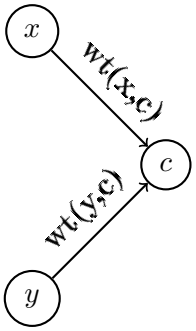
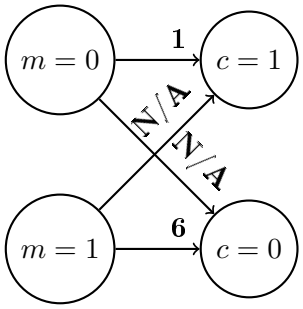
Homework 3

1. **Security of encryption schemes (8+8+8 points).** For each of the encryption schemes below, state whether the scheme is secure or not. Justify your answer in each case.

   (a) The message space is $\mathcal{M} = \{0, 1, \ldots, 6\}$. Algorithm Gen chooses a uniform key from the key space $\mathcal{K} = \{1, \ldots, 6\}$. The encryption algorithm $\mathsf{Enc}_{sk}(m)$ returns $(sk + m) \mod 7$, and the decryption algorithm $\mathsf{Dec}_{sk}(m)$ returns $(c - sk) \mod 7$.
   One of the properties of an encryption scheme states that a scheme is secure if and only if, $\forall x, y$ distinct, $wt(x, c) = wt(y, c)$:



   I will prove this by contradiction - we only need one example where this doesn't hold true. Let's look at the messages 0 and 1. Their graphs for a couple of encodings are shown below. The secret key that witnesses the message's encoding to the given ciphertext is labeled on the connecting line.



   Clearly, we can see that the weights are uneven. Here is the rundown:
   $wt(0, 0) = 0$, $wt(0, 1) = 1$
   $wt(1, 0) = 1$, $wt(1, 1) = 0$
   Since we stated earlier that, for a scheme to be secure, $wt(x, c) = wt(y, c)$. In this example, we can see that $wt(0, 0) = 0$ and $wt(1, 0) = 1$, which contradicts our earlier statement. Therefore, we can conclude that this scheme is insecure.

(b) The message space is $\mathcal{M} = \{0, 1, \ldots, 6\}$. Algorithm Gen chooses a uniform key from the key space $\mathcal{K} = \{0, 1, \ldots, 7\}$. The encryption algorithm $\mathsf{Enc}_{sk}(m)$ returns $(sk + m)$ mod 7, and the decryption algorithm $\mathsf{Dec}_{sk}(m)$ returns $(c - sk)$ mod 7.

Using the same logic as the last part, this scheme is secure. Reiterating, a scheme is secure if and only if $\forall x, y$ distinct, $wt(x, c) = wt(y, c)$. One way that this can be accomplished is if every possible message maps to every possible ciphertext exactly once (aka, all weights are 1). For proof, there is an illustration of the first 2 messages below. Due to the fact that the scheme is uniform (i.e., the encryption method doesn't change for each message), the pattern shown in the sample size below will continue as shown here.

Table 1: As you can see, every element mod 7 appears once, so all of the weights are the same.

M	SK	C
0	1	1
0	2	2
0	3	3
0	4	4
0	5	5
0	6	6
0	7	0
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	0
1	7	1

(c) The message space is $\mathcal{M} = \{1, 3, 5, \ldots, 2019, 2021\}$. Algorithm Gen chooses a uniform key from the key space $\mathcal{K} = \{0, 2, 4, 6, \ldots, 2020\}$. The encryption algorithm $\mathsf{Enc}_{sk}(m)$ returns $(sk + m) \mod 2022$, and the decryption algorithm $\mathsf{Dec}_{sk}(m)$ returns $(c - sk) \mod 2022$.

Again, this part will use the same logic as the previous two parts. In order for a private-key encryption scheme to be secure, for $\forall x, y$ distinct, $wt(x, c) = wt(y, c)$. Much like the previous part, one way that this can be accomplished is if every possible message maps to every possible ciphertext exactly once (aka, all weights are 1). You will see that, because of the circular pattern of the modulus operator, this scheme accomplishes this. Obviously, not all possible outcomes are shown here, but much like part (b), the pattern will continue.

Table 2: As you can see, if this were to be fully listed out, every element $\mod 2022$ would appear once, so all of the weights are the same.

M	SK	C
1	0	1
1	2	3
1	4	5
1	6	7
3	2020	1
3	0	3
3	2	5
3	4	7
5	2018	1
5	2020	3
5	0	5
5	2	7
7	2016	1
7	2018	3
7	2020	5
7	0	7

2. **Equivalent definition of Perfect Secrecy (15 points).** In the lecture we defined the perfect security for any private-key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ as follows. For any message m, cipher-text c, and a priori probability distribution $\mathbb{M}$ over the set of messages, we have:

$$\mathbb{P}\left[\mathbb{M} = m | \mathbb{C} = c\right] = \mathbb{P}\left[\mathbb{M} = m\right]$$

Show that the above definition is equivalent to the following alternative definition. For all messages m, m', cipher-text c, and a priori probability distribution $\mathbb{M}$ over the set of messages, we have:

$$\mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m\right] = \mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m'\right],$$

Remarks: (1) Proving equivalence means that you have to show that the first definition implies the second definition. And, the second definition also implies the first definition.

(2) Additionally, in this problem, for simplicity, assume that in the probability expressions no "division by error" occurs.

First, I will prove that $\mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m\right] = \mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m'\right]$. Using Baye's rule, we can do the following:

$\mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m\right] = \mathbb{P}\left[\mathbb{C} = c | \mathbb{M} = m'\right]$

$\frac{\mathbb{P}[\mathbb{C}=c, \mathbb{M}=m]}{\mathbb{P}[\mathbb{M}=m]} = \frac{\mathbb{P}\left[\mathbb{C}=c, \mathbb{M}=m'\right]}{\mathbb{P}[\mathbb{M}=m']}$

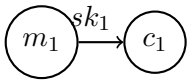
$\frac{\mathbb{P}[\mathbb{C}=c]*\mathbb{P}[\mathbb{M}=m]}{\mathbb{P}[\mathbb{M}=m]} = \frac{\mathbb{P}[\mathbb{C}=c]*\mathbb{P}\left[\mathbb{M}=m'\right]}{\mathbb{P}[\mathbb{M}=m']}$

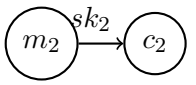
Remember that, since every message m is unique, the probability that $\mathbb{M}$ is some value is equally likely for all messages (aka, $\mathbb{P}\left[\mathbb{P} = m\right]$ is constant). Therefore, we can cancel those statements in the numerators and denominators, and we get:

$\mathbb{P}\left[\mathbb{C} = c\right] = \mathbb{P}\left[\mathbb{C} = c\right]$

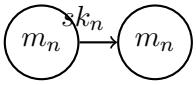
Therefore, the two statements are equivalent.

Next, we need to prove that the first and second definitions are equal. We know that, for perfect security, $\forall m, m'$ distinct, $wt(m, c) = wt(m', c)$. For simplicity's sake, let's say that $wt = 1$. This means that each message m uniquely maps to a ciphertext c, witnessed by a unique secret key sk, as illustrated below:





...



Since all m and c must be in the same message space, and each element in $\mathbb{M}$ and $\mathbb{C}$ are unique, $\mathbb{P}\left[\mathbb{M} = m\right] = \mathbb{P}\left[\mathbb{C} = c\right]$. Therefore, since the first definition evaluates to $\mathbb{P}\left[\mathbb{M} = m\right]$, and the second definition evaluates to $\mathbb{P}\left[\mathbb{C} = c\right]$, we can conclude that the two definitions are equivalent.

3. **Defining Perfect Security from Ciphertexts (15 points).** An upstart in the field of cryptography has proposed a new definition for perfect security of private-key encryption schemes. According to this new definition, a private-key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is perfectly secure, if, for all a priori distribution $\mathbb{M}$ over the message space, and any two cipher-texts c and c', we have the following identity.

$$\mathbb{P}[\mathbb{C} = c] = \mathbb{P}\left[\mathbb{C} = c'\right]$$

Show that the definition in the class does not imply this new definition.

Remark. You need to construct a private-key encryption scheme that is secure according to the definition we learned in the class. However, this scheme does not satisfy the new definition.

In class, we learned that a private key encryption scheme is secure if $\mathbb{P}\left[\mathbb{M} = m|\mathbb{C} = c\right] = \mathbb{P}[\mathbb{M} = m]$. In order to refute this, we need to construct a scheme in which possible ciphertexts are unevenly weighted. However, we still need to satisfy the definition learned in class. If we expand the definition from class using Baye's theorem, we get the following:

$$\mathbb{P}\left[\mathbb{M} = m|\mathbb{C} = c\right] = \frac{\mathbb{P}[\mathbb{M}=m]*\mathbb{P}[\mathbb{C}=c]}{\mathbb{P}[\mathbb{C}=c]} = \mathbb{P}[\mathbb{M} = m]$$

This simply means that, in order for a scheme to fit this definition, the ciphertext c must not depend on the input m. Imagine a private-key encryption scheme similar to one-time pad, except that when encrypting, there is a $\frac{2}{3}$ probability that a 1 will be appended to the resulting encrypted message, and a $\frac{1}{3}$ probability that a 0 will be appended to the resulting encrypted message. This means that the resulting ciphertext will NOT depend on the input message m.

Therefore, we can say that this scheme is completely secure based on the definition given in class ($\mathbb{P}\left[\mathbb{M} = m|\mathbb{C} = c\right] = \mathbb{P}[\mathbb{M} = m]$). However, we know that a ciphertext c with the last digit 1 is twice as likely as a ciphertext c with the last digit 0, so $\mathbb{P}[\mathbb{C} = c] \neq \mathbb{P}\left[\mathbb{C} = c'\right]$.

4. **One-time Pad for 4-Alphabet Words (8+8 points).** We interpret alphabets $\texttt{a}, \texttt{b}, \ldots, \texttt{z}$ as integers $0, 1, \ldots, 25$, respectively. We will work over the group $(\mathbb{Z}_{26}^4, +)$, where $+$ is coordinate-wise integer sum $\mod 26$. For example, $\texttt{abcx} + \texttt{aczd} = \texttt{adba}$.

   Now, consider the one-time pad encryption scheme over the group $(\mathbb{Z}_{26}^4, +)$.

   (a) What is the probability that the encryption of the message `kiwi` is the cipher text `kiwi`?

   In order for the encryption of `kiwi` to be `kiwi`, the secret key would have to be `aaaa`, since the letter `a` doesn't change the value of anything (its value is 0). Since every letter has an equal $\frac{1}{26}$ chance of being chosen, and there are 4 letters, the probability of this is $\frac{1}{26^4}$.

   (b) What is the probability that the encryption of the message `kiwi` is the cipher text `lime`?

   In order for the encryption of `kiwi` to be `lime`, the secret key would have to be `bapv`. Since every letter has an equal $\frac{1}{26}$ chance of being chosen, and there are 4 letters, the probability of this is $\frac{1}{26^4}$.

5. **Lagrange Interpolation(7+7+6 points).** We want to derive a part of the Chinese Remainder Theorem using principles of Lagrange Interpolation. Our goal is the following

Suppose p and q are two distinct primes. Suppose $a \in \{0, \ldots, p-1\}$ and $b \in \{0, \ldots, q-1\}$. We want to find a natural number x such that $$x \pmod p = a \text{ and } x \pmod q = b$$

We shall proceed towards this objective incrementally (similar to the approach of Lagrange interpolation).

(a) Find a natural number x_p satisfying $x_p \pmod p = 1$, and $x_p \pmod q = 0$.

Since $x_p \mod q = 0$, x_p must be a multiple of q, such that $x_p = nq$.
We also know that $x_p \mod p = 1$, so $x_p = np + 1$.
If we set these equations equal, we find the solution for n:

$nq = np + 1$
$nq - np = 1$
$n(q - p) = 1$
$n = \frac{1}{q-p}$

Plugging that into the previous equations, we get the solution for x_p:
$x_p = np + 1 = nq$

$x_p = \frac{p}{q-p} + 1 = \frac{q}{q-p}$

(b) Find a natural number x_q satisfying $x_q \pmod{p} = 0$ and $x_q \pmod{q} = 1$.

Since $x_q \mod p = 0$, x_q must be a multiple of p, such that $x_p = np$.
We also know that $x_q \mod q = 1$, so $x_q = nq + 1$.
If we set these equations equal, we find the solution for n:

$np = nq + 1$
$np - nq = 1$
$n(p - q) = 1$
$n = \frac{1}{p-q}$

Plugging that into the previous equations, we get the solution for x_q:
$x_q = nq + 1 = np$

$x_q = \frac{q}{p-q} + 1 = \frac{p}{p-q}$

(c) Find a natural number x satisfying $x \pmod{p} = a$ and $x \pmod{q} = b$.

From the start, we know that $x = np + a$ and $x = nq + b$, simply by the definition of the mod operator and the equations given to us. Setting these two equations equal gives us the solution for n:
$np + a = nq + b$
$np - nq = b - a$
$n(p - q) = b - a$
$n = \frac{b-a}{p-q}$

Plugging that into the previous equations, we get the solution for x:
$x = np + a = nq + b$

$x = \frac{p(b-a)}{p-q} + a = \frac{q(b-a)}{p-q} + b$

6. **An Illustrative Execution of Shamir's Secret Sharing Scheme (6+10+9 points).** We shall work over the field $(\mathbb{Z}_7, +, \times)$. We are interested in sharing a secret among 6 parties such that any 4 parties can reconstruct the secret, but no subset of 3 parties gain any additional information about the secret.

   Suppose the secret is $s = 3$. The random polynomial of degree < 4 that is chosen during the secret sharing steps is $p(X) = X^3 + 2X + 3$.

   (a) What are the respective secret shares of parties 1, 2, 3, 4, 5, and 6?

Party 1 - $p(X) = 1^3 + 2(1) + 3 = 6$
Party 2 - $p(X) = 2^3 + 2(2) + 3 = 15$
Party 3 - $p(X) = 3^3 + 2(3) + 3 = 36$
Party 4 - $p(X) = 4^3 + 2(4) + 3 = 75$
Party 5 - $p(X) = 5^3 + 2(5) + 3 = 138$
Party 6 - $p(X) = 6^3 + 2(6) + 3 = 231$

(b) Suppose parties 1, 2, 5, and 6 are interested in reconstructing the secret. Run Lagrange Interpolation algorithm as explained in the class.
(*Remark:* It is essential to show the step-wise reconstruction procedure to score full points. In particular, you need to write down the polynomials $p_1(X)$, $p_2(X)$, $p_3(X)$, and $p_4(X)$.)

Recall the secret shares for parties 1, 2, 5, and 6 are the following:
Party 1 - 6 $->$ $(x_1, y_1) = (1, 6)$
Party 2 - 15 $->$ $(x_2, y_2) = (2, 15)$
Party 5 - 138 $->$ $(x_3, y_3) = (5, 138)$
Party 6 - 231 $->$ $(x_4, y_4) = (6, 231)$

By Lagrange,
$p_1(X) = y_1 * \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)} = 6 * \frac{(x-2)(x-5)(x-6)}{(1-2)(1-5)(1-6)} = \frac{-3x^3}{10} + \frac{39x^2}{10} - \frac{78}{5} + 18$

$p_2(X) = y_2 * \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)} = 15 * \frac{(x-1)(x-5)(x-6)}{(2-1)(2-5)(2-6)} = \frac{5x^3}{4} - 15x^2 + \frac{205x}{4} - \frac{75}{2}$

$p_3(X) = y_3 * \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)} = 138 * \frac{(x-1)(x-2)(x-6)}{(5-1)(5-2)(5-6)} = -\frac{23x^3}{2} + \frac{207x^2}{2} - 230x + 138$

$p_4(X) = y_4 * \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)} = 231 * \frac{(x-1)(x-2)(x-5)}{(6-1)(6-2)(6-5)} = \frac{231x^3}{20} - \frac{462x^2}{5} + \frac{3927x}{20} - \frac{231}{2}$

We know that, to solve the equation, all we need is the following formula:
$p(X) = \sum p_i(X)$

$p(X) = p_1(X) + p_2(X) + p_3(X) + p_4(X)$

$p(X) = (\frac{-3x^3}{10} + \frac{39x^2}{10} - \frac{78}{5} + 18) + (\frac{5x^3}{4} - 15x^2 + \frac{205x}{4} - \frac{75}{2}) + (-\frac{23x^3}{2} + \frac{207x^2}{2} - 230x + 138) + (\frac{231x^3}{20} - \frac{462x^2}{5} + \frac{3927x}{20} - \frac{231}{2})$

$p(X) = x^3 + 0x^2 + 2x + 3$
$p(X) = x^3 + 2x + 3$

Which is the original polynomial.

(c) Suppose parties 1, 2, and 5 get together. Let $q_{\widetilde{s}}(X)$ be the polynomial that is consistent with their shares and the point $(0, \widetilde{s})$, for each $\widetilde{s} \in \mathbb{Z}_p$. Write down the polynomials $q_0(X)$, $q_1(X)$, ..., $q_6(X)$.
Party 1 - 6 –> $(x_1, y_1) = (1, 6)$
Party 2 - 15 –> $(x_2, y_2) = (2, 15)$
Party 5 - 138 –> $(x_3, y_3) = (5, 138)$

$q_0(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,0).**
$q_0(X) = \frac{13x^3}{10} - \frac{12x^2}{5} + \frac{71x}{10}$

$q_1(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,1).**
$q_1(X) = \frac{6x^3}{5} - \frac{8x^2}{5} + \frac{27x}{5} + 1$

$q_2(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,2).**
$q_2(X) = \frac{11x^3}{10} - \frac{4x^2}{5} + \frac{37x}{10} + 2$

$q_3(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,3).**
$q_3(X) = x^3 + 2x + 3$

$q_4(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,4).**
$q_4(X) = \frac{9x^3}{10} + \frac{4x^2}{5} + \frac{3x}{10} + 4$

$q_5(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,5).**
$q_5(X) = \frac{4x^3}{5} + \frac{8x^2}{5} - \frac{7x}{5} + 5$

$q_6(X)$ **is consistent with the shares of parties 1, 2, 5, and the point (0,6).**
$q_6(X) = \frac{7x^3}{10} + \frac{12^2}{5} - \frac{31x}{10} + 6$

7. **A bit of Counting (8+8+9 points).** In this problem, we will do a bit of counting related to polynomials that pass through a given set of points in the plane. We already did this counting (slightly informally) in the class. Writing the solution for this problem shall make the solution's intuition more concrete.

   We are working over the field $(\mathbb{Z}_p, +, \times)$, where p is a prime number. Let $\mathcal{P}_t$ be the set of all polynomials in the indeterminate X with degree $< t$ and coefficients in $\mathbb{Z}_p$.

   (a) Let (x_1, y_1), (x_2, y_2), ..., and (x_t, y_t) be t points in the plane $\mathbb{Z}_p^2$. We have that $x_i \neq x_j$ for all $i \neq j$, that is, the first coordinates of the points are all distinct.

   Prove that there exists a *unique polynomial* in $\mathcal{P}_t$ that passes through these t points.

   (Hint: Use Lagrange Interpolation and Schwartz–Zippel Lemma.)

I will prove this by contradiction. So, let's assume that there is NOT a unique polynomial (aka, there is more than one polynomial that passes through all of the points). That means that $\exists p(X), q(X)$ such that $p(x_i) = y_i$ and $q(x_i) = y_i$.

If we have a polynomial $r(X)$ that is the difference between $p(X)$ and $q(X)$ ($r(X) = p(X) - q(X)$), we can assume the following:

i. **$r(X) \neq 0$, since subtracting two distinct polynomials won't equal 0.**

ii. **By Lagrange, we know that $r(X) = \sum_i p(x_i) - q(x_i)$, so $r(x_i) = p(x_i) - q(x_i)$. $p(x_i) = q(x_i) = y_i$, so $r(x_i) = y_i - y_i = 0$.**

iii. **If $r(x_i) = 0$, we can conclude that x_i is a root, and since $r(x)$ is the summation of all $r(x_i)$, we can conclude that $x_1, x_2, x_3, \cdots, x_t$ are ALL roots of $r(X)$.**

iv. **The degree of $r(X)$ is guaranteed to be $< t$, since both $p(X)$ and $q(X)$ have degree $< t$.**

By the Schwartz-Zippel Lemma, a non-zero polynomail of degree t has, at most, t roots. Bullet point iv says that the degree $r(X)$ is guaranteed to be $< t$, so that means the maximum amount of roots for $r(X)$ can have is $< t$. However, bullet point iii says that we have t roots, which is not $< t$, which means that our assumption does not hold. Therefore, by contradiction, there must be a unique polynomial that passes through the points (x_1, y_1), (x_2, y_2), ..., and (x_t, y_t).

(b) Let (x_1, y_1), (x_2, y_2), ..., and (x_{t-1}, y_{t-1}) be $(t-1)$ points in the plane $\mathbb{Z}_p^2$. We have that $x_i \neq x_j$ for all $i \neq j$, that is, the first coordinates of the points are all distinct. Prove that there are p polynomials in $\mathcal{P}_t$ that pass through these $(t-1)$ points.

In the previous problem, we knew there was a unique polynomial because of the fact that we had too many roots ($\geqslant t$) to satisfy our statement. In this case, we have $< t$ roots, so there need not exist a unique polynomial. The question is, how many are there?

Again, from the previous problem, we know that, for t points, there exists a unique polynomial with t roots that passes through them. Say that, for this problem, we have a similar polynomial with t roots. We only have $t-1$ roots that need to be satisfied, so there is 1 root that can be whatever we want. Any root in $\mathbb{Z}_p$ will work! Since there are p elements in $\mathbb{Z}_p$, there are p polynomials in $\mathcal{P}_t$ that pass through these $(t-1)$ points.

(c) Let (x_1, y_1), (x_2, y_2), ..., and (x_k, y_k) be k points in the plane $\mathbb{Z}_p^2$, where $k \leqslant t$. We have that $x_i \neq x_j$ for all $i \neq j$, that is, the first coordinates of the points are all distinct. Prove that there are p^{t-k} polynomials in $\mathcal{P}_t$ that pass through these k points.

We know from part (a) that t points gives us 1 unique polynomial that passes through all t points. This follows the equation here. Since $k = t$, $p^{t-k} = p^0 = 1$ polynomial.

We also know from part (b) that $t-1$ points gives us p unique polynomials that pass through all $(t-1)$ points. This also follows the equation. Since $k = t-1$, $p^{t-k} = p^{t-(t-1)} = p^1 = p$.

Using the same logic as part (b), we know that if there are t points, there exists a unique polynomial with t roots that passes through all t points. If you only have $(t-1)$ roots to satisfy, there is 1 root that can be whatever we want, making a total of p possible polynomials. For every fewer point that we need to satisfy, we gain a power of p possibilities. This means that if we only need to satisfy $(t-2)$ points, we have $p * p = p^2$ possibilities, $(t-3)$ roots gives $p * p * p = p^3$ possibilities, $(t-x)$ roots gives p^x possibilities, and so on. This clearly proves that there are p^{t-k} polynomials in $\mathcal{P}_t$ that pass through a given k points.

Collaborators: N/A