Alice sends messages to Bob, Carla, and Dante. She encrypts each message with a one-time pad, but she’s lazy and uses the same key for each message. In this section, we’ll show how Eve can decrypt all three messages.

Suppose the messages are $M_1, M_2, M_3$ and the key is $K$. The ciphertexts $C_1, C_2, C_3$ are computed as $M_i \oplus K = C_i$. Eve computes

$$C_1 \oplus C_2 = (M_1 \oplus K) \oplus (M_2 \oplus K) = M_1 \oplus M_2.$$ 

Similarly, she obtains $M_1 \oplus M_3 = C_1 \oplus C_3$ and $M_2 \oplus M_3 = C_2 \oplus C_3$. The key $K$ has disappeared, and Eve’s task is to deduce $M_1, M_2, M_3$ from knowledge of $M_1 \oplus M_2, M_1 \oplus M_3, M_2 \oplus M_3$. The following example shows some basic ideas of the method.

Let’s assume for simplicity that the messages are written in capital letters with spaces but with no other punctuation. The letters are converted to ASCII using

$$A = 1000001, B = 1000010, \cdots, Y = 1011001, Z = 1011010, \text{ space } = 0100000$$

(the letters $A$ to $Z$ are the numbers 65 through 90 written in binary, and $\text{ space}$ is 32 in binary).

The XORs of the messages are

$M_1 \oplus M_2 =$

0000000 0001100 1100100 0001101 1100001 0011110 1111001 1100001
1100100 0011100 0001001 0001111 0010011 0010111 0000111

$M_1 \oplus M_3 =$

0000001 1100101 1100001 0000101 1100001 0011011 1100001 1100001
1100101 1110011 0010001 0001100 0010110 0000100 0001101

$M_2 \oplus M_3 =$

0000001 1101011 0000101 0001000 0000000 0000101 0011100 1110010
0000001 1101111 0011000 0000011 0000101 0010011 0001010.

Note that the first block of $M_1 \oplus M_2$ is 0000000. This means that the first letter of $M_1$ is the same as the first letter of $M_2$. But the observation that makes the biggest difference for us is that “space” has a 0 as its leading bit, while all the letters have 1 as the leading bit. Therefore, if the leading bit of an XOR block is 1, then it arises from a letter XORed with “space.”

For example, the third block of $M_1 \oplus M_2$ is 1100100 and the third block of $M_1 \oplus M_3$ is 1100001. These can happen only from “space $\oplus$ 1000100” and “space $\oplus$ 1000001”, respectively. It follows easily that $M_1$ has “space” as its third entry, $M_2$ has 1000100 = $H$ as its third entry, and $M_3$ has $A$ as its third entry. Similarly, we obtain the 7th, 8th, and 9th entries of each message.
The 5th entries cause a little more trouble: $M_1 \oplus M_2$ and $M_1 \oplus M_3$ tell us that there is “space” XORed with $A$, and $M_2 \oplus M_3$ tells us that the 5th entries of $M_2$ and $M_3$ are equal, but we could have “space A A” or “A space space.” We need more information from surrounding letters to determine which it is.

To proceed, we use the fact that some letters are more common than others, and therefore certain XORs are more likely to arise from these than from others. For example, the block 0010001 is more likely to arise from $E \oplus T = 100101 \oplus 1010100$ than from $K \oplus Z = 1001011 \oplus 1011010$. The most common letters in English are $E, T, A, O, I, N, S, H, R, D, L$. Make a table of the XORs of each pair of these letters. If an XOR block occurs in this table, we guess that it comes from one of the pairs yielding this block. For example, 0001001 arises from $A \oplus H$ and from $E \oplus L$, so we guess that the 11th block of $M_1 \oplus M_2$ comes from one of these two combinations. This might not be a correct guess, but we expect such guesses to be right often enough to yield a lot of information.

Rather than produce the whole table, we give the blocks that we need and the combinations of these frequent letters that produce them:

- 0000001 : $D \oplus E$, $H \oplus I$, $R \oplus S$
- 0000111 : $L \oplus O$
- 0001000 : $A \oplus D$, $I \oplus L$
- 0001100 : $D \oplus H$, $E \oplus I$
- 0001101 : $D \oplus S$, $E \oplus R$
- 0010001 : $E \oplus T$
- 0010011 : $A \oplus R$
- 0010111 : $D \oplus R$, $E \oplus S$
- 0011000 : $L \oplus T$
- 0011011 : $H \oplus S$, $J \oplus R$, $O \oplus T$
- 0011110 : $L \oplus R$

Let’s look at the next to last block of each XOR. In $M_1 \oplus M_2$, we have 0010111, which could come from $D \oplus S$ or $E \oplus R$. In $M_1 \oplus M_3$, we have 0010100, which could come from $A \oplus E$ or $H \oplus L$. In $M_2 \oplus M_3$, we have 0010011, which could come from $A \oplus R$. The only combination consistent with these guesses is $E, R, A$ for the next to last letters of $M_1, M_2, M_3$, respectively. This type of reasoning, combined with the information obtained from the occurrence of spaces, yields the following progress (“.” indicates a space, “*” indicates a letter...
to be determined):

\[ M_1 = *E-*R-A-SE*RE* \]
\[ M_2 = *ID-LY-DOL*AR* \]
\[ M_3 = *A-IERE-T*DA* \]

The first letter of \( M_3 \) is a one-letter word. The XOR block 0000001 gives us the possibilities \( D, E, H, I, R, S \), so we guess \( M_3 \) starts with \( I \). The XORs tell us that \( M_1 \) and \( M_2 \) start with \( H \).

Now let’s look at the 12th letters. The block 0001111 for \( M_1 \oplus M_2 \) suggests \( A \oplus N \). The block for \( M_1 \oplus M_3 \) suggests \( D \oplus H \) and \( E \oplus I \). The block for \( M_2 \oplus M_3 \) suggests \( L \oplus O \). These do not have a common solution, so we know that one of the less common letters occurs in at least one of the messages. But \( M_2 \) ends with \( DOL*AR* \), and a good guess is that this should be \( DOLLARS \). If so, then the XOR information tells us that \( M_1 \) ends in \text{SECRET}, and \( M_3 \) ends in \text{TODAY}, both of which are words. So this looks like a good guess. Our progress so far is the following:

\[ M_1 = HE-***R-A-SECRET \]
\[ M_2 = HIDE-LY-DOLLARS \]
\[ M_3 = I-AM-IERE-TODAY \]

It is time to revisit the 5th letters. We already know that they are “space A A” or “A space space.” The first possibility requires \( M_1 \) to have two consecutive one-letter words, which seems unlikely. The second possibility means that \( M_3 \) starts with the two words \( I-A* \), so we can guess this is \( I \ AM \). The XOR information tells us that \( M_1 \) and \( M_2 \) have \( H \) and \( E \), respectively. We now have all the letters:

\[ M_1 = HE-HAR-A-SECRET \]
\[ M_2 = HIDE-LY-DOLLARS \]
\[ M_3 = I-AM-IERE-TODAY \]

Something seems to be wrong in the 6th column. These letters were deduced from the assumption that they all were common letters, and this must have been wrong. But at this point, we can make educated guesses. When we change \( I \) to \( H \) in \( M_3 \), and make the corresponding changes in \( M_1 \) and \( M_2 \) required by the XORs, we get the final answer:

\[ M_1 = HE-HAS-A-SECRET \]
\[ M_2 = HIDE-MY-DOLLARS \]
\[ M_3 = I-AM-HERE-TODAY \]

These techniques can also be used when there are only two messages, but progress is usually slower. More possible combinations must be tried, and more false deductions need to be corrected during the decryption. The process can

The use of spaces and the restriction to capital letters made the decryption in the example easier. However, even if spaces are removed and more symbols are used, decryption is usually possible, though much more tedious.

During World War II, problems with production and distribution of one-time pads forced Russian embassies to re-use some keys. This was discovered, and part of the Venona Project by the U.S. Army’s Signal Intelligence Service (the predecessor to NSA) was dedicated to deciphering these messages. Information obtained this way revealed several examples of Russian espionage, for example, in the Manhattan Project’s development of atomic bombs, in the State Department, and in the White House.