COL872
Problem Set 1

Mallika Prabhakar (2019CS50440)
Sayam Sethi (2019CS10399)
Satwik Jain (2019CS10398)

January 2023

Contents

1 Question 1

Question 1: Lossy Encryption

Question. *Show that if* $\mathcal{E} = (\mathsf{Setup}, \mathsf{Setup\text{-}Lossy}, \mathsf{Enc}, \mathsf{Dec})$ *is a lossy encryption scheme satisfying all the above properties, then* $\mathcal{E}' = (\mathsf{Setup}', \mathsf{Enc}', \mathsf{Dec}')$ *is also a correct and semantically secure public key encryption scheme.*

Proof. **To Prove:** $\mathcal{E}'$ is a correct and semanticaly secure PKE given $\mathcal{E}$ as described in the problem statement.

Correctness

For correctness, $\mathsf{Dec}'(\mathsf{Enc}'(m, pk), sk) = m$

By the definition of $\mathcal{E}'$, $\mathsf{Setup}' = \mathsf{Setup}$, $\mathsf{Dec}' = \mathsf{Dec}$, $\mathsf{Enc}' = \mathsf{Enc}$. In $\mathcal{E}'$, all the keys are a (pk, sk) pair generated by Setup, which is non-lossy hence the encrypted text can always be decrypted (correctness of $\mathcal{E}$).

$\mathsf{Dec}'(\mathsf{Enc}'(m, pk), sk) = \mathsf{Dec}(\mathsf{Enc}(m, pk)) = m$

Therefore the public key encryption scheme $\mathcal{E}'$ has the correctness property.

Semantic Security of PKE

Proof for semantic security of PKE $\mathcal{E}'$ by hybrid world model.

- **World 0**: m_0 is encrypted using Setup
  1. Challenger creates a public key pk and a secret key sk using Setup
  2. Challenger sends pk to the adversary
  3. The adversary sends two messages m_0, m_1 to the challenger
  4. The challenger encrypts m_0 using pk and sends it to the adversary
- **Hybrid 0**: m_0 is encrypted using $\mathsf{Setup\text{-}Lossy}$
  1. Challenger creates a public key pk using $\mathsf{Setup\text{-}Lossy}$
  2. Challenger sends pk to the adversary
  3. The adversary sends two messages m_0, m_1 to the challenger
  4. The challenger encrypts m_0 using pk and sends it to the adversary
- **Hybrid 1**: m_1 is encrypted using $\mathsf{Setup\text{-}Lossy}$
  1. Challenger creates a public key pk using $\mathsf{Setup\text{-}Lossy}$
  2. Challenger sends pk to the adversary
  3. The adversary sends two messages m_0, m_1 to the challenger
  4. The challenger encrypts m_1 using pk and sends it to the adversary
- **World 1**: m_1 is encrypted using Setup
  1. Challenger creates a public key pk and a secret key sk using Setup
  2. Challenger sends pk to the adversary
  3. The adversary sends two messages m_0, m_1 to the challenger

4. The challenger encrypts m_1 using pk and sends it to the adversary

World 0 and Hybrid 0: From indistinguishability of modes, we know that the public keys are sampled from computationally indistinguishable distributions. Therefore, if World 0 and Hybrid 0 were far apart, we would be able to come up with an adversary that breaks the indistinguishability of modes. Hence, these two worlds are close.

Hybrid 0 and Hybrid 1: We know that the distributions for any two messages are statistically indistinguishable in lossy mode. Therefore, no (unbounded) adversary can differentiate between these two hybrids.

Hybrid 1 and World 1: Similar argument follows as for closeness of World 0 and Hybrid 0.

Therefore, we have shown that if $\mathcal{E}$ is a secure lossy encryption scheme, then $\mathcal{E}'$ is a correct and semantically secure encryption scheme. $\square$

2 Question 2

Question 2: A Lossy Encryption Scheme based on LWE

Question. *In this problem, you will have to construct a lossy encryption mode for Regev encryption. The algorithms* Setup, Enc, Dec *are defined as in class (see Lecture Notes). You must define the* Setup-Lossy *algorithm, and then show that it is a secure lossy encryption scheme.*

Proof. Following are the steps to generate the desired lossy encryption scheme-

1. Definition: Lossy Setup Algorithm

$\mathsf{Setup\text{-}Lossy}(1^n)$: $pk = (A, b)$ where $A \leftarrow \mathbb{Z}_q^{n\times m}$ and $b \leftarrow \mathbb{Z}_q^m$

pk is the lossy public key.

2. Indistinguishability of the modes

To Prove: $pk \leftarrow \mathsf{Setup}(1^n)$ is computationally indistinguishable from $pk' \leftarrow \mathsf{Setup\text{-}Lossy}(1^n)$

$pk'(A, b)$ where $A \leftarrow \mathbb{Z}_q^{n\times m}$ and $b \leftarrow \mathbb{Z}_q^m$, therefore pk' is completely random.

$pk = (A, b)$ where $A \leftarrow \mathbb{Z}_q^{n\times m}$ and $b^T = s^T \cdot A + e^T$ where $s \leftarrow \mathbb{Z}_q^n$ is a secret and $e \leftarrow \chi^m$ is random error as per Regev's PKE Scheme.

Following LWE, the distribution of pk and pk' should be computationally indistinguishable as $b^T = s^T \cdot A + e^T$ and $b \leftarrow \mathbb{Z}_q^m$ are computationally indistinguishable due to LWE.

3. Statistical indistinguishability in the lossy mode

To Prove: given m_0, m_1;

$\{pk, \mathsf{Enc}(pk, m_0) : pk \leftarrow \mathsf{Setup\text{-}Lossy}(1^n)\} = \{pk, \mathsf{Enc}(pk, m_1) : pk \leftarrow \mathsf{Setup\text{-}Lossy}(1^n)\}$

Let's take m_0, $\mathsf{Enc}(m_0, pk_{lossy}) = (A \cdot r, b^T \cdot r + m_0 \times \frac{q}{2})$ where $r \leftarrow \{0,1\}^m$

Using Leftover Hash Lemma, $A \cdot r$ is same as a random vector (statistically). Therefore r can not be recovered from $A \cdot r$ and thereby $b^T \cdot r$ is random. $m_0 \times \frac{q}{2}$ + random is still random. Similarly, from m_1, above steps are repeated and we again arrive at a random vector. Therefore having sent encryption of either m_0 or m_1 in lossy encryption mode, it is statistically impossible to figure out which message was encrypted.

□

3 Question 3

3.1 Question 3.1

Question 3.1: Small Secrets LWE - Matrix Version

Question. *In this problem, you have to prove the indistinguishabllity of the matrix version of the ss-LWE assuming that the normal version of ss-LWE is computationally hard.*

Proof. We have the following Distributions:
$\mathcal{D}_1 = \{(A, B) : B = S.A + E, A \leftarrow \mathcal{Z}_q^{n\times m}, S \leftarrow \chi^{n\times n}, E \leftarrow \chi^{n\times m}\}$
and $\mathcal{D}_2 = \{(A, B) : A, B \leftarrow \mathcal{Z}_q^{n\times m}\}$
We have to show that the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are computationally indistinguishable.

Let's create n-1 hybrid distributions $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_{n-1}$ where

$$\mathcal{H}_i = \{(A, B) : B = [b_j] \text{ where } b_j = \begin{cases} j^{th} \text{ row of } S.A + E, & \text{if } j \leq i \\ \mathcal{Z}_q^{1\times m} & i < j \leq n-1 \end{cases}$$

Claim 3.1. *$\mathcal{D}_1$ and $\mathcal{H}_1$ are computationally indistinguishable.*

Proof. The distributions $\mathcal{D}_1$ and $\mathcal{H}_1$ differ only at the first row of the second matrix in an instance of the distribution. If an adversary is able to distinguish between the two distributions, this means it is able to distinguish between the rows of two distributions. This can be used to achieve a reduction that breaks ss-LWE$_{n,m,q,\chi}$. but this contradicts the fact that ss-LWE$_{n,m,q,\chi}$ is a hard computational problem.
So, $\mathcal{D}_1$ and $\mathcal{H}_1$ are computationally indistinguishable Distributions. □

Claim 3.2. *$\mathcal{H}_{i-1}$ and $\mathcal{H}_i$ are computationally indistinguishable.*

Proof. The distributions $\mathcal{H}_{i-1}$ and $\mathcal{H}_i$ differ only at the i^{th} row of the second matrix in an instance of the distribution. If an adversary is able to distinguish between the two distributions, this means it is able to distinguish between the rows of two distributions. This can be used to achieve a reduction that breaks ss-LWE$_{n,m,q,\chi}$. but this contradicts the fact that ss-LWE$_{n,m,q,\chi}$ is a hard computational problem.
So, $\mathcal{H}_{i-1}$ and $\mathcal{H}_i$ are computationally indistinguishable Distributions. □

Claim 3.3. *$\mathcal{H}_{n-1}$ and $\mathcal{D}_2$ are computationally indistinguishable.*

Proof. The distributions $\mathcal{H}_{n-1}$ and $\mathcal{D}_2$ differ only at the last row of the second matrix in an instance of the distribution. If an adversary is able to distinguish between the two distributions, this means it is able to distinguish between the rows of two distributions. This can be used to achieve a reduction that breaks ss-LWE$_{n,m,q,\chi}$. but this contradicts the fact that ss-LWE$_{n,m,q,\chi}$ is a hard computational problem.
So, $\mathcal{H}_{n-1}$ and $\mathcal{D}_2$ are computationally indistinguishable Distributions. □

Using the claims proven above, we can say that the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are computationally indistinguishable.

□

3.2 Question 3.2

Question 3.2: xyz

Question. *ques*

Proof. this proof □

3.3 Question 3.3

Question 3.3: xyz

Question. *ques*

Proof. this proof □

4 Question 4

Question 4: Full Rank Matrices vs Noisy Low Rank Matrices

Question. *Let* $\mathcal{D} = \{(A, A.B + E) : A \leftarrow \mathcal{Z}_q^{2n\times n}, B \leftarrow \mathcal{Z}_q^{n\times 2n}, E \leftarrow \chi^{2n\times 2n}\}$.
Show that $\mathcal{D}$ *is computationally indistinguishable from the uniform distribution over* $\mathcal{Z}_q^{2n\times n} \times \mathcal{Z}_q^{2n\times 2n}$.

Proof. Let us start by defining 3 Distributions over $\mathcal{Z}_q^{n\times 2n} \times \mathcal{Z}_q^{n\times 2n}$:
$\mathcal{D}_0 = \{(A^T, B^T.A^T + E^T) : A^T \leftarrow \mathcal{Z}_q^{n\times 2n}, B^T \leftarrow \mathcal{Z}_q^{2n\times n}, E^T \leftarrow \chi^{2n\times 2n}\}$,

$$\mathcal{D}_1 = \{(A^T, p) \text{ where } p = [p_{ij}] \text{ and } p_{ij} = \begin{cases} e_{ij} + \sum_{k=0}^{n} b_{ik} \times a_{kj}, & \text{if } i \leq n \\ q, q \leftarrow \mathcal{Z}_q, & \text{if } i > n \end{cases}$$

where $b_{ik} \leftarrow B^T, a_{kj} \leftarrow A^T, e_{ij} \leftarrow E^T$
and
$\mathcal{D}_2 = \{(A^T, U^T) : A^T \leftarrow \mathcal{Z}_q^{n\times 2n}, U^T \leftarrow \mathcal{Z}_q^{2n\times 2n}\}$.

Claim 4.1. *Distributions* $\mathcal{D}_0$ *and* $\mathcal{D}_1$ *are computationally indistinguishable.*

Proof. The Distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ differ only at the last n rows of the second matrix which forms a submatrix of dimensions $n \times 2n$.
We can divide the Distribution $\mathcal{D}_0$ into two Distributions $\mathcal{D}_{01}$ and $\mathcal{D}_{02}$ over $\mathcal{Z}_q^{n\times 2n} \times \mathcal{Z}_q^{n\times 2n}$ where $\mathcal{D}_{01}$ consists of first n rows of Second matrix in $\mathcal{D}_0$ and $\mathcal{D}_{02}$ consists of last n rows of Second Matrix.
Similarly we can create Distributions $\mathcal{D}_{11}$ and $\mathcal{D}_{12}$ from $\mathcal{D}_1$.
Now We can see that the Distributions $\mathcal{D}_{01}$ and $\mathcal{D}_{11}$ are identical.
So to show $\mathcal{D}_0$ and $\mathcal{D}_1$ are computationally indistinguishable, it suffices to show that $\mathcal{D}_{02}$ and $\mathcal{D}_{12}$ are computationally indistinguishable.

now, $\mathcal{D}_{02} = \{(A^T, F) : F = B'.A^T + E', A^T \leftarrow \mathcal{Z}_q^{n\times 2n}, B' \leftarrow \mathcal{Z}_q^{n\times n}, E' \leftarrow \mathcal{Z}_q^{n\times 2n}\}$
B' is bottom n rows of B^T, E' is bottom n rows of E^T, and
$\mathcal{D}_{12} = \{(A^T, F) : A^T, F \leftarrow \mathcal{Z}_q^{n\times 2n}\}$

Form Questions 3, we know about the matrix version of Small Secrets LWE. As LWE and Small Secrets LWE are equally hard, we can get the matrix version of LWE.
Using the matrix Version of LWE, $\mathcal{D}_{02}$ and $\mathcal{D}_{12}$ are computationally indistinguishable.
Therefore, Distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ are computationally indistinguishable.

□

Claim 4.2. *Distributions* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *are computationally indistinguishable.*

Proof. The Distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ differ only at the first n rows of the second matrix which forms a submatrix of dimensions $n \times 2n$.
We can divide the Distribution $\mathcal{D}_1$ into two Distributions $\mathcal{D}_{11}$ and $\mathcal{D}_{12}$ over $\mathcal{Z}_q^{n\times 2n} \times \mathcal{Z}_q^{n\times 2n}$ where $\mathcal{D}_{11}$ consists of first n rows of Second matrix in $\mathcal{D}_1$ and $\mathcal{D}_{12}$ consists of last n rows of Second Matrix.
Similarly we can create Distributions $\mathcal{D}_{21}$ and $\mathcal{D}_{22}$ from $\mathcal{D}_2$.
Now We can see that the Distributions $\mathcal{D}_{12}$ and $\mathcal{D}_{22}$ are identical.

So to show $\mathcal{D}_1$ and $\mathcal{D}_2$ are computationally indistinguishable, it suffices to show that $\mathcal{D}_{11}$ and $\mathcal{D}_{21}$ are computationally indistinguishable.

now, $\mathcal{D}_{11} = \{(A^T, F) : F = B'.A^T + E', A^T \leftarrow \mathcal{Z}_q^{n\times 2n}, B' \leftarrow \mathcal{Z}_q^{n\times n}, E' \leftarrow \mathcal{Z}_q^{n\times 2n}\}$
B' is top n rows of B^T, E' is top n rows of E^T, and
$\mathcal{D}_{21} = \{(A^T, F) : A^T, F \leftarrow \mathcal{Z}_q^{n\times 2n}\}$

Using the matrix Version of LWE, $\mathcal{D}_{11}$ and $\mathcal{D}_{21}$ are computationally indistinguishable.
Therefore, Distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are computationally indistinguishable. $\square$

From the above two claims, we can say that the Distributions $\mathcal{D}_0$ and $\mathcal{D}_2$ are computationally indistinguishable.
As Transposing a matrix is just changing the positions of the elements in the matrix, it cannot change the matrix computationally. Using this, we can define two new distributions by just taking the transpose of the matrices in the computationally indistinguishable Matrices which will also be computationally indistinguishable.
Taking Transpose,
$\mathcal{D}'_0 = \{(A, A.B + E) : A \leftarrow \mathcal{Z}_q^{2n\times n}, B \leftarrow \mathcal{Z}_q^{n\times 2n}, E \leftarrow \chi^{2n\times 2n}\}$.
and
$\mathcal{D}'_2 = \{(A, U) : A \leftarrow \mathcal{Z}_q^{2n\times n}, U \leftarrow \mathcal{Z}_q^{2n\times 2n}\}$.
here $\mathcal{D}'_0 = \mathcal{D}$ and $\mathcal{D}'_2$ is a uniform distribution over $\mathcal{Z}_q^{2n\times n} \times \mathcal{Z}_q^{2n\times 2n}$.
Thus $\mathcal{D}$ is computationally indistinguishable from a Uniform Distribution over $\mathcal{Z}_q^{2n\times n} \times \mathcal{Z}_q^{2n\times 2n}$. $\square$

5 Question 5

Question 5: Random-looking t matrices with a special structure

Question. *Define distribution $\mathcal{D}$ that is indistinguishable from $\left(\mathbb{Z}_q^{n\times 2n}\right)^t$ and no subset sum of an element sampled from this distribuion has only small entries.*

Proof. We define the distribution as follows:

$$\begin{aligned}\mathcal{D} &= (\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_t)\\ \mathbf{B}_i &= [\mathbf{A}_i|\mathbf{C}_i], \forall i \in [t]\\ \mathbf{C}_i &= \mathbf{A}_i \cdot \mathbf{S}_i + \mathbf{E}_i + \mathbf{D}_i\\ &\mathbf{A}_i \leftarrow \mathbb{Z}_q^{n\times n}, \mathbf{S}_i \leftarrow \chi^{n\times n}, \mathbf{E}_i \leftarrow \chi^{n\times n}, \mathbf{D}_i \leftarrow d^{n\times n}\end{aligned} \tag{1}$$

We now define χ and d appropriately so that the required properties are satisfied. Consider any matrix $\mathbf{B}_p$ from the tuple sampled from $\mathcal{D}$. This is made up of $\mathbf{A}_p$ and $\mathbf{C}_p$. We will now choose d in such a way that exactly one of these two matrices can have *only small entries*. Now, if $\mathbf{A}_p$ does not have *only small entries*, any value of d will work. Therefore, we consider the case when $\mathbf{A}_p$ is made up of *only small entries*. Also, consider χ to be of the form $\mathsf{Unif}[-m, m]$ where we have to determine the value of m (which has to be some power of q for LWE to remain a hard computational problem, say α). Also, WLOG, we assume d to be positive and it lies between 0 to $q/2$. We now consider the *worst-case scenario* such that d has to be the largest possible value:

$$\begin{aligned}&\sum_{k\in[n]} (a_{ik} \cdot s_{kj}) + e_{ij} + d > q^{0.75}\\ &\implies d > q^{0.75} - e_{ij} - \sum_{k\in[n]} (a_{ik} \cdot s_{kj})\\ &\qquad > q^{0.75} + m + n \times \left(q^{0.75} \cdot m\right)\\ &\qquad > q^{0.75} + q^{\alpha} + n \cdot q^{\alpha+0.75}\\ &\implies d \geq 1 + q^{0.75} \times (1 + n \cdot q^{\alpha}) + q^{\alpha}\\ &\qquad \geq q^{0.75} \times (3 + n \cdot q^{\alpha})\\ &\qquad \geq (n+3) \cdot q^{\alpha+0.75}\\ &\qquad \geq q^{\epsilon+\alpha+0.75} = q^{\alpha'+0.75}\end{aligned} \tag{2}$$

We replace $(n+3)$ by q^{ϵ} since we are dealing with large numbers and $O(\log^2(q)) < O(q^{\gamma})$. Now, since we assumed that d has to be $\leq q/2 = O(q)$, therefore, $\alpha < \alpha' < 0.25$. Also, let $\alpha' + 0.75 = \beta$

$\square$

6 Question 6

6.1 Question 6.1

Question 6.1: Code Obfuscation

Question. *What is the issue with the attempt involving integers?*

Proof. The proposed **Attempt 1** does not ensure correctness since it is possible that for some other integer z', $\Sigma_i a_{i,z'_i} = 0$ since all entries other than a_{t,z_t} are randomly sampled. For instance, consider the following case when $a_{t,1-z_t} = -\Sigma_{i<t} a_{i,z_i} (\mod q)$. Then, $\mathsf{Eval}(\mathsf{Obf}(f_z), z) = \mathsf{Eval}(\mathsf{Obf}(f_z), z \oplus 1) = 1$, which is incorrect. □

6.2 Question 6.2, 6.3, 6.4

Question 6.1: Code Obfuscation

Question. *Propose an attempt at obfuscation of f_z using matrices and ideas developed in Question 5. Describe* Obf *and* Eval*. Show that your scheme satisfies correctness. Prove security of your scheme, assuming* $\mathsf{ss\text{-}LWE}_{n,m,q,\chi}$*. Note that t must be large for this proof to work. How large must t be?*

Proof. □