The University of Melbourne
School of Computing and Information Systems
COMP90043 Cryptography and Security

Assignment 1, Semester 2 2020
Hints and Solutions

Questions

1. Classical Ciphers [20 marks]

   Consider the following version of a classical cipher where plaintext and ciphertext elements are the integers from 0 to 35. Note that this alphabet may be used when plaintexts are 26 English characters and 10 numeric characters. The encryption function, which maps any plaintext p to a ciphertext c, is given by

   $$c = E_{(a,b)}(p) = (ap + b) \bmod 36,$$

   where a and b are integers less than 36.

   (a) What is the decryption function for the scheme?

   Solution:
   $p = D_{(a,b)}(c) = a^{-1}(c - b) \bmod 36$, where a^{-1} is the multiplicative inverse of a mod 36.

   (b) A key is called trivial if $c = p$ for all input p. How many non-trivial keys are possible for this scheme?

   Solution:
   b could be any integer between 0 and 35 (both inclusive).
   a has to be relatively prime to 36 (otherwise $a^{-1}mod36$ does not exist).
   i.e. $a \in \{1, 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35\}$ There are in total $36 \times 12 = 432$ possible keys. Out of which $\{a = 1, b = 0\}$ is a trivial key as

   $$c = (ap + b) \bmod 36 = (1 \times p + 0) \bmod 36 = p.$$

   Thus the number of non-trivial key is 431.

   (c) Would this cipher be considered as mono-alphabetic cipher or poly-alphabetic cipher? Why?

   Solution:
   It is a mono-alphabetic cipher as the encryption is done by mapping each letter in the plaintext into a fixed letter in the ciphertext.

   (d) You are given a large amount of ciphertext characters encrypted using this scheme. Assuming its plaintext was written in English, show how an attacker can retrieve the key.

Solution:
Frequency analysis could be applied to retrieve the mapping between plaintext and ciphertext.
If the encoding of each character is known (e.g. "a" $\leftrightarrow 0$, "b" $\leftrightarrow 1, ...,$ "z" $\leftrightarrow 25$, "0" $\leftrightarrow 26, ...,$ "9" $\leftrightarrow 35$), a bruteforce approach could also be used considering the number of possible keys is relatively small. One can try all possible keys and see which decrypted plaintext is readable.

(e) An oracle is available to you which can output the encrypted ciphertext for arbitrary plaintext you give. Briefly describe an efficient way to retrieve the key using the oracle.

Solution:
We can use two plaintexts p_1 and p_2, to get their corresponding ciphertexts c_1 and c_2. We can then get $(c_1 - c_2) \bmod 36 = a(p_1 - p_2) \bmod 36$. In order to get a unique solution of a, $(p_1 - p_2)$ must have a multiplicative inverse under modulo 36.
As an example, we can feed the oracle with $p_1 = 0$, the encrypted result should be $c_1 = (ap_1 + b) \bmod 36 = (a \times 0 + b) \bmod 36 = b$. Then for $p_2 = 1$, we will get $c_2 = (ap_2 + b) \bmod 36 = (a \times 1 + b) \bmod 36 = (a + b) \bmod 36$. The value of a could be retrieved by $a = c_2 - b \bmod 36$.

2. General Security [8 marks]

   Which of the following factors might be the most concern by the public in regards to using the COVIDSafe app[1]? Justify your answer in a few sentences.

   (a) Confidentiality (b) Integrity (c) Availability

3. Euclid's algorithm [15 marks]

   Perform the following implementation tasks in a language of your choice. You are at free to employ any underlying integer arithmetic library. In order to get full marks, your algorithm has to be able to work in realistic cryptographic environments (consider 10^{1000} as input).

   (a) Implement the extended GCD algorithm as discussed in lectures and print the code here.

[1] `https://www.health.gov.au/resources/apps-and-tools/covidsafe-app`

Solution:
The following function takes in two integers a and b, calculates their GCD and coefficients (x, y) such that $ax + by = gcd(a, b)$.

```
def XGCD(a, b):
    x = temp_y = 0
    y = temp_x = 1
    while b:
        temp_x, x = x, temp_x - a // b * x
        temp_y, y = y, temp_y - a // b * y
        a, b = b, a % b
    return a, temp_x, temp_y # return as (gcd, x, y)
```

(b) Implement a function which takes two positive integers a, n as inputs, and returns the inverse of $(a \mod n)$ based on your extended GCD algorithm (that you just implemented above). Print the code for this function.

Solution:
The following function takes in two integers a and n, finds and returns the inverse of a under modulo n. In case the inverse does not exist, the function will produce no return.

```
def inverse(a, n):
    gcd, x, y = XGCD(a, n)
    if gcd == 1:
        return x % n
```

(c) Use the above function to find the inverse of $(X \mod 16777259)$, where X is your student number. You don't need to show steps for the calculation.

Solution:
For student number 999999, the output is 16402203.

4. Poly-alphabetic Cipher [21 marks]

For this question, we consider the Hill cipher given in the textbook on an alphabet $\mathcal{A}$ consisting of 26 English characters (A-Z), 10 numeric characters (0-9) and space, which corresponds to integers 0 to 36. Here the plaintext is processed successively in blocks of size m. The encryption algorithm takes a block with m plaintext digits and transforms into a cipher block of size m using a key matrix of size $m \times m$ by the linear transformation, which is given by:

$$c_1 = (k_{1,1}p_1 + k_{1,2}p_2 + \cdots + k_{1,m}p_m) \mod 37$$
$$c_2 = (k_{2,1}p_1 + k_{2,2}p_2 + \cdots + k_{2,m}p_m) \mod 37$$

$$\ldots$$
$$c_m = (k_{m,1}p_1 + k_{m,2}p_2 + \cdots + k_{m,m}p_m) \bmod 37$$

Note: For this question, correspondence between plaintext and number modulo 37 are as follows "A" $\leftrightarrow$ 0, "B" $\leftrightarrow$ 1, "C" $\leftrightarrow$ 2, $\ldots$, "Z" $\leftrightarrow$ 25, "0" $\leftrightarrow$ 26, "1" $\leftrightarrow$ 27, "2" $\leftrightarrow$ 28, $\ldots$, "9" $\leftrightarrow$ 35 and " " (space) $\leftrightarrow$ 36

(a) How many different keys are possible in this system?

Solution:
The number of valid keys is the number of invertible matrices of size $m \times m$ over $\mathbb{Z}_{37}$, which can be calculated as:

$$\prod_{i=0}^{m-1}(37^m - 37^i) = (37^m - 1) \times (37^m - 37) \times (37^m - 37^2) \times ... \times (37^m - 37^{m-1})$$

(b) This cipher is easily broken with a known plaintext attack. An adversary discovers the following ciphertext is encrypted using this cipher with $m = 5$ (55 characters in total, no spaces):

`A8VS3XRDEON6JEVXGJID13C07L4C1R4Q965XWRA5DQGYWTNHYO4ND8Z`

If the following combination of plaintext and ciphertext is given (please replace both "?????" by the last five digits of your student number), decrypt the cipher by giving the plaintext as well as both encryption and decryption keys.

Plaintext	`X9B6T6JAW3UEY7FHIW?????5Z`
Ciphertext	`2Q59ZZ1Z?????UMDNY2JHINTS`

You need to show step-by-step details of your working. Make sure to include the details of any package, functions used, and/or programs developed. Simply showing the final result and/or a program would not receive marks.

Solution:
Note: The following solution is based on student number 1000000.

i. The given plaintext and ciphertext combination is

$$P = \texttt{X9B6T6JAW3UEY7FHIW000005Z}$$
$$C = \texttt{2Q59ZZ1Z00000UMDNY2JHINTS}$$

ii. Observe that the provided plaintext and ciphertext combination has 25 characters, which is five times the block size. We will first construct the plaintext matrix and ciphertext matrix.

(solution continued)

$$M_P = \begin{pmatrix} P_1 & P_6 & P_{11} & P_{16} & P_{21} \\ P_2 & P_7 & P_{12} & P_{17} & P_{22} \\ P_3 & P_8 & P_{13} & P_{18} & P_{23} \\ P_4 & P_9 & P_{14} & P_{19} & P_{24} \\ P_5 & P_{10} & P_{15} & P_{20} & P_{25} \end{pmatrix} = \begin{pmatrix} 23 & 32 & 20 & 7 & 26 \\ 35 & 9 & 4 & 8 & 26 \\ 1 & 0 & 24 & 22 & 26 \\ 32 & 22 & 33 & 26 & 31 \\ 19 & 29 & 5 & 26 & 25 \end{pmatrix}$$

$$M_C = \begin{pmatrix} C_1 & C_6 & C_{11} & C_{16} & C_{21} \\ C_2 & C_7 & C_{12} & C_{17} & C_{22} \\ C_3 & C_8 & C_{13} & C_{18} & C_{23} \\ C_4 & C_9 & C_{14} & C_{19} & C_{24} \\ C_5 & C_{10} & C_{15} & C_{20} & C_{25} \end{pmatrix} = \begin{pmatrix} 28 & 25 & 26 & 3 & 7 \\ 16 & 27 & 26 & 13 & 8 \\ 31 & 25 & 26 & 24 & 13 \\ 35 & 26 & 20 & 28 & 19 \\ 25 & 26 & 12 & 9 & 18 \end{pmatrix}$$

iii. Calculate the modular inverse of M_P.

$$M_P^{-1} = \begin{pmatrix} 31 & 19 & 4 & 24 & 31 \\ 29 & 13 & 20 & 23 & 18 \\ 19 & 9 & 23 & 26 & 5 \\ 6 & 31 & 3 & 28 & 11 \\ 26 & 6 & 6 & 14 & 29 \end{pmatrix}$$

iv. The encryption key can be calculated by $K_E = M_C \times M_P^{-1}$.

$$K_E = M_C \times M_P^{-1} = \begin{pmatrix} 30 & 5 & 3 & 33 & 19 \\ 24 & 8 & 31 & 11 & 7 \\ 35 & 9 & 3 & 0 & 36 \\ 32 & 19 & 23 & 11 & 33 \\ 22 & 13 & 32 & 16 & 0 \end{pmatrix}$$

v. The decryption key K_D is the modular inverse of K_E.

$$K_D = K_E^{-1} = \begin{pmatrix} 21 & 18 & 28 & 4 & 4 \\ 3 & 4 & 21 & 16 & 24 \\ 36 & 1 & 19 & 20 & 20 \\ 10 & 7 & 29 & 34 & 16 \\ 19 & 3 & 4 & 11 & 9 \end{pmatrix}$$

vi. Construct the matrix M'_C using ciphertext provided in the question.

$$M'_C = \begin{pmatrix} C'_1 & C'_6 & \dots & C'_{51} \\ C'_2 & C'_7 & \dots & C'_{52} \\ C'_3 & C'_8 & \dots & C'_{53} \\ C'_4 & C'_9 & \dots & C'_{54} \\ C'_5 & C'_{10} & \dots & C'_{55} \end{pmatrix}$$

(solution continued)

vii. The message can be decrypted by $M'_P = K_D \times M'_C$. The decrypted plaintext is (each space is replaced by an underscore):
`THE_QUIETER_YOU_BECOME_THE_MORE_YOU_ARE_ABLE_TO_HEAR___`

5. Probability [11 marks]

Let x be the fourth digit of your student ID (without leading zero), y be the sixth digit of your student ID. The value N used in this task is given by $5x + 6y + 15$.

For the below tasks, you need to show your working by providing formula used, and/or **short** explanation. Also give the numerical final answer (e.g. 1024 instead of 2^{10}).

(a) What is your value of N based on your student ID? You may simply show N, but please make sure that your calculation of N is correct, as you will need this value for the rest of tasks.

Solution:
The following solution is based on student ID 999999. This case $x = 9$ and $y = 9$, hence $N = 5 \times 9 + 6 \times 9 + 15 = 114$.

(b) Assuming that we have 230 students enrolled in this subject, and all student numbers are randomly generated. What's the probability that at least one of your classmate shares the same N with you? Your result should be rounded to three digits after the decimal point.

Solution:
$\{x = 9, y = 9\}$ is the only combination which can give $N = 114$. For any other student shares the same N, (s)he must have 9 as both fourth and sixth digit in his/her student ID. Since student IDs are randomly generated, the value of x and y are independent to each other. For each classmate, the probability of having a different N should be $1 - (0.1 \times 0.1) = 0.99$. Hence for 229 classmates, the probability that all of them having $\{x \neq 9, y \neq 9\}$ would be $0.99^{229} \approx 0.100$. Therefore the probability that at least one of them shares the same N with me would be $1 - 0.100 = 0.900$.

(c) How many ways to place N different balls into five different bins?

Solution:
We can place each ball one-by-one into bins. Each ball has five choice of bins that can be placed into, resulting in $5^N = 481482486096808963263994485646231829634525412053847048809984698891639709472656 25$ different ways.

(d) How many ways to place N identical balls into five different bins, so that all bins are non-empty?

Solution:
We may first place all balls in a row, then place four dividers to separate balls into five groups. As shown by the following graph, each "O" indicates a ball, and each "|" indicates a possible placement of a divider.
O|O|O|O|O|O|O|O|O|O|O|O|O|...|O|O|O|O|O|O|O|O|O|O|O
We need to choose 4 out of $N-1$ available spots to place our dividers, thus the total number of ways would be $\binom{N-1}{4} = \frac{113\times112\times111\times110}{4\times3\times2\times1} = 6438740$.

(e) How many ways to place N identical balls into five different bins?

Solution:
One way of solving this problem is to consider five different cases: How many ways to place N identical balls into five different bins, so that one, two, three, four, five bins are non-empty? If only one bin is non-empty, we will place all N balls into one of five bins, resulting in five different ways. If two of the bins are non-empty, there are $\binom{N-1}{1} = N-1$ different ways for placing N balls, times $\binom{5}{2} = 10$ different ways for selecting two bins out of five. Repeat this analysis with three, four, five non-empty bins. The total number of ways is $\binom{5}{1}\times\binom{N-1}{0} + \binom{5}{2}\times\binom{N-1}{1} + \binom{5}{3}\times\binom{N-1}{2} + \binom{5}{4}\times\binom{N-1}{3} + \binom{5}{5}\times\binom{N-1}{4} = 7673835$.
Another approach involves in adding an additional five balls. We will try to solve "How many ways to place $N+5$ identical balls into five different bins, so that all bins are non-empty?" first. For each valid placement, we can remove one ball from each bin, to make it a valid placement for this task. The total number of ways is $\binom{N+5-1}{4} = 7673835$.

(f) How many ways to place N identical balls into five identical bins, so that at most two bins are non-empty?

Solution:
$N = 0+N = 1+(N-1) = 2+(N-2) = ... = \lfloor\frac{N}{2}\rfloor + \lceil\frac{N}{2}\rceil$
Thus we have $\lfloor\frac{N}{2}\rfloor + 1$ different ways.