The University of Newcastle
School of Electrical Engineering and Computer Science

COMP3260/COMP6360 Data Security
Week 5 Workshop – 22nd and 24th March 2021

Solutions

1. For polyalphabetic substitution cipher with period d estimate the unicity distance, assuming that all keys are equally likely

   Solution:
   The unicity distance is defined as U = H(k)/D, where H(k) is the entropy of the key k, and D is the redundancy of the language. For English, we estimate D=3.2

   There are $(26!)^d$ possible keys so U = lg $(26!)^d$ / 3.2 = d × lg 26! / 3.2 ≈ 27.62d

2. Decipher the following ciphertext, which was enciphered using a Vigenere cipher with key ART: YFN GFM IKK IXA T.

   Solution:
   With a Vigenere cipher the key K is a sequence of letters $K = k_1k_2\ldots k_d$ where k_i gives the amount of shift in the i^{th} alphabet. That is:

   $f_i(X) = (x + k_i) \bmod n$, that is

   $c = (m + k) \bmod 26$, so $m = (c - k) \bmod 26$

YFN	GFM	IKK	IXA	T	Cipher
ART	ART	ART	ART	A	Key
YOU	GOT	ITR	IGH	T	Plaintext

   As numbers for first three characters.

24	05	13	c
00 17	19	k	
24 14	20	m	

   Alternatively, you can use the lookup table in the textbook. For plaintext letter m and key letter k, the ciphertext c is the letter in column m of row k. For ciphertext c, the plaintext m is the column containing c in row k.

   Plaintext is YOU GOT IT RIGHT.

3. Decipher the following ciphertext, which was enciphered using a Beaufort cipher with key ART: CDZ ORQ WRH SZA AHP

Solution:
Beaufort cipher:

$f_i(x) = (k_i - x) \bmod n$, that is

$c = (k - m) \bmod 26$, so $m = (k - c) \bmod 26$

CDZ	ORQ	WRH	SZA	AHP	Ciphertext
ART	ART	ART	ART	ART	Key
YOU	MAD	EAM	IST	AKE	Plaintext

As number for the first three characters:

02	03	25	ciphertext c
00	17	19	key k
24	14	20	plaintext m

Alternatively use Vigenere Tableau. As m + c = k, for plaintext letter m, the ciphertext letter c is the row containing the key k in column m. For ciphertext c, the plaintext m is the column containing k in row c.

Plaintext is YOU MADE A MISTAKE.

4. Consider a linear substitution cipher that uses the transformation $f(a) = ak \bmod 26$. Suppose you know with certainty that the plaintext letter J(*9*) corresponds to the ciphertext letter P(*15*), that is, $9k \bmod 26 = 15$. Break the cipher by solving for *k*.

Solution:
We have an equation of the form ax mod n = b: there are three cases

- When $gcd(a,n) = 1$: find solution x_0 to $ax \bmod n = 1$; then $x = bx_0 \bmod n$.
- When $gcd(a,n) = g$:
  - If g divides *b*, that is, $b \bmod g = 0$, then $ax \bmod n = b$ has *g* solutions of the form: $x = ((b/g)x_0 + t(n/g)) \bmod n$, for $t=0,1,\ldots,g-1$, where x_0 is the solution to $(a/g)x \bmod (n/g) = 1$.
  - If *g* does not divide *b* then there are no solutions.

To solve 9k mod 26 = 15 we need to calculate gcd(9,26). Applying Euclid's algorithm (or simply by inspection, as the values are very small) we show that gcd(9,26)=1, and therefore $k=15k_0 \bmod 26$, where k_0 is a solution of $9k_0 \bmod 26 = 1$, that is, k is a multiplicative inverse of 9 mod 26. Since 26=2x13, we can use CRT and we get
9k mod 2 =1, that is, $k_1 \bmod 2 = 1$, and
9k mod 13 = 1, thus $k_2 \bmod 13 = 3$

Solving these two equations gives us
k mod 2 = 1, $k_1=1$
k mod 13 = 3, $k_2=3$

We are now using CRT to find a common solution in the range [0,25]:

$13y_1 \bmod 2 = 1$, $y_2 = 1$
$2y_1 \bmod 13 = 1$, $y_1=7$

Thus $k_0 = (1x1x13 + 3x7x2) \bmod 26 = (13+42) \bmod 26 = (13+16) \bmod 26 = 3$
Therefore $k=15k_0 \bmod 26 = 45 \bmod 26 = 19$

5. Consider again a linear substitution cipher that uses the transformation *f(a) = ak* mod *26*. Suppose you know with certainty that the plaintext letter N(*13*) corresponds to the ciphertext letter N(*13*), that is, *13k* mod *26 = 13*. Can you break the cipher by solving for *k*? What about if you also know that the plaintext letter C(*2*) corresponds to the ciphertext G(*6*)?

Solution:
We have 13k mod 26 = 13 so we again have an equation of the form ax mod n = b, but unlike in the previous question we have gcd(13,26) = 13. Since 13 mod 13 = 0, the equation has 13 solutions of the form $k=(k_0+ 2t) \bmod 26$, where t=0,1,…,12, and k_0 is the solution of equation $k_0 \bmod 2 = 1$, this $k_0=1$.Therefore, we have k = 1,3,5,…,25 and we cannot break the cipher as we cannot uniquely determine k. However, if we also know that 2k mod 26 = 6, it follows that k also has to satisfy k=3+13t, t=0,1 so k=3 or 16. Thus the common solution is 3. Also note that 16 cannot be chosen for k in any case, as gcd(16,26)=2 and we would not have a one-to-one function, so we could not decrypt messages.

6. Consider again a linear substitution cipher that uses the transformation *f(a) = ak* mod *26*. Suppose that you suspect that the plaintext letter N(*13*) corresponds to the ciphertext letter P(*15*), that is, *13k* mod *26 = 15*. Can you break the cipher by solving for *k*?

Solution:
No, you can't break the cipher but you can tell that your guess is wrong, as the equation 13k mod 26 = 15 has no solutions.

7. Consider the Measure of Roughness $M = \sum_{i=0}^{n-1}(p_i - \frac{1}{n})^2$ and consider the alternative versions $M_1 = \sum_{i=0}^{n-1}(p_i - \frac{1}{n})$ and $M_2 = \sum_{i=0}^{n-1}|p_i - \frac{1}{n}|$. Could each of M_1 and M_2 be used in place of *M*? If yes, which is a better measure and why?

Solution:
M_1 cannot be used, as $M_1 = \Sigma p_i - \Sigma 1/n = 1-1 = 0$ and therefore M_1 does not depend on the frequency distribution of letters. M_2 can be used, but it takes all the probability deviations from 1/n equally into account, while M emphasizes larger deviations (which translate into greater 'roughness'). For example, if we have

a) $p_1=p_2=1/52$, $p_3=p_4=3/52$, and $p_5=\ldots=p_{26}=2/52$, and

b) $p_1=0$, $p_2=4/52$, and $p_3=\ldots=p_{26}=2/52$

(note that $\Sigma p_i = 1$ in both cases as it should be),
we have $M_1=0.076923$ in both cases, while M=0.001479 in a) and M=0.002959 in b).
Therefore, M appears to capture the notion of 'roughness' better than M_1.