Assignment 1: Steganographic Techniques

Instructor: Deepa Kundur

- Please note that this assignment does not have to be handed in. It is merely assigned as an exercise to elucidate the information presented in the lectures.
- Solutions will be provided towards the end of the course to aid in studying for the final exam.

1. It was shown in the lecture that the (15,11) linear block code is capable of correcting one bit per codeword. However, the algorithm discussed in class to embed only one bit of information per codeword is highly inefficient.

   (a) Design a different algorithm which makes use of the same (15,11) code to maximize the data hiding rate. *Hint: use the property that the recoverable bit change in the codeword can occur at any location in the 15-bit codeword sequence.*

   (b) What is the data hiding rate of this scheme? . . . code rate?

   (c) If we applied this (15,11) code to hide information in a data file which is originally 11-kbytes, how many kbytes would the encoded file containing the hidden information be? Assume no other encoding or compression takes place. How many bytes of information can be hidden?

2. Consider the \((n, 1)\) repetition code discussed in class. This code is capable of correcting up to \(\lfloor \frac{n-1}{2} \rfloor\) errors where \(\lfloor \cdot \rfloor\) is the floor operator.

   (a) If \(n = 5\), and we consider the (5,1) code:
      i. What is the maximum number of bits you can hide per codeword?
      ii. What is the maximum data hiding rate?
      iii. What is the code rate?

   (b) Repeat i., ii., and iii. of (a) for the (7,1) repetition code.

   (c) Repeat i., ii., and iii. of (a), for the \((n, 1)\) repetition code where \(n\) is odd.

   (d) As \(n\) is increased what can you say happens to the data hiding rate? . . . the code rate?

   (e) For the \((n, 1)\) code, how many kbytes would it take to encode and store the original 11-kbyte file discussed in Problem 1? If the 11-kbyte innocuous data was actually a saved e-mail message, do you think that the storage requirements of the encoded information would seem odd for large \(n\)? For the repetition code, discuss in a paragraph or two the practical trade-offs between data hiding and code rates highlighting how these rates effect the objectives of a data hiding system as discussed in the lectures.
3. Consider an \((n, k)\) linear block code which corrects up to and including \(m\) errors. Note that such codes are also useful when up to \(m\) bits are missing or unknown. In such a case, the unknown bit values can all be set arbitrarily to zero. If some of the missing bits are in fact ones, then our “guess” will cause at most \(m\) errors which can be corrected by the error-correction properties of the code.

(a) Explain in appropriate detail how the above \((n, k)\) linear block code can be used in a secret sharing system by dividing a secret into \(n\) parts. Discuss the basic steps involved in such a secret sharing system. At least how many parts need to be present to determine the secret?

(b) Does this system of secret sharing necessarily exhibit perfect secrecy? Why or why not? Your answer should refer to the definition of perfect secrecy. Simple probability-based equations to justify your reasoning may help answer the question.

4. Consider a secret sharing scheme to hide information concerning the mass of substances \(A\) and \(B\) that should be added to 1 kg of \(H_2O\) to produce a highly desired “potion”. Let \(x\) and \(y\) be the mass of substances \(A\) and \(B\), respectively.

The geometric secret sharing scheme using lines is employed so that at least 2 part holders are required to figure out the secret. For efficiency, however, both of the \(x\)- and \(y\)-coordinates of the plane are used to store the values of \(x\) and \(y\), respectively (recall that in contrast we discussed in class the situation where the \(x\)-coordinate contains the secret and the associated \(y\)-coordinate is randomly generated). The related \((m_j, b_j)\) values are distributed as usual to all the part holders.

In the design of the secret sharing system it was overlooked, however, that the masses of substances \(A\) and \(B\) are known to be related such that \(x \approx 2y\).

(a) Given this known relationship between \(x\) and \(y\), is the above secret sharing system perfectly secure? Explain why or why not.

(b) Design a method which is perfectly secure to handle the above problem. Outline and discuss the basic steps involved in your approach. Explain why your approach is better and justify why it establishes perfect secrecy. You may want to use simple probability-based equations to justify your reasoning.

5. There are three professors, seven graduate students and eighty-seven undergraduate students all involved in a top secret project. To work on the project a secret key is required to access the relevant computer files.

Four examples of secret sharing schemes were proposed in class: encryption, modulo 2 addition, simple geometry, and polynomials in finite fields. Using this tool-set, design a secret sharing scheme to limit access to the key for each of the following situations.

Accessing the key requires:

(a) at least two professors.
(b) at least one professor and all seven graduate students.
(c) at least one professor and at least four graduate students and at least eleven undergraduate students.
(d) at least eighty-nine members of the project (profs, grads and/or undergrads are not distinguished).
(e) can you implement (d) in another way based on the techniques discussed in class?

For parts (d) and (e), what is the most efficient technique to use for secret sharing in terms of storage requirements for each part holder?