1 Cryptography: An Overview

1.1 Elementary Ciphers

Problem 1. *We mentioned a substitution cipher in which each plaintext letter, represented by an integer x, is replaced by the letter corresponding to the integer $y = ax + b(\mod 26)$, where a and b are integers. If the alphabet we are using has n letters, where n is not necessarily 26, we can generalize this rule to $y = ax + b \pmod n$, where " $\mod n$" means that we take the remainder upon division by n. In answering the following questions, assume that the integers a and b are restricted to the values $0, \ldots, n-1$.*

(a) Suppose that n has the value 26, as it does if the plaintext is in English and we do not encrypt spaces or punctuation marks. Is there a reason not to use certain values of the constant a or of the constant b? If so, which values are the bad ones and what makes them bad?

(b) If we also count "space" as a character to be encrypted, we have $n = 27$. Now what, if any, are the bad values of a? Of b?

(c) For a general n, make a conjecture as to what will be the bad values of a and b, if there are any.

Starting with part (a), there are values of a in particular that break the utility of substitution cipher. The values of a that are *bad* are those that are not relatively prime to 26, in this case $\{[2], [4], [6], [8], [10], [12], [14], [16], [18], [20], [22], [24]\}$. Moving to part (b), our procedure here doesn't change: find all values not relatively prime to 27. For 27, the *bad* values are $\{[3], [6], [9], [12], [15], [18], [21], [24]\}$. Ending with the general case requested by part (c), we can say that unfit values for a in the general substitution cipher formula $y = ax + b \pmod n$ are the zero divisors in the $\mathbb{Z}_n$ ring; conversely, the fit values are the units U_n. It's important to note that these two sets are disjoint and that $\mathbb{Z}_n \backslash \{0\} = U_n \dot{\bigcup} \{\text{zero divisors of } \mathbb{Z}_n\}$ further reinforcing the *bad* vs. *good* paradigm in that a choice cannot be both unitary and a zero divisor. The units are uniquely *good* choices for a because they're invertible, ensuring a bijective (one-to-one and onto) mapping between the plaintext and the ciphertext which is crucial for decryption by Bob. Additionally, the choice $a = 0$ while in $\mathbb{Z}_n$ is not a permitted choice in the first place as it maps every letter to the choice of b, destroying the plaintext for Bob. Finally, there are no restrictions on b as any arbitrary linear shift won't change the bijectivity of the map between two sets of integers $\pmod n$.

Problem 3. *The following ciphertext was generated by a Vigenère cipher with a repeating key. All spaces and punctuation marks were removed from the plaintext, and the resulting ciphertext was broken into six-letter blocks.*

```
NRUATW  YAHJSE  DIODII  TLWCIJ  DIOPRA  DPANTO  EOOPEG
TNCWAS  DOBYAP  FRALLW  HSQNHW  DTDPIJ  GENDEO  BUWCEH
LWKQGN  LVEEYZ  ZEOYOP  XAGPIP  DEHQOX  GIKFSE  YTDPOX
DENGEZ  AHAYOI  PNWZNA  SAOEOH  ZOGQON  AAPEEN  YSWYDB
TNZEHA  SIZOEJ  ZRZPRX  FTPSEN  PIOLNE  XPKCTW  YTZTFB
PRAYCA  MEPHEA  YTDPSA  EWKAUN  DUEESE  YCNJPP  LNWWYO
TSKYEG  YOSDTD  LTPSED  TDZPNK  CDACWW  DCKYSP  CUYEEZ
MYDFMW  YIJEEH  WICPNY  PWDPRA  LSPSEK  CDACOB  YAPFRA
LPLLRA  YTHJCK  XEOQRK  XAOZUN  NEKFTO  TDAZFK  FROPLR
PSWYDE  DMKCEI  JSPPRE  ZUO
```

(a) Look for strings of three or more letters that are repeated in the ciphertext. From the separations of different instances of the same string, try to infer the length of the key.

(b) . . .

After briefly scanning the text, we'll use the trigram TDP as it appears three times in the 399-letter ciphertext.[1] It's important to remember that we aim to find the greatest common divisor of the distances between these instances, as we're predicting the key is cycling a set number of times between each instance.

```
NRUATW  YAHJSE  DIODII  TLWCIJ  DIOPRA  DPANTO  EOOPEG
TNCWAS  DOBYAP  FRALLW  HSQNHW  DTDPIJ  GENDEO  BUWCEH
LWKQGN  LVEEYZ  ZEOYOP  XAGPIP  DEHQOX  GIKFSE  YTDPOX
DENGEZ  AHAYOI  PNWZNA  SAOEOH  ZOGQON  AAPEEN  YSWYDB
TNZEHA  SIZOEJ  ZRZPRX  FTPSEN  PIOLNE  XPKCTW  YTZTFB
PRAYCA  MEPHEA  YTDPSA  EWKAUN  DUEESE  YCNJPP  LNWWYO
TSKYEG  YOSDTD  LTPSED  TDZPNK  CDACWW  DCKYSP  CUYEEZ
MYDFMW  YIJEEH  WICPNY  PWDPRA  LSPSEK  CDACOB  YAPFRA
LPLLRA  YTHJCK  XEOQRK  XAOZUN  NEKFTO  TDAZFK  FROPLR
PSWYDE  DMKCEI  JSPPRE  ZUO
```

The original ciphertext with all TDPs shown.

The first appearance starts at the 67th letter, the second starts at 121st letter, and the final instance of the trigram appears at the 223rd letter. The distance between the first two instances is 54, and the distance between the last two instances is 102. The factors of 54 are $\{1, 2, 3, 6, 9, 18, 27, 54\}$ while the factors of 102 are $\{1, 2, 3, 6, 17, 34, 51, 102\}$. Consequently, $\gcd(54, 102) = 6$, so the periodic key is potentially 6 characters long.

[1] Its important to note that there are five trigrams that appear three times: PRA, RAL, DTD, PSE, and our choice TDP. I quickly eliminated DTD, as the greatest common denominator between instances was 1 as no new information is revealed. While PRA and RAL have a gcd of 3 and 9, respectively, the six-letter partitioning presented by the textbook could be a hint that the gcd would also be 6, which was the case for TDP and PSE. This would also cover a key of period 3, meaning three of the five possible keys would be included in the solution. In the end, this process was *not* brief.

1.2 Enigma

Problem 1. *This is a counting problem focusing on the Enigma plugboard. Recall that the plugboard permutation interchanges some of the letters in pairs. For example, A and F might be interchanged, and M and X might be interchanged.*

(a) Suppose that only one pair of letters are interchanged and the other 24 letters are left unchanged. How many ways are there of choosing the special pair?

(b) In the standard Enigma machine, six pairs of letters were swapped. How many ways are there of choosing these six pairs? Does your answer agree with our rough estimate of 10^{11}*?*

Starting with (a), in order to select 2 plugs from a 26 plug plugboard without regard to their order, we can express this as

$$\binom{26}{2} = \frac{26!}{2! \times 24!} = 325 \text{ ways}$$

to choose special pairs. For (b), we increase the number of pairs to six, meaning we're now picking 12 plugs instead of 2 resulting in $\binom{26}{12}$. However, we also need to eliminate the duplicate instances of pairs with reverse order *and* eliminate the duplicate instances of pairs organized in a different order, both of which aren't relevant to our question. This is then

$$\binom{26}{12} = \frac{26!}{12! \times 12!} \times \frac{12!}{6! \times (2!)^6} \approx 5.5 \times 10^{11} \text{ ways}$$

to choose special pairs, confirming the given estimate of 10^{11} choices.

1.9 RSA

Problem 3. *Prove that* r_{k+1} *in the Euclidean Algorithm is the greatest common divisor of* a *and* b*.*

Proof. Suppose $a, b \in \mathbb{Z}$ such that $a \geq 0$, $b > 0$. We'll show using the Euclidean Algorithm that its penultimate remainder is a common divisor of a and b and that all other common divisors divide this penultimate remainder.

The Euclidean Algorithm is based on the Division Algorithm, which states that for any two integers a and b with $b > 0$, there exist unique integers q_1 and r_1 such that:

$$a = q_1 b + r_1, \quad 0 \leq r_1 < b.$$

If $r_1 = 0$, then b divides a, and $\gcd(a, b) = b$. Otherwise, apply the algorithm iteratively:

$$\begin{aligned} b &= q_2 r_1 + r_2, \quad 0 \leq r_2 < r_1, \\ r_1 &= q_3 r_2 + r_3, \quad 0 \leq r_3 < r_2, \\ &\vdots \\ r_k &= q_{k+2} r_{k+1} + r_{k+2}, \quad 0 \leq r_{k+2} < r_{k+1}, \\ r_{k+1} &= q_{k+3} r_{k+2} + 0. \end{aligned}$$

Since the remainders strictly decrease and are non-negative, the algorithm terminates with some remainder r_{k+1} such that $r_{k+2} = 0$. This means r_{k+1} divides r_k.

First, from the recursive equations above, each remainder is expressed as a linear combination of the previous two terms. Since r_{k+1} divides r_k, and r_k was formed using r_{k-1} and r_{k-2}, by induction, r_{k+1} divides all previous remainders, including b and a. Thus, r_{k+1} is a common divisor of a and b. Next, suppose d is any common divisor of a and b. Then, from the equations $d \mid a,\ d \mid b \implies d \mid r_1$. By induction, since each remainder is formed by a linear combination of the previous two, d must divide each remainder, including r_{k+1}. Since r_{k+1} is itself a divisor of a and b, and it is the largest such divisor in this process, it follows that $\gcd(a, b) = r_{k+1}$. Therefore, the penultimate remainder in the Euclidean Algorithm is the greatest common divisor of a and b. □