1. Suppose you have a magic box which has an input slot and an output slot. The box works as follows: If you write a prime number \( P \), a base \( B \), and an integer \( R \) on a strip of paper, feed the strip into the input slot, and wait one second, the machine will return a different strip through the output slot with a number \( X \). This number \( X \) solves

\[
B^X \equiv R \mod P.
\]

If no such \( X \) exists, then the machine returns a short blank paper strip.

You eavesdrop on two people, Gene and Hilary, using Diffie-Hellman Key Exchange, with \( P = 53 \), \( B = 3 \).

You overhear the public part of the exchange:

\[
3^G \equiv 31 \mod 53, \quad 3^H \equiv 21 \mod 53.
\]

Using the magic box, you discover \( G = 5 \), \( H = 11 \), that is,

\[
3^5 \equiv 31 \mod 53, \quad 3^{11} \equiv 21 \mod 53.
\]

a. (10 pt) Compute the shared secret (also called the agreed key). Explain what you are calculating and the method you use.

The shared secret is \((3^G)^H \equiv (3^H)^G \equiv 3^{GH}\).

One way to compute is by repeated squaring to find \( 21^5 \).

Another way is to use Fermat’s little Theorem:

\[
3^{55} \equiv 3^{32}3^3 \equiv 1 \cdot 3^3 \equiv 27.
\]

b. (20 pt) Explain why Diffie-Hellman Key Exchange seems to be secure, in real life, and why this magic box compromises the security of this cryptosystem.

The security of Diffie-Hellman rests in the apparent difficulty of the discrete log problem. There is no known polynomial time (relative to the number of digits of the input) algorithm to solve discrete logs.

The magic box solves the discrete log problem in constant time plus however long it takes to write the problem and the answer. This makes solving the discrete log easy.
2. Let $H(n)$ be the number of dots in a hexagonal grid with $n$ dots on a side. Shown below are the grids with 1, 2, 3, and 4 dots on a side.

\[ \begin{array}{c c c c c c}
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\end{array} \]

a. (10 pt) Compute the differences $H(3) - H(2)$, $H(4) - H(3)$, and $H(5) - H(4)$.

$H(3) - H(2) = 12$, $H(4) - H(3) = 18$, $H(5) - H(4) = 24$, by counting the dots.

b. (20 pt) Use an induction argument to show that for natural numbers $n$,

$$H(n) = 3n(n - 1) + 1.$$ 

The dots of the hexagon with $n$ dots on a side are covered by one dot in the center plus six triangles with $n-1$ dots on a side. Let $T(n-1)$ be the number of dots in a triangle with $n-1$ dots on a side. We see $H(n) = 6T(n-1)+1$.

Claim: $T(n-1) = n(n-1)/2$.

Proof, by induction. Base case is $T(0) = 0$. If there are no dots, there are no dots. Inductive hypothesis: suppose for a given $k > 0$, that $T(k-1) = k(k-1)/2$. Inductive step: I will show that $T(k) = (k+1)k/2$, the next case. $T(k)$ is the number of dots in a triangle with $k$ dots on a side. This is just one more row than the triangle with $k-1$ dots on a side, and that last row has $k$ dots. Thus, $T(k) = T(k-1) + k$. Using the inductive hypothesis, $T(k) = k(k-1)/2 + k = (k(k-1) + 2k)/2$. Using the distributive and commutative laws of multiplication, the result $T(k) = (k+1)k/2$ follows. By the principle of mathematical induction, $T(n-1) = n(n-1)/2$ for all natural numbers $n$.

$H(n) = 6n(n - 1)/2 + 1 = 3n(n - 1) + 1$. 

3. The World War I era ADFGVX cipher is a two step method. The first step is a substitution using two letters among ADFGVX to stand for each plaintext letter or digit. The second step is a keyword columnar transposition. Below the key word, the result of the first step is written out, using one column for each letter of the key word. The result is copied in columns, in alphabetical order of the letters of the key word.

Here is the substitution table:

<table>
<thead>
<tr>
<th>A</th>
<th>D</th>
<th>F</th>
<th>G</th>
<th>V</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>F</td>
<td>L</td>
<td>1</td>
<td>A</td>
<td>O</td>
</tr>
<tr>
<td>D</td>
<td>J</td>
<td>D</td>
<td>W</td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>F</td>
<td>C</td>
<td>I</td>
<td>Y</td>
<td>B</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>R</td>
<td>5</td>
<td>Q</td>
<td>8</td>
<td>V</td>
</tr>
<tr>
<td>V</td>
<td>6</td>
<td>K</td>
<td>7</td>
<td>Z</td>
<td>M</td>
</tr>
<tr>
<td>X</td>
<td>S</td>
<td>N</td>
<td>H</td>
<td>∅</td>
<td>9</td>
</tr>
</tbody>
</table>

Each letter or digits is substituted with the pair using first the letter at the left end of the row and second the letter at the top of the column. So L enciphers as AD.

a. (10 pt) Encipher THE using the substitution table.

XV XF GX

b. (20 pt) The following message was enciphered first by applying the ADFGVX substitution and second, using keyword columnar transposition with the key word PARIS:

DAVFAADX VXFGVXVX FFXXXVFX XGXGXA VA DAGXAVGG

What is the first word of the message?

<table>
<thead>
<tr>
<th>P</th>
<th>A</th>
<th>R</th>
<th>I</th>
<th>S</th>
<th>Plain text: ITWASTHEBESTOFTIMES∅</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>First word: IT</td>
</tr>
<tr>
<td>F</td>
<td>D</td>
<td>X</td>
<td>V</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>A</td>
<td>G</td>
<td>X</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>V</td>
<td>X</td>
<td>F</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>A</td>
<td>X</td>
<td>V</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>A</td>
<td>A</td>
<td>X</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>D</td>
<td>V</td>
<td>V</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>A</td>
<td>X</td>
<td>G</td>
<td></td>
</tr>
</tbody>
</table>
4. Farmer Magog comes home from the Neolithic Revolution. Magog has learned about the latest inventions: planting crops, and also writing numbers in base ten (using fashionable Arabic numerals).

a. (5 pt) Farmer Magog has a square field measuring 200 cubits on a side. If Farmer Magog harvests \( \frac{1}{50} \) bushels of grain per square cubit, to the nearest whole bushel, how much does Magog harvest?

\[
(200 \text{ cubits})(200 \text{ cubits})\left(\frac{1}{50} \text{ bushels / square cubit}\right) = 800 \text{ bushels.}
\]

b. (5 pt) Farmer Magog scratches this number in base ten on a stone tablet, taking 1 minute per digit. How long does it take Farmer Magog to record the harvest?

\[
\lceil \log_{10} 800 \rceil + 1 = 3, \text{ so 3 minutes (or just count the digits).}
\]

Magog is prosperous, and periodically acquires larger lands. Suppose Magog has a square field measuring \( C \) cubits on a side. Crop yield is the same \( \frac{1}{50} \) bushels per square cubit.

c. (10 pt) Let \( h(C) \) be the amount of Magog’s harvest, in bushels. Which complexity class, from the list below, is the smallest containing \( h(C) \), and which word best describes this class?

As suggested above, \( h(C) = C^2 \cdot \frac{1}{50} \text{ bushels/ \text{ cubic}} \).

Thus, \( h(C) \) is in \( O(C^2) \), which is type I and polynomial order.

d. (10 pt) Magog still records the harvest on a stone tablet. It takes time \( r(C) \) to write the number in base ten. Which complexity class, from the list below, is the smallest containing \( r(C) \), and which word best describes this class?

The time to record \( h(C) \) is \( r(C) = (\lceil \log_{10} h(C) \rceil + 1) \text{ minutes} \). Using the result for \( h(C) \) above, we see \( r(C) = (\lceil 2 \log_{10} C - \log_{10} 50 \rceil + 1) \text{ minutes} \). For large \( C \), \( r(C)/\log C \) is about 2, so \( r(C) \) is type III, which is logarithmic.

**Complexity classes:**

I. \( O(C^K) \), \( K > 0 \)  
   for example: \( O(C^1) \), \( O(C^2) \), \( O(C^{10}) \),

II. \( O(K^C) \), \( K > 1 \)  
   for example: \( O(2^C), O(e^C), O(10^C) \),

III. \( O(\log C) \),

IV. \( O(1) \).

In cases I and II, \( K \) is a some constant independent of \( C \).

**Descriptions:** constant, exponential, logarithmic, polynomial.
5. A eight bit linear feedback shift register generates a key for a binary stream cipher. If the register has contents $\begin{bmatrix} b_7 & b_6 & b_5 & b_4 & b_3 & b_2 & b_1 & b_0 \end{bmatrix}$ at one step, then at the next step, it shifts to $\begin{bmatrix} b_5 + b_2 + b_0 & b_7 & b_6 & b_5 & b_4 & b_3 & b_2 & b_1 \end{bmatrix}$. The new leftmost bit is the output of the machine.

a. (6 pt) Start with the register fill $\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$. Compute the first eight bits $z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8$ of the key stream by running the shift register eight steps. (The starting fill does not produce an output bit.)

<table>
<thead>
<tr>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Key stream is first column, 01011111.

b. (6 pt) Encipher the short plaintext 10011011 by computing $\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \oplus z_1 z_2 z_3 z_4 z_5 z_6 z_7 z_8$. The $\oplus$ is the usual bitwise exclusive or.

key 01011111
plain 10011011
cipher 11000100

c. (6 pt) Convert the result of part (b) to hexadecimal.

1100 0100 = $\$ C4.$

Suppose you have this shift register box, and you receive a long string of bits enciphered by the method above, possibly with a different key stream.

d. (12 pt) What information do you need to determine the key stream, and how do you decipher the message? More specifically: which bits and how many bits do you require to determine the entire key stream, and how do you decipher the message?

I need 8 bits of information. The 8 bit initial fill will work, or any 8 consecutive bits of the keystream. Exclusive or is its own inverse, so I decipher the message by computing $P = K \oplus C$, where $K$ is a keystream bit, $C$ is a ciphertext bit, and $P$ is the corresponding plaintext bit.
6. Imagine working for the cryptographic corps of your favorite country in the days before digital computers.

Fact 1. You use a mechanical polyalphabetic substitution cipher (maybe like an Enigma machine).
Fact 2. The machine takes a four letter key, reset at the beginning of each message.
Fact 3. Half of the messages you encipher begin “To the ministry of....”
Fact 4. Your superiors will not change their writing style.
Fact 5. Your superiors will not replace the cryptographic hardware.

a. (15 pt) Assume the adversary has a copy of your machine, but has not stolen your keys. Explain why Fact 3 helps your adversary decipher a lot of whole messages.

   The opening phrase “To the ministry of” is a crib. For messages which start that way, the adversary will have matched plaintext and ciphertext. The adversary could catalog all $26^4$ encryptions, arising from the four letter key, which is only about half a million, doing that work just once, not per message, to decrypt all the messages which begin with the crib. Possibly some careful analysis of the machine would allow a smaller catalog to work.

b. (15 pt) You learn Fact 4 and Fact 5 after complaining to the higher-ups. You can work within the crypto corps to improve the cryptographic protocol. What would you change about your cryptographic practice to make your system more secure?

   The crib could be eliminated by agreeing on a codebook. A dummy first word of the enciphered message could be used to indicate whether the opening phrase should be expanded to be “To the ministry of,” e.g. if the first word has an odd number of letters.

   The crib could be made less important by agreeing to reset the machine with a new key after the salutation. That way, although the adversary will read the crib, the adversary will not have easy access to the sensitive part of the message.
7 a. (15 pt) Suppose the text of each day’s New York Times is a sample of standard English prose, and each day is enciphered using a daily monoalphabetic substitution. Let $p$ be the probability that two letters circled at random in one day’s ciphertext will match. The value of $p$ is computed each day. Only letters count, and anything which is not a letter is ignored.

Which of the following is true, and why?

I. The value of $p$ will be near $\frac{1}{26}$. The exact value will vary above and below $\frac{1}{26}$, but overall will be approximately $\frac{1}{26}$.

II. The value of $p$ will be always be less than $\frac{1}{26}$. The exact value will typically be near a definite value less than $\frac{1}{26}$.

III. The value of $p$ will never be less than $\frac{1}{26}$. The exact value will typically be near a definite value more than $\frac{1}{26}$.

IV. There is not enough information to tell how $p$ compares with $\frac{1}{26}$.

The distribution of letters in standard English determines the value of $p$. If $p_A, p_B, \ldots, p_Z$ are the frequencies of the letters A through Z on a given day, then $p = p_A^2 + p_B^2 + \ldots + p_Z^2$, where we assume it is possible for the same position to be circled twice. The letter frequencies will not usually vary much from day to day, because the sample is so large.

The monoalphabetic substitution affects which letter occurs with which frequency, but the set of frequencies is unchanged by relabeling the letters. So $p$ is independent of the substitution of the day.

If the letter frequencies of the day were exactly uniform, then $p$ would be $\frac{1}{26}$. The letter frequencies of English are not uniform. E is much more common than Q. For this reason, $p$ is always at least $\frac{1}{26}$, and typically near a definite value more than $\frac{1}{26}$. Choice III.
7 b. (15 pt) A Vigenère encryption using a keyword produced the ciphertext below.

ICJEVAQIPW BCIIJRQFVIF AZCPQYMJAH NGFYDHWQQR NARELKBRYG PCSPKWBPUG 60
KBKZWDQZXS AFZLOIETV PSITQISOTF KKVTQPSOW KPVRJLIECH OHITFPSUDX 120
XRACLJSNLU BOIPRJHYPI EFJERBTVMU QOIJZAGYLO HSEOHWFJCL JGGTWACWEK 180
EGKZNASGEK AIEWTARJED PSJYHQQHLO EBSKHAJYWW KTXSLOBFEV QQQPHZWERZ 240
AARVHISOTF KOGCRCJLO KTRYDHHZLQ YSFYWDKZQD HCNTQCPRLD OARVHOSIER 300
CSKSHNARVH LSRNHPXLPW DSILPLZVQWO JOCNLDJRRXY JRCVPOEOL 360
JUFYRQFGLU PHYLWISOTF KJWERNSTZQ MIVCDEZCZV PHVCUEHFCB EBPFWGEPZ 420
ISOTFKOEOD NWQZQYHYPY AHKhISEEGB AKHTOECPH JFPRQ

The cryptanalyst received the ciphertext as a string of capital letters, and transcribed it in lines of sixty letters, numbered on the right edge, and put the letters in blocks of ten. The cryptanalyst underlined some short repeated sequences.

Determine the length of the keyword.

The repeated sequences are probably start at positions which differ by a multiple of the length of the keyword.

ISOTKF begins at positions 96, 246, 376, and 421.
The differences are: 246 - 96 = 150, 376-246 = 130, and 421 - 376 = 45.
Any set of three independent differences will do, and one can also use the repeated sequence WDS and do the same analysis.

The greatest common divisor of 150, 130, and 45 is 5, so that is the most likely length for the keyword.

Thanks to Treven Wall (CU Math) who produced this cipher text.
THE BONUS.

Recall Euler’s function $\phi(N)$ is the number of positive integers less than $N$ which are relatively prime to $N$.

A Carmichael number is a positive integer $N$ which is not prime and has the property that for any integer $a$ relatively prime to $N$, $a^{N-1} \equiv 1 \mod N$.

\begin{enumerate}
\item[(a)] (3 pt) The prime factorization of 561 is $3 \times 11 \times 17$. Compute $\phi(561)$.

The number 561 is the product of distinct primes to the first power, so $\phi(561) = (3 - 1)(11 - 1)(17 - 1) = 2 \times 10 \times 16 = 320$.

Alternatively, the prime divisors of 561 are 3, 11, and 17. Among the 561 natural numbers from 1 to 561, there are 187 multiples of 3, 51 multiples of 11 and 33 multiples of 17.

But 17 numbers are multiples of both 3 and 11, 11 numbers are multiples of both 3 and 17, and 3 numbers are multiples of both 11 and 17.

Finally, we must exclude the one multiple of 3, 11, and 17, that is, 561.

$$\phi(561) = 561 - 187 - 51 - 33 + 17 + 11 + 3 - 1.$$ 

This is a classic inclusion-exclusion argument, which we never covered in class. But we did do the first way.

\item[(b)] (12 pt) The number 561 is a Carmichael number, meaning that 561 is not prime, and for any integer $a$ relatively prime to 561, the congruence $a^{560} \equiv 1 \mod 561$ holds. (Notice that $560 = 561 - 1$.) Justify that 561 is a Carmichael number.

By Fermat’s little Theorem, if 3 does not divide $b$, then $b^2 \equiv 1 \mod 3$. If $(a, 561) = 1$, then 3 does not divide $a$, and so does not divide $a^{280}$ either. By Fermat’s little Theorem, $a^{560} \equiv (a^{280})^2 \equiv 1 \mod 3$.

Ditto if 11 does not divide $b$. Ditto if $(a, 561) = 1$, then 11 does not divide $a$, and so does not divide $a^{56}$. Ditto $a^{560} \equiv (a^{28})^{10} \equiv 1 \mod 11$. Ditto 17.

The only integer from 0 to 560 which is congruent to 1 mod 3, 11, and 17 is 1, so $a^{560} \equiv 1 \mod 561$.

In fact, let $\lambda$ be the least common multiple of 2, 10, and 16. Then $a^{\lambda} \equiv 1 \mod 561$. The value of $\lambda$ is 80, and 560 is a multiple of 80.

The relevance is that one cannot use the (false) converse of Fermat’s little Theorem to prove that $N$ is prime. There are a few composite numbers which pass that test.