CPSC 418 / MATH 318 — Introduction to Cryptography
ASSIGNMENT 1

Name: Harjee Johal
Student ID: 30000668

Problem 1 — Linear Feedback Shift Register Key Streams (10 marks)

(a) The first 19 bits generated by this sequence are: $z_0 = 1, z_1 = 0, z_2 = 1, z_3 = 0, z_4 = 1, z_5 = 1, z_6 = 0, z_7 = 0, z_8 = 1, z_9 = 0, z_{10} = 0, z_{11} = 0, z_{12} = 1, z_{13} = 1, z_{14} = 1, z_{15} = 1, z_{16} = 0, z_{17} = 1, z_{18} = 0$

(b) We can use linear recurrence (1) provided in the question, along with the known sequence:

$$z_i, z_{i+1}, \ldots, z_{i+2m-1}$$

to generate the following system of m equations:

$$\begin{aligned}
z_{i+m} &= c_{m-1}z_{i+m-1} + c_{m-2}z_{i+m-2} + \ldots + c_0 z_i \\
z_{i+m+1} &= c_{m-1}z_{i+m} + c_{m-2}z_{i+m-1} + \ldots + c_0 z_{i+1} \\
&\ldots \\
z_{i+2m-2} &= c_{m-1}z_{i+2m-3} + c_{m-2}z_{i+2m-4} + \ldots + c_0 z_{i+m} \\
z_{i+2m-1} &= c_{m-1}z_{i+2m-2} + c_{m-2}z_{i+2m-3} + \ldots + c_0 z_{i+m-1}
\end{aligned}$$

Since there are m unknown coefficients $c_0, c_1, \ldots, c_{m-1}$, and we have m equations utilizating these coefficients, we can generate a m x m matrix to solve for these coefficients.

(c)

Problem 2 — Password Counts (20 marks plus 5 bonus marks)

(a) The total number of passwords of length 8 that can be generated using these 94 printable characters is 94^8

(b) There are two scenarios in this question that need to be considered.

Case 1: The first four letters of the child's name are all lowercase:
If the child's name is in all lowercase, then there are $26^3 * 366 = 6,432,816$ possible password candidates. We know that the first four characters of the password are the first four letters of the child's name, and since the first one is known to be 'L', that leaves the next three slots unknown. Each of these slots could be any of the lowercase letters of the alphabet, so there are 26^3 possibilities.

We also know that the last four characters of the password are the date and month that the user was born in, in DDMM format. The number of unique DDMM combinations is equal to the number of days in the year - there is one possible DDMM combination for each day. Since we know that the user was born in 2008, that means that there are 366 unique DDMM possibilities, since 2008 was a leap year.

By multiplying the number of possibilities for the first four characters of the password by the number of possibilities for the second half of the password, we get $26^3 * 366 = 6,432,816$.

Case 2: The first four letters of the child's name are all uppercase:
The number of possibilities in this case is the same as case one: $26^3 * 366 = 6,432,816$. This scenario is almost identical to case one, with the exception that we're working with uppercase letters instead of lowercase letters. However, since there are an equal number of uppercase and lowercase letters in the alphabet, the calculations are the exact same.

Therefore, the total number of password candidates is simply the sum of the number of possibilities of each case: $6,432,816 + 6,432,816 = 12,865,632$ total candidates.

(c) The total number of passwords of length 8 that have at least one numerical digit and at least one special character can be computed by finding the number of passwords that violate this condition, and subtracting them from the total number of passwords of length 8.

First, we compute the total number of passwords of length 8: 94^8. Let us represent this value with n_t.

Next, we find the number of passwords of length 8 without a numerical digit in them. Since there are 10 numerical digits, the total number of passwords of length 8 that use none of them is: $(94 - 10)^8 = 84^8$. Let us represent this value as n_n

Next, we find the number of passwords of length 8 without a special character in them. There are 32 special characters, so the number of passwords of length 8 that don't use them is: $(94 - 32)^8 = 62^8$. Let us represent this value as n_s.

Within these two computed subsets of the total passwords of length 8, we've double-counted a specific category. Namely, we've double counted the total number of passwords of length 8 that have neither a numerical digit nor a special characer in them. Therefore, to avoid these subset being subtracted from the total number of passwords of length 8 twice, it must be added once. The total number of passwords of length 8 that don't use a numercal digit or a special character is $(94 - 32 - 10)^8 = 52^8$. Let us represent this value as $n_{n,s}$.

Combining all of this information together, we can compute the number of passwords of length 8 with at least one numerical digit and at least one special character as:

$$n_t - n_n - n_s + n_{n,s}$$

Substituting in the values that have been computed, we get $94^8 - 84^8 - 62^8 - 52^8$ passwords of length 8 that have at least one numerical digit and at least one special character.

(d) The percentage of 8-character passwords that satisfy the requirements of part (c) is simply a ratio of the value computed in the previous question, and the total number of passwords of length 8:

$$(94^8 - 84^8 - 62^8 - 52^8)/94^8 = 0.5487698\ldots \approx 54.9\%.$$

Therefore, approximately 54.9% of passwords satisfy the requirement in (c).

(e) (**Bonus**) In order to find the number of passwords of length 8 with at least one numerical digit and at least one special character and at least one uppercase letter, it is easier to find all of the passwords that violate this condition, and then subtract that from the total of passwords. We'll represent the number of passwords with at least one uppercase, at least one special character, and at least one numerical digit as n_{target}.

The total number of passwords of length 8 using 94 printable characters is 94^8. Let us represent this value as n_t.

Next, we must find the number of passwords of length 8 without any uppercase letters. There are 26 uppercase letters, so the total number of passwords of length 8 without an uppercase letter is $(94 - 26)^8 = 68^8$. We'll represent this value as n_u.

We must also find the number of passwords of length 8 without any special characters. This was computed in part (c) and found to be 62^8. We'll represent this value as n_s.

The next value to calculate is the number of passwords of length 8 without any numerical digits. This was also computed in part (c) and found to be 84^8. We'll represent this value as n_n.

Also similarly to part (c), if we subtract n_u, n_s, and n_n from n_t, we'll end up subtracting passwords that fall into two of these categories twice. Namely, passwords that fall into both n_u and n_s, passwords that fall into both n_u and n_n, and passwords that fall into both n_s and n_n will end up being subtracted twice. In order to counter this, each of these subsets must be added back. Thus, we must compute the total number of passwords that fall into each of these subsets.

The number of passwords of length 8 with neither an uppercase letter nor a special character is $(94 - 26 - 32)^8 = 34^8$. We'll represent this as $n_{u,s}$.

The number of passwords of length 8 with neither an uppercase letter nor a numerical digit is $(94 - 26 - 10)^8 = 58^8$. We'll represent this as $n_{u,n}$.

The number of passwords of length 8 with neither an special character nor a numerical digit is $(94 - 32 - 10)^8 = 52^8$. We'll represent this as $n_{s,n}$.

With this, we have now accounted for the passwords that fall into two categories. However, there is one more category of passwords that hasn't been considered yet - passwords that have neither an uppercase letter nor a special character nor a numerical digit. Let us represent this value as $n_{u,s,n}$. This subset is subtracted from the total count three times: once when n_u is subtracted from n_t, once when n_s is subtracted from n_t, and once when n_n is subtracted from n_t.
$n_{u,s,n}$ is also added back to n_t three times: once when $n_{u,s}$ is added back to n_t, once when $n_{u,n}$ is added back to n_t, and once when $n_{s,n}$ is added back to n_t. Since $n_{u,s,n}$ been subtracted three times and added three times, this subset is still present in the count of passwords with neither an uppercase letter nor a special character nor a numerical digit. Therefore, it must be subtracted one more time to be removed from this count.

The value of $n_{u,s,n}$ is $(94 - 26 - 32 - 10) = 26^8$.

Combining all of this information, we can compute n_{target} as:

$$\begin{aligned} n_{target} &= n_t - n_u - n_s - n_n + n_{u,s} + n_{u,n} + n_{s,n} - n_{u,s,n} \\ &= 94^8 - 68^8 - 62^8 - 84^8 + 34^8 + 58^8 + 52^8 - 26^8 \end{aligned}$$

Therefore, the value of n_{target} is $94^8 - 68^8 - 62^8 - 84^8 + 34^8 + 58^8 + 52^8 - 26^8$.

(f) If each permissable character in a password is chosen with equal likelihood, that means that each password in that password space is chosen with equal likelihood. Therefore, the entropy of the password space is $log_2(n)$, where n is the number of passwords in the password space.

For part (a): In part (a) we asserted that the number of passwords of length 8 was 94^8. Therefore, the entropy of this password space is:

$$\begin{aligned}
entropy &= log_2(94^8) \\
&= 52.43671\ldots \\
&\approx 52.4
\end{aligned}$$

For part (c): In part (c) we asserted that the number of passwords of length 8 with at least one numerical digit and at least one special character was $(94^8 - 84^8 - 62^8 - 52^8)$. Therefore, the entropy of this password space is:

$$\begin{aligned}
entropy &= log_2(94^8 - 84^8 - 62^8 - 52^8) \\
&= 51.5709\ldots \\
&\approx 51.6
\end{aligned}$$

(g) Assuming that each password candidate is chosen with equal likelihood, we would need 2^{128} passwords in order to achieve an entropy of 128. Since we have 94 printable characters to use in our passwords, that means we just have to determine x such that $94^x = 2^{128}$. Solving this equation, we get:

$$\begin{aligned}
94^x &= 2^{128} \\
log_2(94^x) &= log_2(2^{128}) \\
log_2(94^x) &= 128 \\
xlog_2(94) &= 128 \\
x &= 128/log_2(94) \\
x &= 19.5283\ldots \\
x &\approx 19.5.
\end{aligned}$$

However, there's no such thing as a password that's 19.5 characters long. Therefore, we must take the ceiling of this value. We cannot take the floor, because a password length of 19 would have a entropy below 128. Therefore the minimum password length that guarantees a password space with entropy 128 is 20.

Problem 3 — Probabilities of Non-Collisions (26 marks)

(a) If there are n numbers, the chance of your favourite number N being chosen is $1/n$.

(b) If there are n numbers, the chance of your favourite number N not being chosen is $(n-1)/n$

(c) If there are n numbers, and k participants, the probability of none of the participants being assigned your favourite number N is:

$$(\frac{n-1}{n})^k$$

given that the assignment of numbers are independent events.

(d) In order to find the maximal value of k when $n = 10$ such that the probability P of your favourite number N not being chosen is 50%, we use:

$$(\frac{n-1}{n})^k \geq 0.50$$

However, we know that as k increases, the value of P decreases. Therefore, we must solve for k at the boundary where $P = 50\%$:

$$\begin{aligned}
(\frac{n-1}{n})^k &= 0.50 \\
(\frac{10-1}{10})^k &= 0.50 \\
(\frac{9}{10})^k &= 0.50 \\
log_2(0.90)^k &= log_2(0.50) \\
klog_2(0.90) &= -1 \\
k &= \frac{-1}{log_2(0.90)} \\
k &= 6.5788\ldots \\
k &\approx 6.58
\end{aligned}$$

However, we cannot have such a thing as 6.58 participants. Since the value of P decreases as k increases, we must take the floor of k, 6, in order to ensure that the value of P remains above 50%. Therefore, the maximal number of participants this experiment can have to ensure at least a 50% chance that none of the pariticipants are assigned your favourite number N, given that $n = 10$ is 6.

(e) If there are k participants, and there are n numbers that are to be independently assigned to the participants, the probabilty of them all receiving different numbers can be represented as a product of multiple events:

$$P_{unique} = P_{firstpickunique} * P_{secondpickunique} * \ldots * P_{k^{th}pickunique}$$

The probability of the first participant choosing a unique number is 100%, since no numbers have been assigned yet. The probability of the second participant choosing a unique number is the probability of them not choosing the same number as the first participant. The probability of the second participant choosing the same number as the first participant is $(\frac{1}{n})$, since only one number out of n has been assigned. Therefore, the probability of them choosing a different number than the first probability is:

$$P_{secondpickunique} = (1 - \frac{1}{n})$$

The probability of the third participant choosing one of the numbers already assigned is $\frac{2}{n}$, since 2 of the n numbers have been assigned. The probability of the third pariticipant choosing a different number than the first two participants is $(1 - \frac{2}{n})$.

There is a pattern here, were the probability of the i^{th} participant choosing a unique number is $(1 - \frac{i-1}{n})$. Using this information, we can define P_{unique} as:

$$\begin{aligned}
P_{unique} &= P_{firstpickunique} * P_{secondpickunique} * \ldots * P_{k^{th}pickunique} \\
&= (1) * (1 - \frac{1}{n}) * (1 - \frac{2}{n}) * (1 - \frac{3}{n}) * \ldots * (1 - \frac{k-2}{n}) * (1 - \frac{k-1}{n}) \\
&= (\frac{n}{n}) * (\frac{n-1}{n}) * (\frac{n-2}{n}) * (\frac{n-3}{n}) * \ldots * (\frac{n-k-2}{n}) * (\frac{n-k-1}{n}) \\
P_{unique} &= \frac{n!}{(n-k)! * n^k}
\end{aligned}$$

Therefore, the probability of all k participants being assigned different numbers, when there are n numbers to chose from, is $\frac{n!}{(n-k)! * n^k}$, given that $k \leq n$.

(f) If we suppose that $n = 10$, we can find the maximal value of k such that $P_{unique} \geq 0.50$ by testing the values of k in order starting from 1, until the condition $P_{unique} \geq 0.50$ is violated.

Case $k = 1$: If $k = 1$, we can sub the values of k, n into the P_{unique} formula derived in part (e):

$$\begin{aligned}
P_{unique} &= \frac{n!}{(n-k)! * n^k} \\
&= \frac{10!}{(10-1)! * 10^1} \\
P_{unique} &= 1 \\
1 &\geq 0.5
\end{aligned}$$

Therefore, $k = 1$ is our current maximal value.

Case $k = 2$: If $k = 2$, we can sub the values of k, n into the P_{unique} formula derived in part (e):

$$\begin{aligned} P_{unique} &= \frac{n!}{(n-k)! * n^k} \\ &= \frac{10!}{(10-2)! * 10^2} \\ P_{unique} &= 0.9 \\ 0.9 &\geq 0.5 \end{aligned}$$

Therefore, $k = 2$ is our current maximal value.

Case $k = 3$: If $k = 3$, we can sub the values of k, n into the P_{unique} formula derived in part (e):

$$\begin{aligned} P_{unique} &= \frac{n!}{(n-k)! * n^k} \\ &= \frac{10!}{(10-3)! * 10^3} \\ P_{unique} &= 0.72 \\ 0.72 &\geq 0.5 \end{aligned}$$

Therefore, $k = 3$ is our current maximal value.

Case $k = 4$: If $k = 4$, we can sub the values of k, n into the P_{unique} formula derived in part (e):

$$\begin{aligned} P_{unique} &= \frac{n!}{(n-k)! * n^k} \\ &= \frac{10!}{(10-4)! * 10^4} \\ P_{unique} &= 0.504 \\ 0.504 &\geq 0.5 \end{aligned}$$

Therefore, $k = 4$ is our current maximal value.

Case $k = 5$: If $k = 5$, we can sub the values of k, n into the P_{unique} formula derived in part (e):

$$\begin{aligned} P_{unique} &= \frac{n!}{(n-k)! * n^k} \\ &= \frac{10!}{(10-5)! * 10^5} \\ P_{unique} &= 0.252 \\ 0.252 &\ngeq 0.5 \end{aligned}$$

In this case, $k = 5$ has a P_{unique} value less than 0.5. As can be seen from the past 5 calculations, the value of P_{unique} decreases as k approaches n. Therefore, all subsequent values of k after $k = 5$ will also have P_{unique} values below 0.5. Thus, the maximal value of k such that $P_{unique} \geq 0.50$ when $n = 10$, is $k = 4$.

(g) In part (e), we defined the probability P of k participants being assigned a different number from a list of numbers n as:

$$P = (1 - \frac{1}{n}) * (1 - \frac{2}{n}) * (1 - \frac{3}{n}) * \ldots * (1 - \frac{k-2}{n}) * (1 - \frac{k-1}{n})$$

Given that k is very large, and that k is very small compared to n, we can surmise that n is very large. This means that the fractions

$$P = \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{k-2}{n}, \frac{k-1}{n}$$

will all be very small positive values, since n is defined as a positive integer. Therefore, we can apply the approximation:

$$e^{-x} \approx 1 - x$$

which applies when $x > 0$ is very small, to

$$P = (1 - \frac{1}{n}) * (1 - \frac{2}{n}) * (1 - \frac{3}{n}) * \ldots * (1 - \frac{k-2}{n}) * (1 - \frac{k-1}{n})$$

to obtain

$$P \approx (e^{-\frac{1}{n}}) * (e^{-\frac{2}{n}}) * (e^{-\frac{3}{n}}) * \ldots * (e^{-\frac{k-2}{n}}) * (e^{-\frac{k-1}{n}}).$$

Using power rules, we obtain:

$$\begin{aligned} P &\approx e^{-(1+2+3+\ldots+(k-2)+(k-1))/n} \\ &= e^{-\frac{k(k-1)}{n}}. \end{aligned}$$

Since k is very large, we can use the approximation $k(k-1) \approx k^2$. Applying this approximation, we get the result

$$P \approx e^{-k^2/n}.$$

(h) The result of part (g) is that the probability P of all k participants being assigned different numbers from a list of n numbers can be approximated as $P \approx e^{\frac{-k^2}{2n}}$. We are asked to prove that the number of participants k needed to ensure a roughly 50%

chance that all participants are assigned different numbers is approximately $1.177\sqrt{n}$. Substituing $P = 0.50$, we can solve the approximately of part (f) to find k as a function of n:

$$\begin{aligned}
P &= e^{\frac{-k^2}{2n}} \\
0.50 &= e^{\frac{-k^2}{2n}} \\
ln(0.50) &= ln(e^{\frac{-k^2}{2n}}) \\
ln(0.50) &= \frac{-k^2}{2n} \\
-ln(0.50) * 2n &= k^2 \\
k &= \sqrt{-ln(0.50) * 2n} \\
k &= \sqrt{-ln(0.50) * 2} * \sqrt{n} \\
k &= (1.1774\ldots)\sqrt{n} \\
k &\approx 1.177\sqrt{n}
\end{aligned}$$

From this, it can be seen that by using the approximation from part (g) we can represent the number of participants k needed to ensure a rougly 50% chance of all participants being assigned different numbers as a function of the amount of assignable numbers n: $k \approx 1.177\sqrt{n}$.

Problem 4 — Equiprobability maximizes entropy for two outcomes, 10 marks

(a) If $p(X_1) = p = \frac{1}{8}$, and $p(X_2) = (1 - p) = \frac{7}{8}$, we can plug these values into the entropy equation to obtain a value for $H(X)$:

$$\begin{aligned} H(X) &= -p * log_2(p) - (1 - p) * log_2(1 - p) \\ &= -\frac{1}{8} * log_2(\frac{1}{8}) - \frac{7}{8} * log_2(\frac{7}{8}) \\ &= 0.5435\ldots \\ H(X) &\approx 0.54. \end{aligned}$$

(b) In order to find the maximal value of $H(X)$, you must find the critical points of the function. That is, you must find the value of X when the derivative of $H(X)$ is 0, $H'(X) = 0$.

First, we must derive $H'(X)$:

$$\begin{aligned} H'(X) &= (-p * log_2(p) - (1 - p) * log_2(1 - p))' \\ &= -(1 * log_2(p) + \frac{p}{p * ln(2)}) - (-1 * log_2(1 - p) + \frac{(1 - p)}{(1 - p) * ln(2) * -1}) \\ &= -(log_2(p) + \frac{1}{ln(2)}) - (-log_2(1 - p) - \frac{1}{ln(2)}) \\ &= -log_2(p) - \frac{1}{ln(2)} + log_2(1 - p) + \frac{1}{ln(2)} \\ &= log_2(1 - p) - log_2(p) \\ H'(X) &= log_2(\frac{1 - p}{p}). \end{aligned}$$

From this, we see that $H'(X) = log_2(\frac{1-p}{p})$. Since both p and $1 - p$ are defined as being positive, we don't need to worry about this derivative being undefined. Next, we set $H'(X) = 0$ and solve for p:

$$\begin{aligned} H'(X) &= 0 \\ 0 &= log_2(\frac{1 - p}{p}) \\ 2^0 &= \frac{1 - p}{p} \\ p &= 1 - p \\ 2p &= 1 \\ p &= \frac{1}{2}. \end{aligned}$$

From this we identify $p = \frac{1}{2}$ as a critical point of $H(X)$. However, in order to determine whether $H(X)$ has a maximum value when $p = \frac{1}{2}$, we must perform a second derivative test. If the value of the second derivative is > 0 at the critical point, the function has a minimum value at that point. Else, if the second derivative is < 0 at the critical point, then that critical point is a maximum of the original function.

First, we must find the second deriviative of $H(X)$, denoted by $H''(X)$. We know that $H'(X) = log_2(1-p) - log_2(p)$. Therefore, $H''(X)$ is:

$$\begin{aligned} H''(X) &= (log_2(1-p) - log_2(p))' \\ &= \frac{-1}{(1-p) * ln(2)} - \frac{1}{p * ln(2)} \end{aligned}$$

Next, we substitute in $p = \frac{1}{2}$ into $H''(X)$ in order to perform the second derivative test:

$$\begin{aligned} H''(X) &= \frac{-1}{(1-p) * ln(2)} - \frac{1}{p * ln(2)} \\ H''(X) &= \frac{-1}{(1-\frac{1}{2}) * ln(2)} - \frac{1}{\frac{1}{2} * ln(2)} \\ H''(X) &= -2.7725\ldots \\ H''(X) &< 0. \end{aligned}$$

Since $H''(X)$ is less than 0 when $p = \frac{1}{2}$, and $H'(X) = 0$ when $p = \frac{1}{2}$, that means that $H(X)$ has a maxiumum value when $p = \frac{1}{2}$.

When $p = \frac{1}{2}$, the value of $p(X_1) = p = \frac{1}{2}$, and the value of $p(X_2) = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$. Therefore, when H(X) has a maximum value, $p(X_1) = p(X_2)$.

(c) $H(X)$ has a maximal value when $p = \frac{1}{2}$. If we substitute $p = \frac{1}{2}$ into $H(X)$, we find that the maximal value is:

$$\begin{aligned} H(X) &= -p * log_2(p) - (1-p) * log_2(1-p) \\ &= -(\frac{1}{2}) * log_2(\frac{1}{2}) - (1 - \frac{1}{2}) * log_2(1 - \frac{1}{2}) \\ &= -(\frac{1}{2}) * (-1) - (\frac{1}{2}) * log_2(\frac{1}{2}) \\ &= (\frac{1}{2}) - (\frac{1}{2}) * (-1) \\ H(X) &= 1 \end{aligned}$$

Thus, the maximal value of $H(X)$ is 1.

Problem 5 — Cryptanalysis of a class of linear ciphers, 34 marks)

***** Remove the text for this problem if you don't attempt it. *****

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Problem 7 — Mixed Vigenère cipher cryptanalysis, 10 marks

***** Remove the text for this problem if you don't attempt it. *****