MISSION 2

Thao Bui, Ian Martiny, Colin Sidey, Bharadwaj Thirumal

December 25, 2016

1. **Task 1:** Find a Hill cipher key (block size 4) which encrypts both plaintexts (JESS and JEFF) to the same ciphertext.

   Because we know the methodology behind the Hill cipher it isn't too difficult for us to develop a key to confuse attempts at cryptanalysis. The Hill cipher interprets the plaintext as vectors then multiplies those vectors with a matrix which serves as the encryption key. The plaintext "jess" and "jeff" have very similar vectors Jess $=\vec{v_1} = \begin{pmatrix}9 & 4 & 18 & 18\end{pmatrix}$ and Jeff $=\vec{v_2} = \begin{pmatrix}9 & 4 & 5 & 5\end{pmatrix}$.

   If we want to hide the identity of Jeff or Jess, then we need the product of the vector multiplied with the matrix to be the same for vectors $\vec{v_1}$ and $\vec{v_2}$. We can accomplish this by doing a little linear algebra. For an encryption key matrix M and a cipher text vector $\vec{w}$

$$\vec{v_1} \times M = \vec{w} = \vec{v_2} \times M$$
$$\vec{v_1} \times M = \vec{v_2} \times M$$
$$\vec{v_1} \times M - \vec{v_2} \times M = 0$$
$$(\vec{v_1} - \vec{v_2}) \times M = 0$$
$$[\begin{pmatrix}9 & 4 & 18 & 18\end{pmatrix} - \begin{pmatrix}9 & 4 & 5 & 5\end{pmatrix}] \times M = 0$$
$$\begin{pmatrix}0 & 0 & 13 & 13\end{pmatrix} \times M = 0$$

   Let's now consider our encryption key matrix M. For a general M we have.

$$M = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} \end{pmatrix}$$

   If we multiply the difference of the vectors and M we have $(\vec{v_1} - \vec{v_2}) \times M =$

$$\begin{pmatrix}13(x_9 + x_{13}) & 13(x_{10} + x_{14}) & 13(x_{11} + x_{15}) & 13(x_{12} + x_{16})\end{pmatrix} = \vec{0} \quad (1)$$

   By examining (1), it's clear that in order to keep "jess" and "jeff" ambiguous we need not concern ourselves with the top two rows of M and thus can keep them random, with the only restriction being that they aren't linearly dependent with any other row in M. The simplest way to solve (1) would be to create a matrix in which the bottom two entries of each column were the same value but opposite in sign. For example $x_9 = -x_{13}$. There is however a problem with this, this would create two linear dependent rows and thus $det(M) = 0$.

The way to resolve this is to remember that we're working modulo 26. Each entry of our vector in (1) is multiplied by 13; since $13(2) \equiv 0 \pmod{26}$, we only need the sum of bottom two entries of each column of M to be equivalent to 2 modulo 26. Here's an example of a suitable matrix for disguising the vectors "jess" and "jeff."

$$M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 15 & 11 & 24 & 6 \\ 13 & 17 & 4 & 22 \end{pmatrix}$$

2. **Task 2:** Which pairs of coprime numbers have all quotients 1 in the Euclidean algorithm? It turns out they are a well-known collection of numbers with a familiar name. Conjecture a theorem, then state it and write a nice proof.

Theorem 1. *Consecutative Fibonacci numbers are co-prime and have all quotients equal to 1 in the Euclidean algorithm*

We prove this in two steps: first a Lemma to demonstrate that consecutive Fibonacci numbers are co-prime and then we use that to show that the quotients will always be one in the Euclidean algorithm.

Lemma 1. *Let $f_0, f_1, f_2, \ldots, f_n$ be the Fibonacci sequence then*

$$\forall n \geq 0,\ gcd(f_n, f_{n-1}) = 1$$

Proof. We prove this using induction.

Base Case: For $n = 0$, we get

$$\begin{aligned} gcd(f_1,\ f_0) &= gcd(1,\ 1) \\ gcd(f_1,\ f_0) &= 1 \end{aligned} \tag{2}$$

Inductive Step: $\forall k \geq 0$, we assume $gcd(f_{k+1},\ f_k)$ is true. That is,

$$gcd(f_{k+1},\ f_k) = 1 \tag{3}$$

Now we have to show that $gcd(f_{k+2},\ f_{k+1}) = 1$

By the definition of Fibonacci sequence, we know that

$$f_{k+2} = f_k + f_{k+1} \tag{4}$$

We also know that for any 2 integers a and b,

$$gcd(a,\ b) = gcd(a,\ a \pm b) \tag{5}$$

From (4) and (5), we can write

$$gcd(f_{k+2},\ f_{k+1}) = gcd(f_{k+1},\ f_k) \tag{6}$$

From (3), we know that $gcd(f_{k+1},\ f_k) = 1$

So (6) becomes

$$gcd(f_{k+2},\ f_{k+1}) = 1$$

□

We also know that between any two consecutive Fibonacci numbers the quotient (in the Euclidean algorithm) must be one. We can see this in a couple different ways. First we could recognize that the quotient of consecutive Fibonacci numbers approximates the golden ratio ≈ 1.618 meaning that our quotient in the Euclidean algorithm is one. Or we could recognize that the Fibonacci numbers are defined in a way such that $f_n = f_{n-1} + f_{n-2}$ where f_{n-2} satisfies the properties of necessary for our remainder ($0 \leq r < f_n$), meaning that $f_n = f_{n-1} + f_{n-2}$ is our proper decomposition via the Euclidean algorithm.

Thus this mostly completes the proof of our theorem. We now have that any two consecutive Fibonacci numbers are co-prime and that their quotient is 1 in the Euclidean algorithm. It just remains to show that the next step in the Euclidean algorithm will also have consecutive Fibonacci numbers.

Since the next step in the Euclidean algorithm takes the divisor and the remainder (which are consecutive Fibonacci numbers) and makes them the dividend and divisor (respectively), we are effectively done. Our Euclidean algorithm will have the following form:

$$\begin{aligned} f_n &= 1 \cdot f_{n-1} + f_{n-2} \\ f_{n-2} &= 1 \cdot f_{n-1} + f_{n-3} \\ &\vdots \\ 2 &= 1 \cdot 1 + 1 \\ 1 &= 1 \cdot 1 + 0 \end{aligned}$$

□

3. **Task 3:** Explain to your student how to perform a chosen plaintext attack on this system, if the block length is known. Give a full explanation: which plaintexts do you use, and how to recover the key.

Either by working with an example block size or by using the fact that all three ciphers are essentially linear functions of the plaintext we can come to the realization that our overall encryption function is a linear function of the plaintext as well.

This means that when we encrypt we will get an output such as the following (assuming a block size of n):

$$(x_1, x_2, \ldots, x_n) \mapsto (\alpha_{1,1}x_1 + \alpha_{1,2}x_2 + \cdots \alpha_{1,n}x_n + C_1, \alpha_{2,1}x_1 + \cdots + C_2, \ldots, \alpha_{n,1}x_1 + \cdots C + n)$$

our goal to figure out the "key" then is to determine the values of $\alpha_{i,j}, C_i$. In particular this means that we do not need to determine the values of each individual key for the Affine, Hill or Caesar shift cipher.

In order to find these coefficients we need $n + 1$ plaintexts, as follows: first we use the plaintext $(0, 0, \ldots, 0)$. This will give us the following output:

$$(0, 0, \ldots, 0) \mapsto (C_1, C_2, \ldots, C_n) = \vec{C}$$

so that we have the shifts of each component. (Note that these shifts are not just the Caesar shifts it includes components from the Affine shift, Caesar shift and part of the Hill key as well).

Now we use unit vectors to determine the values of our $\alpha_{i,j}$:

$$(1, 0, \ldots, 0) \mapsto (\alpha_{1,1} + C_1, \alpha_{2,1} + C_2, \ldots, \alpha_{n,1} + C_n)$$

Since we already know $(C_1, \ldots, C_n)$ we can subtract that vector to determine the values for $(\alpha_{1,1}, \ldots, \alpha_{n,1})$. Using these unit vectors we can determine the values of our $\alpha_{i,j}$ in the same way. That is $\vec{v_i}$ (unit vector in direction i) will give us the values of $\alpha_{j,i}$ (after subtracting $\vec{C}$).

Thus we have our key for encryption with a known plaintext attack.

4. **Task 4:** We need to determine the "key" for the given system. While there are three parts to the key, for the Affine, Hill, and Caesar ciphers we will just simply find a single key which will allow us to encrypt any message, rather than finding all parts of the key.

   To find the key we take advantage of the fact that we are composing 3 linear transformations, so we are getting a single linear transformation. Which tells us that our encryption function will behave as follows:

$$\begin{aligned}(w_1, w_2, w_3) \mapsto &(\alpha_1 w_1 + \alpha_2 w_2 + \alpha_3 w_3 + c_1, \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 + c_2, \\ &\gamma_1 w_1 + \gamma_2 w_2 + \gamma_3 w_3 + c_3) \pmod{26}\end{aligned} \tag{7}$$

   So our goal is to find all the coefficients, and we have the key.

   We begin with plugging in the zero vector: $(0\ 0\ 0) \mapsto (25,14,25)$. This gives us the overall shift involved from the cipher. Again we cannot break this into parts of the Affine, or Caesar cipher but this is the effective shift due to all of the encryptions combined. These are the c_i in (7).

   Now we find the coefficients on each input. To this end we plug in the unit vectors which gives:

$$\begin{aligned}(1,0,0) &\mapsto (18,5,11) = (\alpha_1 + c_1, \beta_1 + c_2 + \gamma_1 + c_3) \\ (0,1,0) &\mapsto (11,12,18) = (\alpha_2 + c_1, \beta_2 + c_2 + \gamma_2 + c_3) \\ (0,0,1) &\mapsto (4,0,4) = (\alpha_3 + c_1, \beta_3 + c_2 + \gamma_3 + c_3)\end{aligned}$$

   Since we need to find the actual value of our coefficients we need to subtract our c_i's `mod` 26. Which leaves us with:

$$\begin{aligned}(\alpha_1, \beta_1, \gamma_1) &= (19, 17, 12) \\ (\alpha_2, \beta_2, \gamma_2) &= (12, 24, 19) \\ (\alpha_3, \beta_3, \gamma_3) &= (5, 12, 5)\end{aligned}$$

   Which tells us our encryption function is:

$$\begin{aligned}(w_1, w_2, w_3) \mapsto &(19 \cdot w_1 + 12 \cdot w_2 + 5 \cdot w_3 + 25, 17 \cdot w_1 + 24 \cdot w_2 + 12 \cdot w_3 + 14, \\ &12 \cdot w_1 + 19 \cdot w_2 + 5 \cdot w_3 + 25) \pmod{26}\end{aligned}$$

   We can confirm this by plugging in arbitrary values for our plaintext and confirming we get the same ciphertext, for example:

$$(7, 12, 20) \mapsto (12, 11, 21)$$

   As a side note it is worth noting that none of this was necessary, examining the source code on your webpage gave away the keys as the Affine key: $(19, 3)$ the Hill cipher key: $\begin{pmatrix} 1 & 2 & 3 \\ 5 & 4 & 2 \\ 2 & 1 & 3 \end{pmatrix}$ and the Caesar shift cipher key: $c = 7$.