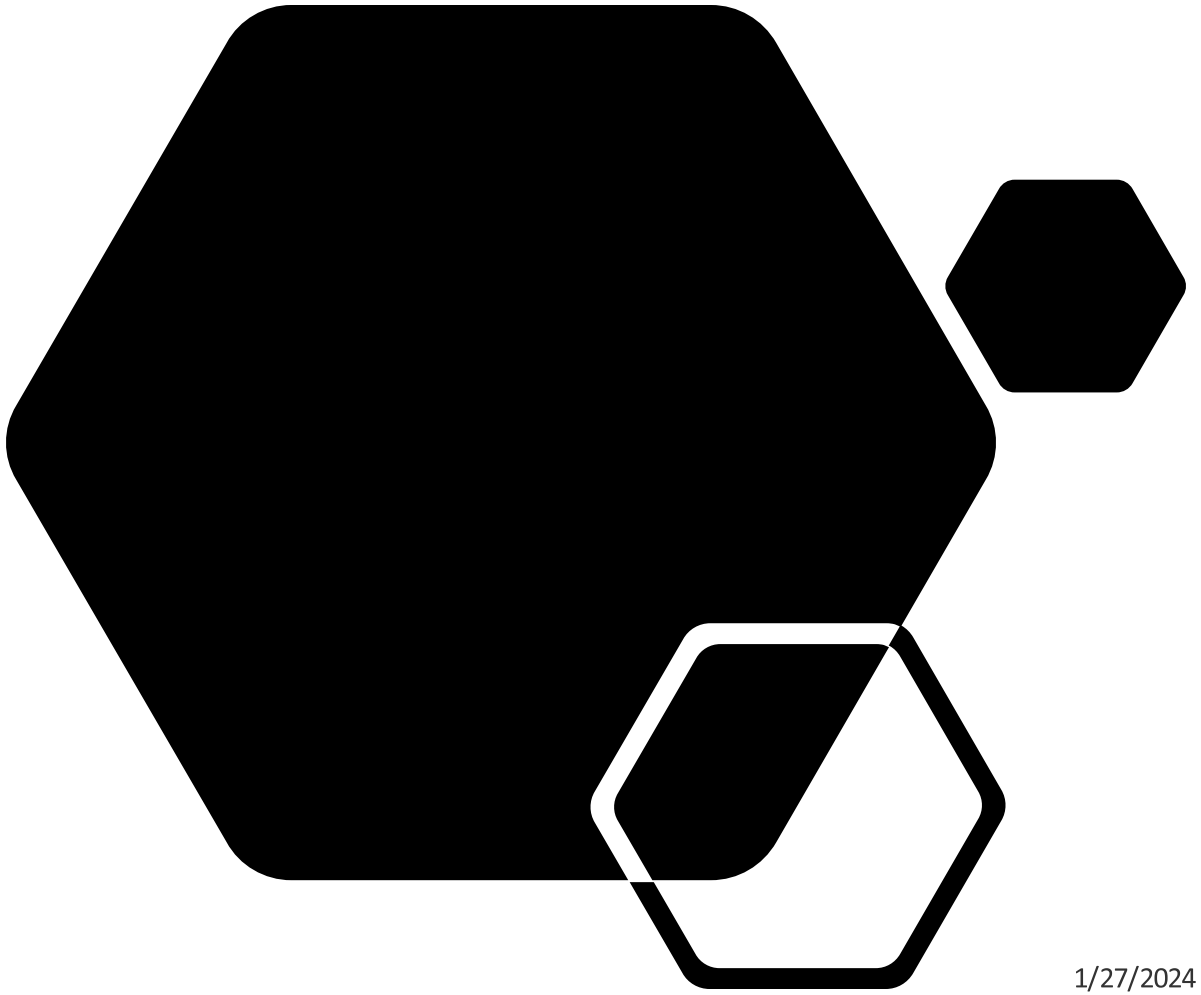


Figure 1: Location-based Alert System

Limitations of HE schemes

- Support for multiple users
- Large computational overhead

Private Information Retrieval



Private Information Retrieval

- A privacy setting
- Allows an user to download a messages from a set of messages from a system of databases without revealing the index of the required message to the databases
- Applications in medicine, finance, national defence

The basic PIR setting

- N non-colluding databases with
  - K independent messages
- Replicated databases
- Two constraints
  - User privacy constraint
  - Correctness constraint

The basic PIR setting (2)

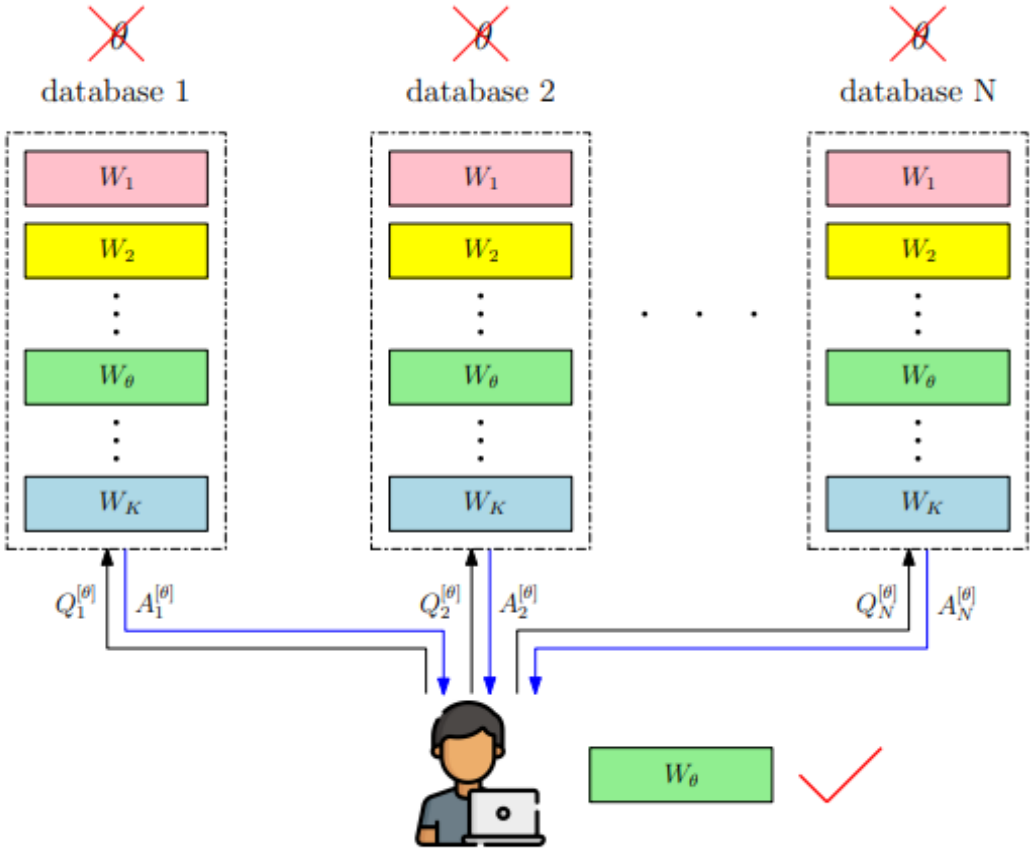


Fig. 2: The system model of PIR.

Vithana, S., Wang, Z., & Ulukus, S. (2023). Private Information Retrieval and Its Applications: An Introduction, Open Problems, Future Directions. arXiv preprint arXiv:2304.14397.

The easiest PIR implementation

- How can the user retrieve the needed data in the easiest way?

Advantages and Disadvantages

- Advantages:
  - High privacy level

- Disadvantages:
  - Huge communication costs
  - Retrieving of the message is done on the user side

PIR parameters

- PIR rate
  - Number of useful bits relative to number of downloaded bits
- PIR capacity
  - Maximum possible PIR rate

First PIR scheme proposed

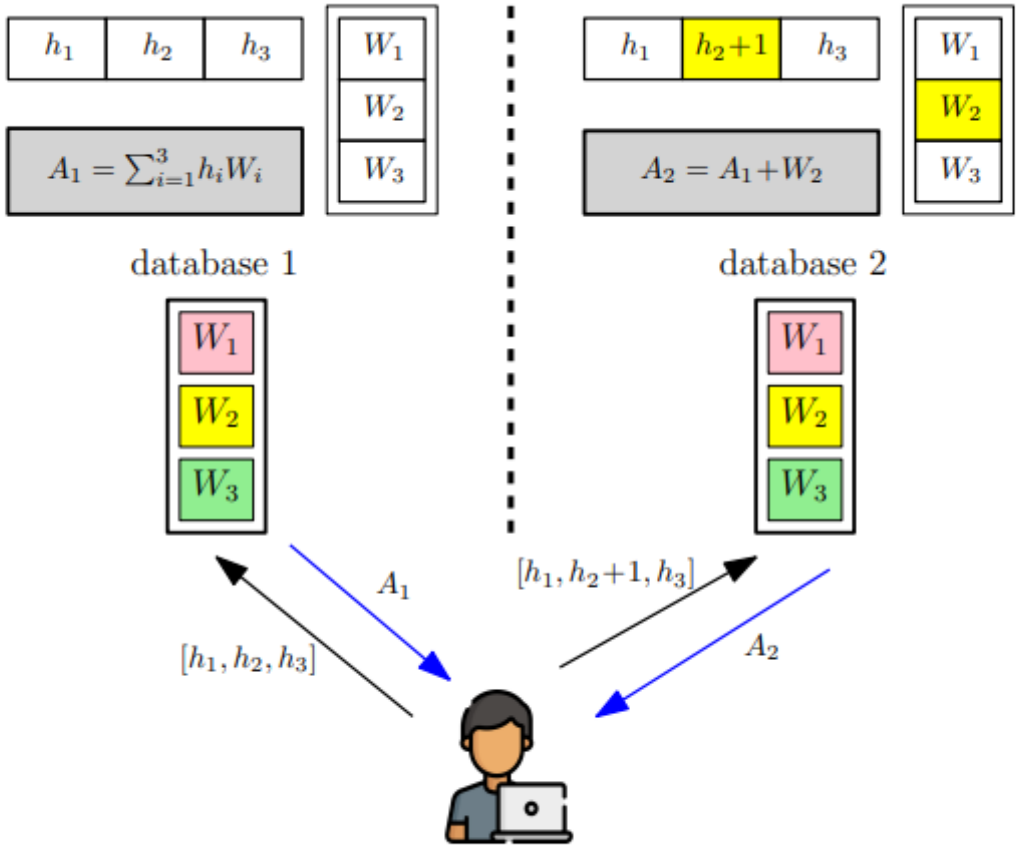


Fig. 3: PIR scheme in [1] for $N = 2$ and $K = 3$.

Vithana, S., Wang, Z., & Ulukus, S. (2023). Private Information Retrieval and Its Applications: An Introduction, Open Problems, Future Directions. arXiv preprint arXiv:2304.14397.

First PIR scheme proposed (2)

- What is the PIR rate of the scheme in the previous slide?
  - 2 downloaded messages – A1, A2
  - Only 1 needed message – W2
  - R = ½ = 50% → Can be improved

PIR approaches to maximize PIR rate

- Deterministic approach
- Probabilistic approach

Deterministic PIR

- Reuse bits from previous queries
  - Side information
- Divides messages in subpackets
- Enforces a message symmetry
  - A way to reduce the possibility of databases to derive the message index
- Reduces communication costs
- Increases PIR rate

Probabilistic PIR

- Based upon the idea that any query can be used to retrieve any message from the databases

- With or without message symmetry

SPIR – Symmetric Private Information Retrieval

- An extension of classical PIR
- Adds a new constraint
  - Database privacy constraint
- Reduces PIR rate

SPIR – Symmetric Private Information Retrieval (2)

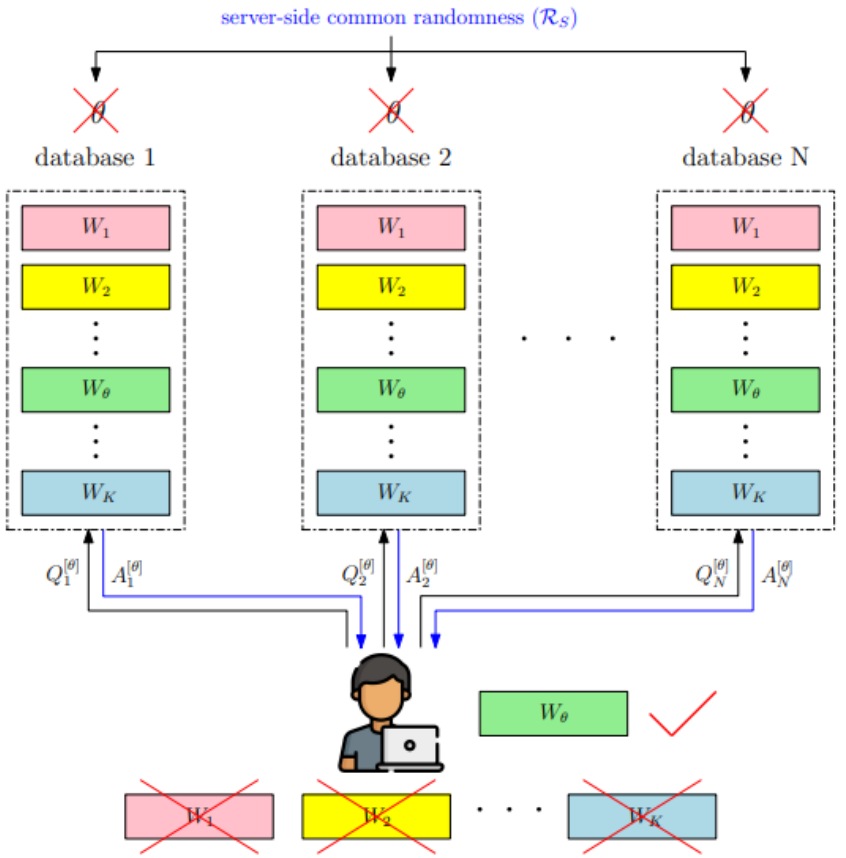


Fig. 6: The system model of SPIR.

Vithana, S., Wang, Z., & Ulukus, S. (2023). Private Information Retrieval and Its Applications: An Introduction, Open Problems, Future Directions. arXiv preprint arXiv:2304.14397.

Conceptual extensions of PIR

- PSI – Private Set Intersection
- PSU – Private Set Union
- PRUW – Private Read Update Write

PSI

- Two parties, P1 and P2, want to determine common elements without disclosing any additional information

- Can be reduced to a flavour of SPIR
  - The one that starts the communication is considered the user
  - The other one is considered the databases system

PSU

- Two parties, P1 and P2, want to determine the union of their two datasets without revealing any additional information

- Introduces two more privacy constraints:
  - P1 remaining elements privacy constraint
  - P2 remaining elements privacy constraint

- Can be reduced to PSI by using De Morgan's law

PRUW

- An user reads data from the databases system, updates this data and writes back update into the databases system

- Used in distributed machine learning

Case study – PIR on GPUs

Towards Efficient Private Spatial Information Retrieval Using GPUs

Mihai Maruseac[1], Gabriel Ghinita[1], Ming Ouyang[1], Razvan Rughinis[2]
[1]UMass Boston, [2]Politehnica University
{mmarusea,gghinita,mouyang}@cs.umb.edu, razvan.rughinis@cs.pub.ro

ABSTRACT

Latest generation mobile devices allow users to receive services tailored to their current locations. Location-based service providers perform spatial queries based on the user locations, but may also share them with various third parties. User whereabouts may disclose sensitive details about an individual's health status, political views or lifestyle choices, and therefore must be thoroughly protected. *Private information retrieval (PIR)* methods support blind execution of range and NN queries with cryptographic-strength security, but incur significant performance overhead. We employ graphical processing units (GPUs) to speed up the crypto operations required by PIR. We identify the challenges that arise when using GPUs for this purpose, and we propose solutions to address them. To the best of our knowledge, this is the first work to use GPUs for efficient private spatial information retrieval, and an important first step towards GPU-based acceleration of a broader range of secure spatial data operations.

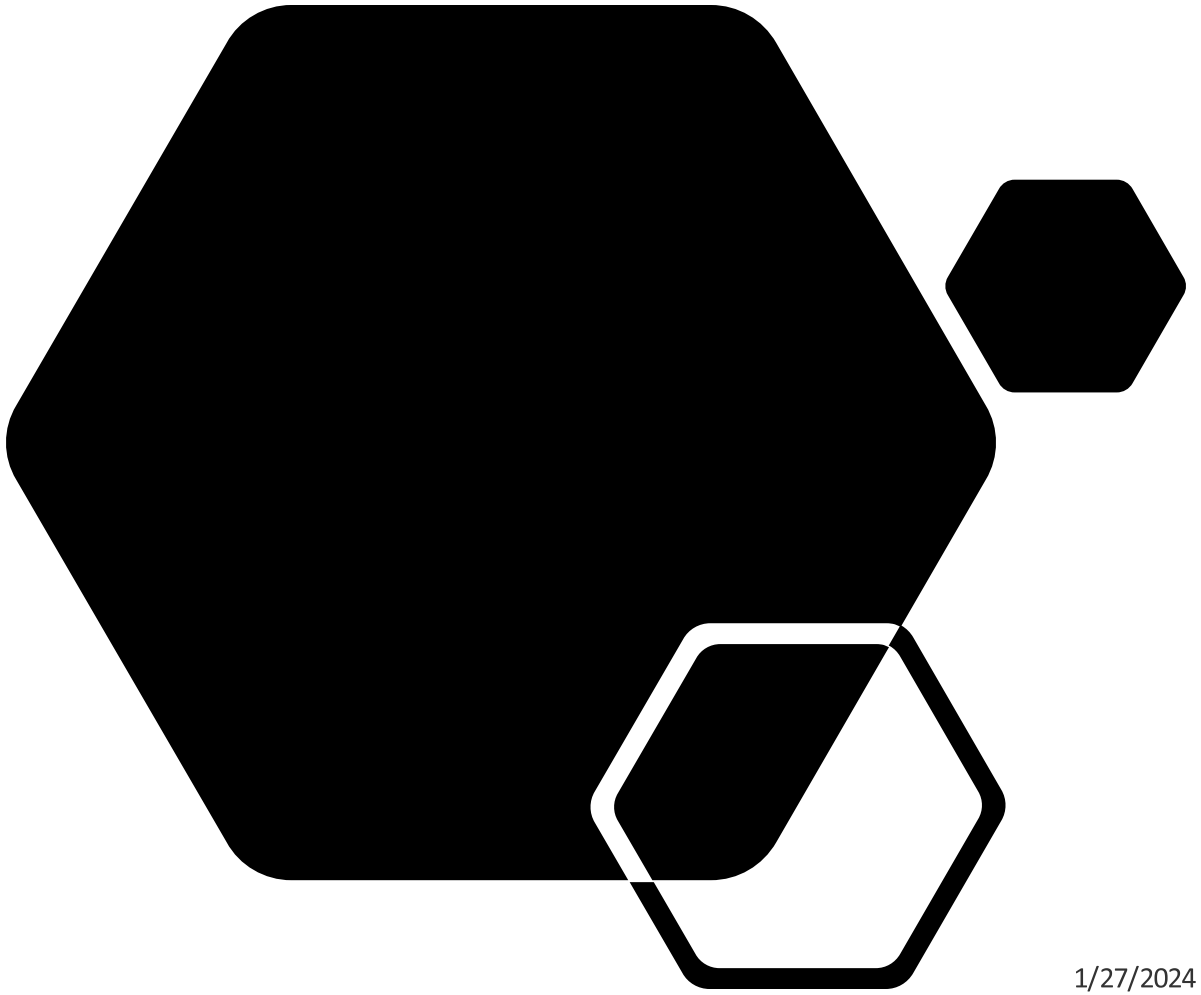
fail to protect against adversaries with background knowledge, and are not suitable for continuous queries by moving users. In cryptographic techniques [2, 1] users send to the SP only their encrypted coordinates. Next, the SP executes a *private information retrieval (PIR)* protocol that blindly processes the query, and returns as result an encrypted token that only the client can decode to obtain the query answer. Techniques in this category are provably secure, but are computationally expensive, as they require large amounts of cryptographic primitive evaluations.

We investigate the use of *graphics processing units (GPUs)* and *Compute Unified Device Architecture (CUDA)* to speed up the execution of private spatial information retrieval techniques. GPU devices consist of large numbers (i.e., thousands) of simple computing cores that are able to perform basic operations in parallel. However, the execution and programming model of GPUs is significantly different than general-purpose CPUs. GPUs have small amounts of memory resources, and rigid patterns of data access that

Open directions for future PIR research

- Multi-client PRUW
- Fundamental limits of PRUW in FL
- Separated PR / PW per model in DML / FL

Conclusions



1/27/2024

Conclusions

- HE protects data from adversaries by using encryption schemes
  - Eval algorithm
  - PHE, SHE, FHE
  - Outsourced computing
  - Anonymous Data Processing

- PIR offers privacy guarantees to an user that does not want to disclose the data he needs from a database
  - PIR rate, PIR capacity
  - Deterministic PIR, Probabilistic PIR
  - SPIR
  - PSI, PSU, PRUW

References

1. https://venafi.com/blog/homomorphic-encryption-what-it-and-how-it-used/

2. https://bit-ml.github.io/blog/post/homomorphic-encryption-toy-implementation-in-python/

3. Ghinita, G., & Rughinis, R. (2014, March). An efficient privacy-preserving system for monitoring mobile users: making searchable encryption practical. In Proceedings of the 4th ACM conference on Data and application security and privacy (pp. 321-332).

4. Vithana, S., Wang, Z., & Ulukus, S. (2023). Private Information Retrieval and Its Applications: An Introduction, Open Problems, Future Directions. arXiv preprint arXiv:2304.14397.

5. Maruseac, M., Ghinita, G., Ouyang, M., & Rughinis, R. (2014, November). Towards efficient private spatial information retrieval using gpus. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 405-408).