Discussion Question & Answers

Privacy by Design

A Strategic Framework for Choosing Between Encryption and Noise

Group Members:

Max Forrest	(30301436)	Sukhmanveer Singh	(30291188)
Chirag Kumar	(30273579)	Karanveer Singh	(30282286)

Question 1:

Explain the principle of composition in Differential Privacy, focusing on how privacy loss accumulates across repeated analyses of the same dataset. Explain why the privacy budget (ϵ) is considered as both an operational challenge and a unique strength of Differential Privacy framework. In your response, analyze and discuss the strategic trade-offs organizations face when deciding whether to allocate more of the privacy budget to immediate, high-accuracy aggregate reporting or to preserve it for future exploratory analysis.

Answer 1:

The principle of composition is a core idea in Differential Privacy and explains the cumulative nature of privacy loss. It describes how privacy guarantees degrade in a quantifiable and predictable way as multiple queries are run on the same dataset. Each analysis consumes a portion of the privacy budget (ϵ), and when several analyses are performed, mathematically the total privacy loss is simply the sum of the individual ϵ values. This makes the privacy budget a finite and non-renewable resource that represents the maximum amount of information that can be revealed about any individual (maximum allowable information leakage from a dataset). Unlike traditional anonymization methods that fail unpredictably under linkage attacks, DP uses composition to provide a formal upper bound on how much any individual's data can influence the total output over time.

The privacy budget is both an operational challenge and a strength of Differential Privacy. It poses practical difficulties and challenges because every query, regardless of how minor, consumes part of the same limited budget, making the framework less suited to open-ended or exploratory analysis where the number of queries cannot be anticipated. At the same time, this constraint what makes Differential Privacy robust. By requiring organizations to track and account for cumulative privacy loss, it transforms privacy from a vague promise into a measurable and auditable risk-management process with clear guarantees.

Organizations face significant strategic trade-offs when determining how to allocate the privacy budget (ϵ). Choosing a large ϵ value provides higher data accuracy and less injected noise for immediate reports, which is beneficial for stakeholders requiring precise aggregate statistics. A small ϵ value provides stronger privacy but requires more noise, potentially reducing the usefulness of the results. Deciding whether to exhaust the privacy budget

upfront or reserve it for future, unforeseen analytical needs is ultimately a policy and risk-management choice rather than a purely technical decision, requiring a balance between the business value of the data and the organization's tolerance for privacy risk.

Question 2:

A small medical clinic wants to use a cloud service to analyze their patient health records and identify patterns in treatment outcomes. The clinic is worried about patient health because they would be uploading sensitive medical data to someone else's servers. A technology consultant suggests using Homomorphic Encryption to keep the data safe.
In your opinion, would you recommend Homomorphic Encryption for this clinic? Why or why not? Explain your reasoning.

Answer 2:

I would not recommend Homomorphic encryption for this small medical clinic, despite its strong privacy protections because of the following reasons:

1. **Scale mismatch:** A small clinic likely has limited financial resources and technical expertise, making HE's high implementation and operational cost impractical.
2. **Performance concerns:** If the clinic needs timely insights to improve patient care, HE's slow processing speed would create unacceptable delays.
3. **Similar alternatives exist:** The clinic could use Differential Privacy which provides good privacy protection with much better performance and lower costs or they could use HIPAA-compliant cloud service with strong security measures and proper legal agreements.

When HE would make sense: HE would be more appropriate if this was a large hospital system or research facility with substantial resources, working with extremely sensitive data, needing to collaborate across multiple institutions and able to accept batch processing delays. In those high-stake scenarios, HE's absolute confidentiality guarantees justify the cost and complexity.

Better recommendation for the small clinic: The clinic should look into HIPAA-compliant cloud services that use strong encryption system(encrypting data at rest and in transit), proper access controls and clear legal agreements about data handling. If they need privacy-preserving analytics, Differential Privacy might offer a better balance of privacy protection, cost and performance for their needs.

Question 3:

Privacy-preserving technologies provide safeguards against specific threat models that are relevant to an environment. By adding noise to Homomorphic Encryption in a hybrid approach, what threat models might be addressed that Homomorphic Encryption alone does not? Why is it worth consideration?

Answer 3:

Homomorphic encryption is primarily designed to address the threat model of semi-trusted data processors accessing private data during processing. Alone, HE does nothing address the threat model of a data requester using their requests to identify a particular data donor whose data exists within the dataset. By adding noise to HE in a hybrid implementation, the techniques used in Differential Privacy can address the threat models addressed by DP.

Highly sensitive data that is accessed by trusted third parties and processed in the cloud is especially worth considering the use of a hybrid HE/DP approach for. Datasets suitable for HE can have DP noise added relatively simply, making it possible to add for any HE use-case. A genetics lab, for example, might use this format to share data with researchers while keeping patient data confidential and ensuring that their cloud host lacks access to the dataset.