Cryptography challenge: facilitator notes

The Alan Turing Institute

Last updated: October 27, 2023

General comments

Timings, etc.

Whom to ask if you need help.

1 Introduction

For any students who are struggling to crack the initial cipher, the best place to start is the first and last lines, which are respectively "Hello Bob" and "Alice". Hints to this include the capitalisation of the names, as well as the repeated letters in Bob's name. With these letters filled in, the rest of the message becomes

> Hello Bob,
>
> Ho* a*e *o*?
> Le*'* *ee* o* *o**a*!
>
> Alice

(The answer is Monday.)

Symmetric ciphers

2 Caesar cipher

The Caesar cipher page did not seem to pose any problems in the 2023 sessions. The answer for the question at the bottom is `HIDDENMESSAGE`.

For very advanced students, you may wish to ask them why there are only 26 possibilities (or 25 useful ones) for a Caesar cipher. The answer is that there are only 26 letters in the English alphabet, so shifting by 27 is the same as shifting by 1—this is a very early example of the modular arithmetic which will be covered in the second half.

3 Monoalphabetic ciphers

For why texts may deviate from the standard frequency distribution: a simple one is the use of different spellings, e.g. American vs British English. More examples can be found at https://en.wikipedia.org/wiki/Letter_frequency.

One thing worth pointing out is that frequencies are only really stable if the text is long enough.

The decoding challenge is probably the part of the first half that students will spend the most time on, if only because there is a lot of trial-and-error involved. It is generally obvious that the most common letter in English is E, but the next most common letters tend to depend on the text from which it is drawn, so are less immediately useful.

From the frequency analysis, students should be able to guess that ciphertext `I` corresponds to plaintext `E`. Furthermore, in the ciphertext, the trigram `PVI` appears regularly. Together this implies that `P` corresponds to `T` and `V` corresponds to `H`. Sensible next steps would be to try to match up the remaining common letters, namely `Y` and `A`: these correspond to `A` and `O` in plaintext respectively.

The full answers are:

ciphertext	A	B	C	D	E	F	G	H	I	J	K	L	M
plaintext	O	B	S	G	P	I	M	C	E	F	Q	K	V
ciphertext	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
plaintext	X	R	T	W	N	Z	D	L	H	J	Y	A	U

4 Vigenère cipher

Reading the Vigenère square the other way around works because the square is symmetric: at a more fundamental level this is because the encoding is effectively carried out by adding two numbers, and addition is a commutative operation.

We found that providing a concrete example of calculating the IoC for a short piece of text was helpful for aiding students' understanding.

To prove that the IoC for long, random text is 1: let the length of the text be N. Then, we have that $n_\mathrm{A} = n_\mathrm{B} = \cdots = N/26$, and so:

$$\mathrm{IoC} = 26 \cdot \frac{N/26(N/26-1)\cdot 26}{N(N-1)} = \frac{N-26}{N-1},$$

and as $N \to \infty$, this tends to 1. (Strictly, we only need that $N \gg 26$ or that $N/26 \gg 1$, which the question states.)

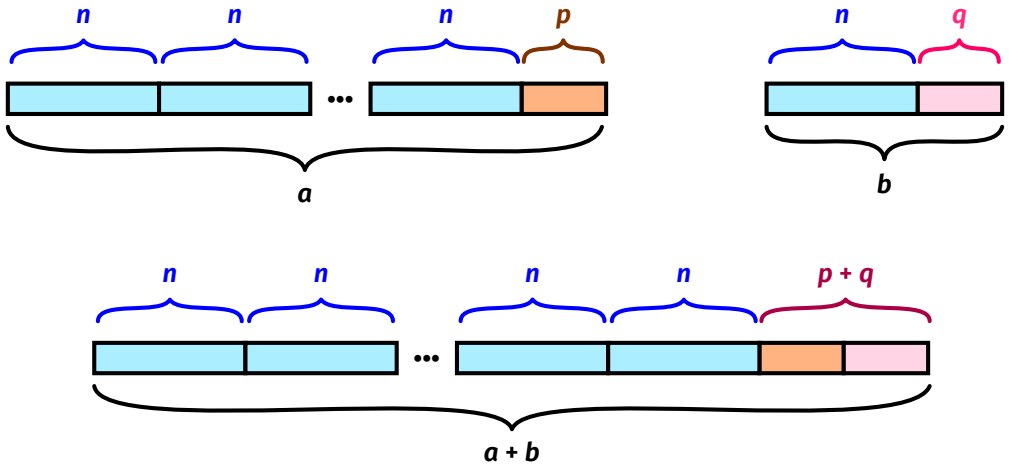
For the decoding challenge, the key is (quite literally) `KEY`. The text is the opening of Turing's 1936 paper *On Computable Numbers, with an Application to the Entscheidungsproblem* (DOI: 10.1112/plms/s2-42.1.230).

5 Enigma

Asymmetric ciphers

6 Modular arithmetic I

Generally, it really helps if you can use pen-and-paper diagrams to illustrate the idea of the 'remainder'. For example, this is a pictorial argument for $(a + b) \equiv (p + q) \bmod n$:



Mathematically, the proof for $(ab) \equiv (pq) \bmod n$ follows the same structure as for the earlier one. Let $a = p + cn$, and $b = q + dn$. Then

$$\begin{aligned} ab &= (p + cn)(q + dn) \\ &= pq + (pd + qc)n + cdn^2 \\ &\equiv pq \bmod n. \end{aligned}$$

One thing that students are very tempted to do is to 'factorise out' the $\bmod n$ part, like this:

$$a + b \equiv (p \bmod n) + (q \bmod n) = (p + q) \bmod n.$$

This is not valid, because the mod operator is not a function, and so does not distribute over addition. To explain this, you could use the example of subtraction: say $a = p - x$ and $b = q - x$. You certainly cannot 'factorise out' the $-x$ part like this:

$$a + b = (p - x) + (q - x) = (p + q) - x.$$

The proof by induction goes as follows:

- *Base case.* Consider $m = 1$. Then $a^1 \equiv a \equiv p \equiv p^1 \bmod n$.
- *Inductive step.* Assume that $(a^m) \equiv (p^m) \bmod n$ for some m. Then

$$a^{m+1} = a^m \cdot a \equiv p^m \cdot p \bmod n = p^{m+1} \bmod n,$$

  where the middle equality follows from statement (2).

For the challenge question, the answer is $7^{175} \bmod 16 = 7$. The suggestion to find a pattern should lead students to notice that

$$\begin{aligned} 7^1 \bmod 16 &= 7 \\ 7^2 \bmod 16 &= 1 \\ 7^3 \bmod 16 &= 7 \end{aligned}$$

and so on, which implies that any odd power of 7 is congruent to 7 modulo 16.

Formally, we can show this by noting that $7^2 = 49 \equiv 1 \bmod 16$. So,

$$\begin{aligned} 7^{175} \bmod 16 &= (7^{173} \cdot 7^2) \bmod 16 \\ &\equiv (7^{173} \cdot 1) \bmod 16 \quad \text{(using statement (2) in the notes)} \\ &= 7^{173} \bmod 16 \\ &\equiv (7^{171} \cdot 7^2) \bmod 16 \\ &\vdots \\ &\equiv 7^1 \bmod 16 = 7. \end{aligned}$$

7 RSA scheme I

8 Modular arithmetic II

The answer for the exercise at the bottom of the page is $d = 325$.

Specifically: we need $ed + y\phi = 1$, or substituting in the values, $13d + 352y = 1$.

If we divide 352 by 13 we find that $352 = (13 \cdot 27) + 1$, or equivalently, $(13 \cdot -27) + (352 \cdot 1) = 1$. This suggests a value of $d = -27$ and $y = 1$. As before, we need to add and subtract $(13 \cdot 352)$ to make d positive:

$$\begin{aligned}
1 &= (13 \cdot -27) && + (352 \cdot 1) \\
&= (13 \cdot -27) + (13 \cdot 352) && - (13 \cdot 352) + (352 \cdot 1) \\
&= (13 \cdot (-27 + 352)) && + (352 \cdot (1 - 13)) \\
&= (13 \cdot 325) && + (352 \cdot (-12)) \\
&= (13d) && + (352y),
\end{aligned}$$

which lets us read off $d = 325$ (and the unneeded $y = -12$).

9 RSA scheme II

To crack the RSA scheme, we need to find d (using the Euclidean algorithm), which requires us to know the value of ϕ. Because $\phi = (p-1)(q-1)$, this in turn means that we need to know the prime factors p and q which are multiplied together to form n.

In the first case where n is small (143), this is trivial: we have that $p = 11$ and $q = 13$. Thus $\phi = 10 \cdot 12 = 120$, and we can find $d = 103$ using exactly the same steps as on the previous page. Plugging this value of d into the script gives the correct answer $m = 50$.

In the second case, factorising 373577 is much harder (and so is calculating d, at least without a computer). The answer is $m = 500$, but the students are *not* expected to complete this.

For the final section involving the logarithmic graph, the exact value of n found will vary on the computer (and browser) being used. On one occasion, Chrome outperformed Safari by a factor of 3. In any case, when tested on Firefox 119.0 on a 2021 MacBook Pro (32 GB RAM, Apple M1 Pro), the regression equation obtained was

$$\log_{10}(t/\text{ms}) = 0.43 \log_{10} n - 4.00$$

and plugging in a time $t = 365 \cdot 24 \cdot 60 \cdot 60 \cdot 1000$ ms (approximately 1 year) gives $\log_{10} n = 33.72$, i.e., somewhere around $5 \cdot 10^{33}$.

It should be noted that the prime factorisation algorithm used is hardly the most performant (it is a brute-force search). However, it is perfectly serviceable in illustrating the point that the RSA scheme requires large numbers.

10 Cryptography Game Resources