2015 STEGANOGRAPHY

Steganography concerns techniques for embedding hidden messages into innocuous-seeming communication, such as email messages or images, so that nobody except you and your intended receiver can even tell that there is a hidden message — let alone the message.

The Tale of Demeratus

In ancient Greece people used to send each other messages by writing on wax tablets. These were wooden tablets that were coated in wax and people engraved their messages in the wax using a sharp stylus. Wax tablets were commonly used because they were portable and reusable: after a message has been read, the tablet can be heated so that the wax softens and then the wax can be smoothed out so that the tablet becomes blank and can be used again.

Around 500 BC King Demeratus of Sparta was living in exile in Persia and learned of a plan by the King of Persia to invade Greece. He wanted to warn his countrymen of the impending attack but in those days all messages were delivered by messengers and checked by guards along the way. He could not just send a plain message stating what he wanted to say because it would be stopped by the guards. Instead he scraped the wax off a wax tablet, wrote his warning on the wood beneath, then re-coated the tablet in wax. This resulted in a seemingly blank tablet, which was delivered to Sparta without any trouble. When the tablet was received it took them a while to figure out that there was a hidden message under the wax but they discovered it early enough that they were able to be quite prepared for the Persian invasion!

These days Demeratus's strategy would probably not work — it is strange to send wax tablets instead of letters or email and suspicious border guards might inspect the wax tablet further. Suppose you were in Demeratus's situation in more modern times and wanted to send a warning message to your country by sending a postcard through the postal mail service. Imagine that all postal mail is inspected by voracious border guards who look for signs of dubious activity and eat any post that looks suspicious or unusual.

In Demeratus's case, he had to hope that the recipients would realise something was fishy and think to look underneath the wax. That was probably why he left the tablet blank. If he had already had an agreement with the recipients in Sparta he could have written an innocent-looking message on the wax to make the tablet look really just like a normal communication. Had he done that however his recipients would probably have discarded the tablet as unimportant. For this question you may assume that you can agree on a plan with your recipient beforehand about how you will hide the message. You can then try to make your postcard look as much like a normal (handwritten) postcard as possible!

Question 1

Suppose that the guards do not let you write your own letters. You dictate what to write on your postcard and they will write your message for you. How could you send your message so that it would reach its destination safely?

Question 2

Suppose you are now allowed to choose your own postcard at a souvenir shop and write a message on the postcard yourself. Suggest another way to hide your message, given this extra freedom, that was not previously possible.

Child's Play

When Demeratus was a child he had a good friend Leonidas and they both enjoyed playing with coloured tiles. Each tile was square (of identical size) and they came in four colours: scarlet, ruby, lime and emerald. Demeratus and Leonidas used to arrange these into mosaics of 10×10 tiles before showing their mosaics to each other.

Scarlet and ruby are *red colours,* and lime and emerald are *green colours*. Two mosaics A and B are considered to look similar to each other if at each position in the mosaic, either both A and B have red-coloured tiles, or both A and B have green-coloured tiles.

Question 3

Demeratus and Leonidas have devised an algorithm for communicating favourite numbers (between 1 and 10^{30}) to each other, by taking an existing mosaic and producting a similar mosaic with the number embedded. Suggest a suitable algorithm.

Question 4

Leonidas wishes to communicate larger favourite numbers (between 1 and 10^{60}) but has been unable to come up with a suitable algorithm. Does such an algorithm exist? Justify your answer.

Question 5

Leonidas's little niece Gorgo likes to look at the mosaics he exchanges with Demeratus. When Leonidas and Demeratus start using your steganographic system, do you think Gorgo will be able to tell that something is different from before? What are the clues that Gorgo might get?

The Digital Age

Demeratus's great-great-great-great-great-great-great-great-granddaughter Ariadne gets a digital camera for her birthday. She learns that a digital photograph is made up of small squares called *pixels*, much like the coloured mosaics that Demeratus made as a child. The pixels are so small that the human eye, rather than seeing the individual squares, sees the picture as a whole.

In grayscale pictures the colour of each pixel is represented by a number between 0 and 255 (inclusive); 0 corresponds to black, 255 corresponds to white, and the numbers in between correspond to (increasingly lighter) shades of gray. When the camera takes a photograph light enters the lens of the camera from the scene that the camera is facing (and taking a picture of). The light from the scene lands on a small rectangular sensor which detects the intensity of light that lands on each pixel of the rectangle. The intensities are then converted into values between 0 and 255 which are stored as the digital image.

Question 6

Ariadne takes several photos of a scene in quick succession without moving her camera. The scene is a stationary landscape which does not change while she is taking the photographs. In what way would you expect the resulting digital images to vary? What do you think would stay the same and what would change?

Suppose that Ariadne dropped her camera and it started malfunctioning in a strange way: specifically, the intensity value of every tenth pixel was rounded down to the closest multiple of 3. That is, if the real pixel intensity is *x*, then the camera records the pixel intensity as *x - (x mod 3)*.

Question 7

What would the photos taken with the malfunctioning camera look like to the human eye?

If you had a computer and could analyze the list of intensities in the image then it would be very easy to identify a picture taken with Ariadne's malfunctioning camera: just check every tenth pixel and if all the corresponding intensities are divisible by 3 then it is probably a picture from Ariadne's camera. (It would be very unlikely that every tenth pixel would have an intensity divisible by 3 just by chance, but it could happen! So, this method is not always successful but it has a very high success probability.)

Suppose that Ariadne dropped her camera and it started malfunctioning in a different way, affecting only certain pixels, as follows: if the real pixel intensity is 100 then the malfunctioning camera records the pixel intensity as 100 with probability half and as 200 with probability half. If the real pixel intensity is 200 it behaves in exactly the same way.

Question 8

What would the photos taken with this malfunctioning camera look like to the human eye?

Question 9

Suppose you have a computer and can analyze the list of intensities in the image. Suggest a method to identify pictures taken with Ariadne's malfunctioning camera. Your method need not be always correct so briefly explain why you think that your method is better than random guessing.

Ariadne drops her camera again and to her delight it starts working properly again. She takes some photos while on holiday in Persia and decides to send them to a friend along with an embedded steganographically hidden message.

Question 10

Some images are more suitable for embedding hidden messages than others. Rank the image types listed below in order of how easy you think it would be for an enemy to detect that there is a message hidden in the image:

a photograph of a forest
a slide from a PowerPoint presentation
a photograph of a blank whiteboard
a photograph of a teacher standing in front of a blank whiteboard
a smiley emoticon

What properties of an image do you think make it good for embedding messages in a hard-to-detect way?

Ariadne takes her text (expressed as a string of 0s and 1s, with as many bits as the number of pixels) without encrypting or garbling it, and uses the following algorithm to embed it into the photograph:

```
For i = 1 to message_length_in_bits:
  Let b_i denote the i^th bit of the message
  Let p_i denote the i^th pixel intensity (between 0 and 255) in the photograph
    If p_i is even and b_i = 1 then p_i = p_i + 1
    If p_i is odd and b_i = 0 then p_i = p_i - 1
```

Question 11
Suppose you are given a photo where Ariadne has embedded a message using the above algorithm. Give an algorithm that takes the list of pixel intensities in the photo as input, and outputs the embedded message.

Question 12
Suppose you are given a photo and you are asked to guess whether it is a unmodified photo taken by Ariadne's camera or if it is a photo where Ariadne has embedded a message. Do you think you could tell the difference *just by looking* at the photo? Justify your answer.

Question 13
Suppose you have a computer and can analyze the list of intensities in the photo you are given. How can you tell the difference between an unmodified photo and a photograph with an embedded message?

Ariadne is worried that other people will discover that she is steganographically hiding messages in her photos and that they will be able to read her private messages. In order to prevent this, she decides to encrypt her messages before embedding them in her photos.

In steganography if an enemy knows the steganographic system that you are using then he can read your messages. The aim of steganography is to prevent the enemy from even realising a secret message is being sent — then he will not be able to read your messages. Encryption is different in that even if the enemy knows your system and sees an encrypted message he will not be able to learn the contents of the message. In order to achieve this, encryption makes use of *secret keys* which are known to the sender and the recipient, but not to the enemy.

The following is a simple encryption scheme for encrypting messages of length n:

> Generate a random bit-string s of length n. (Random here means each bit is 0 with probability 1/2 and 1 with probability 1/2. Each bit is independent of every other bit.) The encryption of a message m is $(s \text{ XOR } m)$.

In the above scheme, s is the secret key. If the recipient knows s, then when they receives an encrypted message they can calculate the original message.

Question 14
What is the probability that the first bit of an encrypted message will be 0? Does this probability depend on the original message?

Ariadne now embeds the encrypted version of her message in her photograph.

Question 15
How might you distinguish an unmodified photograph from one containing an embedded message?

(Hint: Consider the histogram, with intensities between 0 and 255 on the horizontal axis, showing how many pixels in the image have each intensity.)