CPSC 418 / MATH 318 — Introduction to Cryptography
ASSIGNMENT 1 Problems 1-4

Name: Karmvir Singh Dhaliwal
Student ID: 30025474

Problem 1 — Password length and entropy

(a) There are 128^8 total *ASCII* encodings of 8 character strings. This is because for each character, we have 2^7, or 128 possible options for characters, and 8 character spaces total.

(b) i. There are 94^8 total passwords of length 8 consisting of any printable character. This is because there are 94 total printable characters, and 8 characters in the password, so we have 94 choices for each spot.

ii. There are 26^8 total passwords of length 8 consisting of only lowercase letters. This is because there are 26 total lowercase characters, and 8 total characters in the password, so we have 26 choices for each spot.

(c) i. The approximate percentage of 8 character *ASCII* encodings that consist of any printable character is

$$\frac{94^8}{128^8} \approx 8.46\%$$

as we have 94^8 total passwords consisting of printable characters, and 128^8 total *ASCII* encodings.

ii. The approximate percentage of 8 character *ASCII* encodings that consist of any lowercase character is

$$\frac{26^8}{128^8} \approx 0.0002898\%$$

as we have 26^8 total passwords consisting of lowercase characters and 128^8 total *ASCII* encodings.

(d) i. The entropy of the space of passwords consisting of 8 printable characters is

$$\begin{aligned} H(x) &= \log_2(\frac{1}{P(x)}) \\ &= \log_2(\frac{1}{94^8}) \\ &= \log_2(94^8) \\ &\approx 52.44 \end{aligned} \tag{1}$$

We can use the equation $H(x) = \log_2(\frac{1}{P(x)})$ because the passwords are equally likely, and therefore the entropy is maximal.

ii. The entropy of the space of passwords consisting of 8 lowercase characters is

$$\begin{aligned} H(x) &= \log_2(\frac{1}{P(x)}) \\ &= \log_2(\frac{1}{26^8}) \\ &= \log_2(26^8) \\ &\approx 37.60 \end{aligned} \tag{2}$$

We can use the equation $H(x) = \log_2(\frac{1}{P(x)})$ because the passwords are equally likely, and therefore the entropy is maximal.

(e) i. The minimum password length of passwords consisting of any printable character to achieve an entropy of 128 is

$$\begin{aligned} 128 &= \log_2(\frac{1}{P(x)}) \\ 128 &= \log_2(\frac{1}{94^8}) \\ 128 &= \log_2(94^x) \\ 128 &= x\log_2(94) \\ x &= \frac{128}{log_2(94)} \\ x &\approx 19.53 \end{aligned} \tag{3}$$

Therefore, you would need a password length of approximately 20 characters to have an entropy of 128, given all passwords occur equally likely.

ii. The minimum password length of passwords consisting of any lowercase character to achieve an entropy of 128 is

$$\begin{aligned} 128 &= \log_2(\frac{1}{P(x)}) \\ 128 &= \log_2(\frac{1}{26^8}) \\ 128 &= \log_2(26^x) \\ 128 &= x\log_2(26) \\ x &= \frac{128}{log_2(26)} \\ x &\approx 27.23 \end{aligned} \tag{4}$$

Therefore, you would need a password length of approximately 28 characters to have an entropy of 128, given all passwords occur equally likely.

Problem 2 — One-time pad without the all-zeros key

(a) Let our message space be $\{0,1\}^2$ that is, all possible strings of length 2 containing only 0s and 1s. Suppose all messages are chosen with equal likelihood. Then, our $P(M) = \frac{1}{4}$, as we have four possible combinations of 0s and 1s; namely 00, 01, 10, 11. We choose our message to be 01. We choose our cipher-text to be 11. Since we know that our key cannot be 00, we know that our cipher-text can not be the message itself. This removes 1 possible option for the message. This means that our $P(M|C)$ is $= \frac{1}{3}$, as there are only 3 possible options for our message knowing that the cipher-text is 11, namely 00, 01, 10. We have that $P(M) \neq P(M|C)$, thereby proving using the definition of perfect secrecy that the one-time pad without the 0 key no longer has perfect secrecy. □

(b) Since we know the one-time pad is perfectly secret, we know that $P(M) = P(M|C)$ always holds true. This means that knowing the cipher-text gives us no information about the original message itself. With this, given a cipher-text that has been encrypted using the 0 key, if an eavesdropper doesn't know what the key is they have no way of knowing that the cipher-text is the message; it is impossible to figure out the cipher-text is the message without figuring out the key, so the eavesdropper's best bet is simply guessing.

Problem 3 — Weak collisions

(a) Since the numbers are randomly assigned, the chance that someone is assigned my favourite number N is $\frac{1}{n}$.

(b) The probability that a participant is not assigned my favourite number N is simply the total probability without the probability of being assigned my number N, or $1 - \frac{1}{n}$.

(c) The probability that none of the K participants are assigned my favourite number N is $(1 - \frac{1}{n})^k$. Each participants probability of not getting N is $1 - \frac{1}{n}$, so to calculate the total probability that no one gets N we multiply $1 - \frac{1}{n}$ by itself, k times.

(d) The threshold K in this case is:

$$\begin{aligned} n &= 10 \\ \frac{1}{2} &= (1 - \frac{1}{n})^k \\ \frac{1}{2} &= (1 - \frac{1}{10})^k \\ \frac{1}{2} &= (\frac{9}{10})^k \\ k &= \log_{\frac{9}{10}}(\frac{1}{2}) \\ k &\approx 6.58 \end{aligned} \tag{5}$$

So, the minimum number of participants needed to have a 50 percent chance of a weak collision is 7. Essentially, we find the lowest integer k where the probability that no one has my favourite number N is less than or equal to 50% in the equation $(1 - \frac{1}{n})^k$.

(e) Suppose that the number of participants k $> \log_e(2)n$. Now:

$$\begin{aligned} &1 - (1 - \frac{1}{n})^k \quad \text{-From our probability equation} \\ &> 1 - (e^{\frac{-1}{n}})^k \quad \text{-From given Taylor inequality} \\ &= 1 - (e^{\frac{-k}{n}}) \quad \text{-From power rules} \\ &> 1 - (e^{\frac{-log_e(2)n}{n}}) \quad \text{-From our assumption} \\ &= 1 - (e^{-log_e(2)}) \\ &= \frac{1}{2} \end{aligned} \tag{6}$$

As required, thereby proving that if the number of participants is above $log_e(2)n \approx 0.69n$, then there is at least a 50% chance of a weak collision. □

Problem 4 — (Strong) collisions

(a) The probability that no collisions occur is

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

Essentially, when we have 1 participant the probability of no collisions is 1, then for each extra person, there is one less possible number that they can select, as the number has already been selected by someone else. We multiply the probability for each person together to get our total probability.

(b) The probability that there is a collision is

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

That is, total probability minus the probability of no collisions.

(c) Using the above equation:

$$\begin{aligned}
1-\prod_{i=1}^{k-1}(1-(\frac{i}{n})) &\geq \frac{1}{2} \\
1-\prod_{i=1}^{k-1}(1-(\frac{i}{10})) &\geq \frac{1}{2} \\
\prod_{i=1}^{k-1}(1-(\frac{i}{10})) &\leq \frac{1}{2} \\
(1-(\frac{0}{10})(1-(\frac{1}{10})(1-(\frac{2}{10})(1-(\frac{3}{10})(1-(\frac{4}{10}) &\leq \frac{1}{2} \\
(1)(\frac{9}{10})(\frac{8}{10})(\frac{7}{10})(\frac{6}{10}) &\leq \frac{1}{2} \\
\frac{3024}{10000} &\leq \frac{1}{2} \\
k &= 5
\end{aligned} \tag{7}$$

So, the threshold K in this case is 5.

(d) We want to prove that P $\leq e^{-\frac{k(k-1)}{2n}}$. Now:

$$\begin{aligned}
P &= \prod_{i=1}^{k-1} 1-\frac{i}{n} \\
&< \prod_{i=1}^{k-1} e^{-\frac{i}{n}} \quad \text{-From our Taylor inequality} \\
&= e^{-\sum_{i=1}^{k-1}\frac{i}{n}} \quad \text{-From the power rules} \\
&= e^{-\frac{\sum_{i=1}^{k-1} i}{n}} \\
&\leq e^{-\frac{k(k-1)}{2n}}
\end{aligned} \tag{8}$$

Since $i < k$, as required. Thereby proving that P $\leq e^{-\frac{k(k-1)}{2n}}$.□

(e) Suppose that the number of participants k $> \sqrt{\log_e(4)n}$. Now:

$$
\begin{aligned}
& e^{\frac{-k^2}{2n}} \quad \text{-From our probability equation} \\
\log_e(x) &= \frac{-k^2}{2n} \\
\log_e(x) &> \frac{-\sqrt{\log_e(4)n}^2}{2n} \quad \text{-From our assumption} \\
\log_e(x) &> \frac{-\log_e(4)n}{2n} \\
\log_e(x) &> \frac{-\log_e(2^2)n}{2n} \\
\log_e(x) &> \frac{-\log_e(2)2n}{2n} \\
\log_e(x) &> -\log_e(2) \\
\log_e(x) &> \log_e(\frac{1}{2}) \\
x &> \frac{1}{2}
\end{aligned}
\tag{9}
$$

As required, thereby proving that if the number of participants is above $\sqrt{\log_e(4)n} \approx 1.177\sqrt{n}$, then there is at least a 50% chance of a strong collision. □