MAS 433 Tutorial 1

Wang Xueou (087199E16)

August 31, 2011

Question 1 Solution:
1.1. In the modified one-time pad C_i is either 0, 1, or 2.
If $C_i = 0$, we know both the plaintext and the key are 0, i.e., $P_i = 0, K_i = 0$.
If $C_i = 2$, then both the plaintext and the key are 0, i.e., $P_i = 1, K_i = 1$.
If $C_i = 1$, we have either $P_i = 0, K_i = 1$, or $P_i = 1, K_i = 0$. As there are only 2 possibilities, we just try each case and check in which case the plaintext makes sense.
1.2. A cryptosystem has perfect secrecy if knowing ciphertext reveals no information about the plaintext. i.e., $\mathbf{Pr}\left[\mathbf{P} = p|\mathbf{C} = c\right] = \mathbf{Pr}\left[\mathbf{P} = p\right]$. Now, suppose $\mathbf{P} = \mathbf{K} = \{0,1\}^n$. After encrypting using the modified one-time pad, the cipher $\mathbf{C}$ has m 1's and the rest 0's ($0 < m < n$ because $m = 0$ or n is trivial). Then we have:
$\mathbf{Pr}\left[\mathbf{P} = p\right] = \frac{1}{2^n}$
while $\mathbf{Pr}\left[\mathbf{P} = p|\mathbf{C} = c\right] = \frac{1}{2^m} > \mathbf{Pr}\left[\mathbf{P} = p\right]$
Thus, the modified one-time pad is not perfectly secure.

Question 2 Solution:
2.1. Entropy $= -2^{995} \times \left(\frac{1}{2^{995}} \times \log_2 \frac{1}{2^{995}}\right) = 995$
The risk of using the key is that the first five bits will be known to the attacker. If there are sensitive information in the five bits, it would be very dangerous.
2.2. Entropy $= -1000 \times (0.54 \times \log_2 0.54 + 0.46 \times \log_2 0.46) = 995.378439$
The risk of using the key is that the key will have around 54% of zeros. Suppose the computer is subjected to unlimited computing ability, then all the keys with 540 zeros can be tried to break the cipher.

Question 3 Solution:
3.1. Entropy $= 128 \times (-0.5 \times \log_2 0.5 - 0.5 \times \log_2 0.5) = 128$
On average, the guesses needed to guess the value correctly is:

$$\begin{aligned} & \textstyle\sum_{i=0}^{2^{128}-1} \frac{2^{128}-i}{2^{128}} \times \frac{1}{2^{128}-i} \times (i+1) \\ = & \textstyle\sum_{i=0}^{2^{128}-1} \frac{i+1}{2^{128}} \\ = & \frac{1}{2^{128}} \times \frac{\left(1+2^{128}\right)\times 2^{128}}{2} \\ = & 2^{127} + 0.5 \end{aligned}$$

3.2.
3.2.1. Entropy $= -\frac{1}{2} \times \log_2 \frac{1}{2} - 2^{127} \times \left(\frac{\frac{1}{2}}{2^{127}} \log_2 \frac{\frac{1}{2}}{2^{127}}\right) = 127\frac{1}{2}$
3.2.2. Since "$010101 \cdots 0101''$ appears half of the time, we have half of all the keys being this key. Thus we only need to guess the which half of the keys is this key. We have

$$\left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{n}{2} + \sum_{i=1}^{\frac{n}{2}-1} \left(\begin{array}{c} \frac{n}{2} \\ i \end{array}\right) \left(\frac{1}{2}\right)^{i} \left(\frac{1}{2}\right)^{\frac{n}{2}-i} \left(\frac{n}{2} + i\right)$$

3.2.3. The average number of guesses is

$$\begin{array}{rl} & \frac{1}{2} \times 1 + \sum_{i=0}^{2^{127}-1} \frac{2^{127}-i}{2^{127}} \times \frac{1}{2^{127}-i} \times (i+1) \\ = & 0.5 + \frac{1}{2^{127}} \sum_{i=0}^{2^{127}-1} \frac{i+2}{2^{127}} \\ = & 0.5 + \frac{1}{2^{127}} \times \frac{\left(2+2^{127}+1\right) \times 2^{127}}{2} \\ = & 0.5 + \left(1.5 + 2^{126}\right) \\ = & 2^{126} + 2 \end{array}$$

Question 4 Solution:
4.1. $\mathbf{Pr}\left[C = c_i\right] = \Sigma_{j,s} \mathbf{Pr}\left[P = p_j\right] \cdot \mathbf{Pr}\left[K = k_s\right], \text{where} E_{k_s}\left(p_s\right) = c_i$. Thus,
$\mathbf{Pr}\left[C = c_1\right] = \mathbf{Pr}\left[P = p_1\right] \cdot \mathbf{Pr}\left[K = k_1\right] = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$
$\mathbf{Pr}\left[C = c_2\right] = \mathbf{Pr}\left[P = p_1\right] \cdot \mathbf{Pr}\left[K = k_2\right] + \mathbf{Pr}\left[P = p_2\right] \cdot \mathbf{Pr}\left[K = k_1\right] = \frac{1}{4} \times \frac{1}{4} + \frac{3}{4} \times \frac{1}{2} = \frac{7}{16}$
$\mathbf{Pr}\left[C = c_3\right] = \mathbf{Pr}\left[P = p_1\right] \cdot \mathbf{Pr}\left[K = k_3\right] + \mathbf{Pr}\left[P = p_2\right] \cdot \mathbf{Pr}\left[K = k_2\right] = \frac{1}{4} \times \frac{1}{4} + \frac{3}{4} \times \frac{1}{4} = \frac{1}{4}$
$\mathbf{Pr}\left[C = c_4\right] = \mathbf{Pr}\left[P = p_2\right] \cdot \mathbf{Pr}\left[K = k_3\right] = \frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$

4.2.
Entropy of $\mathbf{P} = -\frac{1}{4} \times \log_2 \frac{1}{4} - \frac{3}{4} \times \log_2 \frac{3}{4} = 0.811278124$
Entropy of $\mathbf{K} = -\frac{1}{2} \times \log_2 \frac{1}{2} - \frac{1}{4} \times \log_2 \frac{1}{4} - \frac{1}{4} \times \frac{1}{4} = 1.5$
Entropy of $\mathbf{C} = -\frac{1}{8} \times \log_2 \frac{1}{8} - \frac{7}{16} \times \log_2 \frac{7}{16} - \frac{1}{4} \times \log_2 \frac{1}{4} - \frac{3}{16} \times \log_2 \frac{3}{16} = 1.84960175$

4.3.
$\mathbf{Pr}\left(p_1|c_1\right) = \frac{\mathbf{Pr}(P=p_1,C=c_1)}{P(C=c_1)} = \frac{\mathbf{Pr}(P=p_1)\mathbf{Pr}(K=k_1)}{P(C=c_1)} = \frac{\frac{1}{4} \times \frac{1}{2}}{\frac{1}{8}} = 1$
$\mathbf{Pr}\left(p_1|c_2\right) = \frac{\mathbf{Pr}(P=p_1,C=c_2)}{P(C=c_1)} = \frac{\mathbf{Pr}(P=p_1)\mathbf{Pr}(K=k_2)}{P(C=c_2)} = \frac{\frac{1}{4} \times \frac{1}{4}}{\frac{7}{16}} = \frac{1}{7}$
$\mathbf{Pr}\left(p_1|c_3\right) = \frac{\mathbf{Pr}(P=p_1,C=c_3)}{P(C=c_3)} = \frac{\mathbf{Pr}(P=p_1)\mathbf{Pr}(K=k_3)}{P(C=c_3)} = \frac{\frac{1}{4} \times \frac{1}{4}}{\frac{1}{4}} = \frac{1}{4}$
$\mathbf{Pr}\left(p_1|c_4\right) = \frac{\mathbf{Pr}(P=p_1,C=c_4)}{P(C=c_4)} = 0$
$\mathbf{Pr}\left(p_2|c_1\right) = 1 - \mathbf{Pr}\left(p_1|c_1\right) = 0$
$\mathbf{Pr}\left(p_2|c_2\right) = 1 - \mathbf{Pr}\left(p_1|c_2\right) = \frac{6}{7}$
$\mathbf{Pr}\left(p_2|c_3\right) = 1 - \mathbf{Pr}\left(p_1|c_3\right) = \frac{3}{4}$
$\mathbf{Pr}\left(p_2|c_4\right) = 1 - \mathbf{Pr}\left(p_1|c_4\right) = 1$

4.4.

$$\begin{array}{rcl} \text{Given } c_1 : \text{Entropy of } \mathbf{P} & = & -\mathbf{Pr}\left(p_1|c_1\right) \log_2 \mathbf{Pr}\left(p_1|c_1\right) - \mathbf{Pr}\left(p_2|c_1\right) \log_2 \mathbf{Pr}\left(p_2|c_1\right) \\ & = & 0 + 0 \\ & = & 0 \\ \text{Given } c_2 : \text{Entropy of } \mathbf{P} & = & -\mathbf{Pr}\left(p_1|c_2\right) \log_2 \mathbf{Pr}\left(p_1|c_2\right) - \mathbf{Pr}\left(p_2|c_2\right) \log_2 \mathbf{Pr}\left(p_2|c_2\right) \\ & = & -\frac{1}{7} \times \log_2 \frac{1}{7} - \frac{6}{7} \times \log_2 \frac{6}{7} \\ & = & 0.591672779 \\ \text{Given } c_3 : \text{Entropy of } \mathbf{P} & = & -\mathbf{Pr}\left(p_1|c_3\right) \log_2 \mathbf{Pr}\left(p_1|c_3\right) - \mathbf{Pr}\left(p_2|c_3\right) \log_2 \mathbf{Pr}\left(p_2|c_3\right) \\ & = & -\frac{1}{4} \times \log_2 \frac{1}{4} - \frac{3}{4} \times \log_2 \frac{3}{4} \\ & = & 0.811278124 \\ \text{Given } c_4 : \text{Entropy of } \mathbf{P} & = & -\mathbf{Pr}\left(p_1|c_4\right) \log_2 \mathbf{Pr}\left(p_1|c_4\right) - \mathbf{Pr}\left(p_2|c_4\right) \log_2 \mathbf{Pr}\left(p_2|c_4\right) \\ & = & 0 + 0 \\ & = & 0 \end{array}$$

These results are different from the entropy of $\mathbf{P}$ because the keys are not randomly generated and thus they are not perfectly secure, i.e. $\mathbf{Pr}\left[P = p_i | C = c_j\right] \neq \mathbf{Pr}\left[P = p_i\right]$

Question 5 Solution:
For plaintext spaces consisting of 26^m $m-$ grams, we have $|\mathbf{K}| = 26^m$, i.e. $n = 26^m$

$$\begin{array}{rcl}\log_2 n & \approx & \log_2 \left(\sqrt{2\pi n}\right)\left(\frac{n}{e}\right)^n) \\ & = & \log_2 \sqrt{2\pi} + 0.5 \log_2 n + n \log_2 \left(\frac{n}{e}\right) \\ & = & (0.5+n)\log_2 n - n \log_2 e + \log_2 \sqrt{2\pi} \\ & = & (0.5+n)\log_2 -1.44n + 1.33\end{array}$$

The unicity distance is defined by

$$n_0 \approx \frac{\log_2 |K|}{R_L \log_2 |P|}$$

Now $n = 26^m$,

$$\begin{array}{rcl} n_0 & \approx & \frac{(0.5+26^m)\log_2 26^m 1.44\times 26^m+1.33}{R_L \log_2 |P|} \\ & = & \frac{m(0.5+26^m)\log_2 26-1.44\times 26^m+1.33}{0.75\times \log_2 26} \\ & = & \frac{m(0.5+26^m)\log_2 26-1.44\times 26^m+1.33}{0.75\times 4.7}\end{array}$$

If $m=1, n_0 = \frac{(0.5+26)\log_2 26-1.44\times 26+1.33}{0.75\times 4.7} = 25.0926674$

If $m=2, n_0 = \frac{2\times\left(0.5+26^2\right)\log_2 26-1.44\times 26^2+1.33}{0.75\times 4.7} = 1528.39289$

If $m=3, n_0 = \frac{3\times\left(0.5+26^3\right)\log_2 26-1.44\times 26^3+1.33}{0.75\times 4.7} = 63132.9719$

If $m=4, n_0 = \frac{4\times\left(0.5+26^4\right)\log_2 26-1.44\times 26^4+1.33}{0.75\times 4.7} = 2250756.84$

If $m=5, n_0 = \frac{5\times\left(0.5+26^5\right)\log_2 26^5-1.44\times 26^5+1.33}{0.75\times 4.7} = 74362919.1$

Question 6 Solution:

$\mathbf{P} = \mathbf{C} = \mathbf{K} = \{0,1,2,\cdots,25\}$, and key is chosen randomly. We need to prove $\mathbf{Pr}[P=p|C=c] = \mathbf{Pr}[P=p]$

Proof. $\mathbf{Pr}[C=c|P=p] = \mathbf{Pr}[K = c\ominus p] = \frac{1}{26}$

$$\begin{array}{rcl}\mathbf{Pr}[C=c] & = & \sum_{p\in\mathbf{P}}\left(\mathbf{Pr}[P=p]\,\mathbf{Pr}[C=c|P=p]\right) \\ & = & \sum_{p\in\mathbf{P}}\left(\mathbf{Pr}[P=p]\cdot\frac{1}{26}\right) \\ & = & \frac{1}{26}\end{array}$$

$$\begin{array}{rcl}\mathbf{Pr}[P=p|C=c] & = & \frac{\mathbf{Pr}[P=p]\cdot\mathbf{Pr}[C=c|P=p]}{\mathbf{Pr}[C=c]} \\ & = & \frac{\mathbf{Pr}[P=p]\cdot\frac{1}{26}}{\frac{1}{26}} \\ & = & \mathbf{Pr}[P=p]\end{array}$$

Question 7 Solution:

7.1.

Step1. Suppose we guess the key length is m. We pick out the $1st, (1+m)th, (1+2m)th, \cdots$ to constitute a new cipher C^{new}. Then C^{new} is just a shift cipher.

Step2. For $k = 0,1,2,\cdots,25$, we shift C^{new} using k to get the decrypted message D^{new}.

Step3. Then we calculate the frequency of letter $A, B, C, \cdots, Z$ in D^{new}.

Step4. Let $X = (x_0, x_1, \cdots, x_{25}), Y = (y_0, y_1, \cdots, y_{25})$, where x_i, y_i are the frequency of letter i in D^{new} and Englishi language respectively. We calculate the correlation between them as:

$$Corr_{xy} = \frac{26\sum x_i y_i - \sum x_i y_i}{\sqrt{26\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{26\sum y_i^2 - \left(\sum y_i\right)^2}}$$

Then the smallest $Corr_{xy}$ indicates that our key length and the shift key are both correct.

7.2. Suppose we want to try the key length of m. Then we pick up the $i-th\,(1 \leq im)$ letter every m letters in the original cipher and thus get m new ciphers $C_1, C_2, \cdots, C_m$. For each new cipher C_i, we calculate the frequency of every cipher letter and get their entropy by

$$H\left(C_i\right) = -\sum_{x \in C_i} \mathbf{Pr}\left[x\right] \cdot \log_2 \mathbf{Pr}\left[x\right]$$

If all the m entropies are around 4.19, then our length is correct. □