2.8 Summary

In this chapter we covered several classic cryptosystems, including the simple substitution, the double transposition, codebooks, and the one-time pad. Each of these illustrates some important points that we'll return to again in later chapters. We also discussed some elementary aspects of cryptography and cryptanalysis.

In the next chapter we'll turn our attention to modern symmetric key ciphers. Subsequent chapters cover public key cryptography, and hash functions. Cryptography will appear again in later parts of the book. In particular, crypto is a crucial ingredient in security protocols. Contrary to some authors' misguided efforts, the fact is that there's no avoiding cryptography in information security.

2.9 Problems

1. In the field of information security, Kerckhoffs' principle is like motherhood and apple pie, all rolled up into one.
   a) Define Kerckhoffs' principle in the context of cryptography.
   b) Give a real-world example where Kerckhoffs' principle has been violated. Did this cause any security problems?
   c) Kerckhoffs' principle is sometimes applied more broadly than its strict cryptographic definition. Give a definition of Kerckhoffs' principle that could apply more generally.
2. Edgar Allan Poe's 1843 short story, "The Gold Bug," features a cryptanalytic attack.
   a) What type of cipher is broken and how?
   b) What happens as a result of this cryptanalytic success?
3. Given that the Caesar's cipher was used, find the plaintext that corresponds to the ciphertext

   ```
   VSRQJHEREVTXDUHSDQWV.
   ```

4. Find the plaintext and the key, given the ciphertext

   ```
   CSYEVIXIVQMREXIH.
   ```

   Hint: The message was encrypted with a simple substitution, where the key is a shift of the alphabet.
5. Suppose that we have a computer that can test 2^{40} keys each second.
   a) What is the expected time (in years) to find a key by exhaustive search if the keyspace is of size 2^{88}?
   b) What is the expected time (in years) to find a key by exhaustive search if the keyspace is of size 2^{112}?

c) What is the expected time (in years) to find a key by exhaustive search if the keyspace is of size 2^{256}?

6. The weak ciphers used during the election of 1876 employed a fixed permutation of the words for a given length sentence. To see that this is weak, find the permutation of $(1, 2, 3, \ldots, 10)$ that was used to produce the scrambled sentences below, where "San Francisco" is treated as a single word:

   first try try if you and don't again at succeed
   only you you you as believe old are are as
   winter was in the I summer ever San Francisco coldest spent

   Note that the same permutation was used for all three sentences, i.e., the three sentences are in depth.

7. This problem deals with the concepts of confusion and diffusion.
   a) Define "confusion" and "diffusion" as used in cryptography.
   b) Which classic cipher discussed in this chapter employs only confusion?
   c) Which classic cipher discussed in this chapter employs only diffusion?
   d) Which cipher discussed in this chapter employs both confusion and diffusion?

8. Recover the plaintext and key for the simple substitution example that appears in (2.2) on page 20.

9. Determine the plaintext and key for the ciphertext that appears in the quote at the beginning of this chapter. Hint: The message was encrypted with a simple substitution cipher and the plaintext contains no spaces or punctuation.

10. Decrypt the following message, which was encrypted using a simple substitution cipher:

```
GBSXUCGSZQGKGSQPKQKGLSKASPCGBGBKGUKGCEUKUZKGGBSQEICA
CGKGCEUERWKLKUPKQQGCIICUAEUVSHQKGCEUPCGBCGQOEVSHUNSU
GKUZCGQSNLSHEHIEEDCUOGEPKHZGBSNKCUGSUKUASERLSKASCUGB
SLKACRCACUZSSZEUSBEXHKRGSHWKLKUSQSKCHQTXKZHEUQBKZAEN
NSUASZFENFCUOCUEKBXGBSWKLKUSQSKNFKQQKZEHGEGBSXUCGSZQ
GKGSQKUZBCQAEIISKOXSZSICVSHSZGEGBSQSAHSGKHMERQGKGSKR
EHNKIHSLIMGEKHSASUGKNSHCAKUNSQQKOSPBCISGBCQHSLIMQGKG
SZGBKGCGQSSNSZXQSISQQGEAEUGCUXSGBSSJCQGCUOZCLIENKGCA
USOEGCKGCEUQCGAEUGKCUSZUEGBHSKGEHBCUGERPKHEHKHNSZKGGKAD
```

11. Write a program to help an analyst decrypt a simple substitution cipher. Your program should accept the ciphertext as input, compute letter

frequency counts, and display these for the analyst. Your program should then allow the analyst to guess a key and display the results of the putative decryption using the specified putative key. Of course, you may add other features to your program that you consider useful. Use your program to help solve Problem 10, and comment on the usefulness of your program, as compared to working only with pencil and paper.

12. Extend the program described in Problem 11 so that it includes the following features:
    i) Make an initial decryption of the message. The recommended way to proceed is to use monograph (i.e., individual letter) frequencies to make an initial guess for the key. Call this the "best key."
    ii) Use digraph frequencies to compute a score for any putative key.
    iii) Generate new putative keys by swapping each pair of letters in the best key—if the score from ii) improves for a given swap, update the best key; if not, leave the best key unchanged.
    iv) Iterate the process in iii) until the score does not improve for an entire pass through the key (i.e., all pairs have been swapped). The best key is your putative solution.

    Some errors in the key will likely remain, so your program must also include all of the functionality of the program in Problem 11. Use your program to solve Problem 10 and give the fraction of the key that is correctly recovered automatically, and the fraction of plaintext letters that are determined correctly.
13. Jakobsen's algorithm [59] is an extremely efficient and effective simple substitution solver. Implement Jakobsen's algorithm and test your program on 10 distinct simple substitution ciphertext messages of each of the lengths $L \in \{100, 200, 300, \ldots, 1000\}$, that is, 10 messages of length $L = 100$, 10 messages of length $L = 200$, and so on. On the same axes, graph the average fraction of the key that is correctly recovered, and the average fraction of plaintext letters that are correctly determined for each of these lengths.
14. Decrypt the following ciphertext:

```
IAUTMOCSMNIMREBOTNELSTRHEREOAEVMWIH
TSEEATMAEOHWHSYCEELTTEOHMUOUFEHTRFT
```

    This message was encrypted with a double transposition (of the type discussed in this chapter) using a matrix of 7 rows and 10 columns. Hint: The first word is "`there`."
15. Outline an automated attack on a double transposition cipher (of the type discussed in the text), assuming that the size of the matrix is known.

16. A double transposition cipher can be made much stronger by using the following approach. First, the plaintext is put into an $n \times m$ array, as described in the text. Next, permute the columns, and then write out the intermediate ciphertext column by column. That is, column 1 gives the first n ciphertext letters, column 2 gives the next n, and so on. Then repeat the process, that is, put the intermediate ciphertext into an $n \times m$ array, permute the columns, and write out the ciphertext column by column. Use this approach, with a 3×4 array, and permutations $(2, 3, 4, 1)$ and $(4, 2, 1, 3)$ to encrypt the plaintext `attackatdawn`.
17. Using the letter encodings in Table 2.1, the two ciphertext messages

    `KHHLTK` and `KTHLLE`

    were encrypted with the same one-time pad. Find all dictionary words that are possible plaintext pairs and in each case, give the corresponding one-time pad.
18. Using the letter encodings in Table 2.1, the following ciphertext message was encrypted with a one-time pad:

    `KITLKE`.

    a) If the plaintext is "thrill," what is the key?
    b) If the plaintext is "tiller," what is the key?
19. Suppose that the following is an excerpt from the decryption codebook for a classic codebook cipher:

123	once
199	or
202	maybe
221	twice
233	time
332	upon
451	a

    Decrypt the ciphertext

    $$242,\ 554,\ 650,\ 464,\ 532,\ 749,\ 567$$

    assuming that the following additive sequence

    $$119,\ 222,\ 199,\ 231,\ 333,\ 547,\ 346$$

    was used to encrypt the message.

20. An affine cipher is a type of substitution where each letter is encrypted according to the rule $c = (a \cdot p + b) \pmod{26}$ (see the Appendix for a discussion of the mod operation). Here, p, c, a, and b are each numbers in the range 0 to 25, where p represents the plaintext letter, c the ciphertext letter, and a and b are constants. For the plaintext and ciphertext, the number 0 corresponds to "`a`," 1 corresponds to "`b`," and so on. Consider the ciphertext `QJKES REOGH GXXRE OXEO`, which was generated using an affine cipher. Determine the constants a and b and decipher the message. Hint: Plaintext "t" encrypts to ciphertext "`H`" and plaintext "`o`" encrypts to ciphertext "`E`."
21. A Vigenère cipher uses a sequence of shift-by-n simple substitutions, where the shifts are indexed using a keyword, with "A" representing a shift-by-0, "B" representing a shift-by-1, etc. For example, if the keyword is "DOG," then the first letter is encrypted using a simple substitution with a shift-by-3, the second letter is encrypted using a shift-by-14, the third letter is encrypted using a shift-by-6, and the pattern is repeated—the fourth letter is encrypted using a shift-by-3, the fifth letter is encrypted using a shift-by-14, and so on. Cryptanalyze the following ciphertext, i.e., determine the plaintext and the key:

    ```
    CTMYR DOIBS RESRR RIJYR EBYLD IYMLC CYQXS RRMLQ FSDXF
    OWFKT CYJRR IQZSM X
    ```

    This particular message was encrypted using a Vigenère cipher with a 3-letter English keyword:
22. Suppose that on the planet Binary, the written language uses an alphabet that contains only two letters X and Y. Also, suppose that in the Binarian language, the letter X occurs 75% of the time, while Y occurs 25% of the time. Assume that you have two messages in the Binary language, and the messages are of equal length.
    a) If you compare the corresponding letters of the two messages, what fraction of the time will the letters match?
    b) Suppose that one of the two messages is encrypted with a simple substitution, where X is encrypted as Y and Y is encrypted as X. If you now compare the corresponding letters of the two messages—one encrypted and one not—what fraction of the time will the letters match?
    c) Suppose that both of the messages are encrypted with a simple substitution, where X is encrypted as Y and Y is encrypted as X. If you now compare the corresponding letters of the two messages—both of which are encrypted with the same key—what fraction of the time will the letters match?

d) Suppose instead that you are given two randomly generated sequences consisting of the two letters X and Y. If you compare the corresponding letters of the two messages, what fraction of the time will the letters match?

e) Briefly describe the index of coincidence (IC), as described, for example, in [42].

f) How can the index of coincidence be used to determine the length of the keyword in a Vigenère cipher (see Problem 21 for the definition of a Vigenère cipher)?

23. In this chapter, we discussed a forward search attack on a public key cryptosystem.

a) Explain how to conduct a forward search attack.

b) How can you prevent a forward search attack against a public key cryptosystem?

c) Why can't a forward search attack be used to break a symmetric cipher?

24. Consider a "one-way" function h, that is, a function where given the value $y = h(x)$, it is computationally infeasible to find x directly from y.

a) Suppose that Alice computes $y = h(x)$, where x is Alice's salary, in dollars. If Trudy obtains y, how can she determine Alice's salary x? Hint: Adapt the forward search attack to this problem.

b) Why does your attack in part a) not violate the one-way property of h?

c) How could Alice prevent this attack? We assume that Trudy has access to the output of the function h, Trudy knows that the input includes Alice's salary, and Trudy knows the format of the input. Also, no keys are available, so Alice cannot encrypt the output value.

25. Suppose that a particular cipher uses a 40-bit key, and the cipher is secure, i.e., there is no known shortcut attack.

a) How much work, on average, is an exhaustive search attack?

b) Outline an attack, assuming that known plaintext is available.

c) How would you attack this cipher in the ciphertext-only case?