Review of[1]
Data Privacy and Security
Author: David Salomon
Publisher: Springer–Verlag, 2003
$51.48, Hardcover

Reviewer: Nick Papanikolaou
(Dept. of Computer Science,
University of Warwick, U.K.)

1 Introduction

The field of cryptology and data security hardly needs any introduction; numerous popular accounts of the subject have appeared over the years, and it is already a core topic in undergraduate computer science. The very term "cryptology" is testimony to the long history of the field; the term is derived from the words $\kappa\rho\upsilon\pi\tau\acute{o}\varsigma$ (meaning hidden), and $\lambda\acute{o}\gamma o\varsigma$ (meaning speech), which have retained their meaning in the Greek language for many centuries.

Cryptology is the study of codes and ciphers, mechanisms through which data can be transformed so as to make their content unreadable to anyone but specially authorized persons. It is traditionally divided into *cryptography,* the development of new codes and ciphers, and *cryptanalysis,* the art of subverting existing ones. While cryptanalysis is regarded as a sort of 'black magic' that has always required special skill (in Alan Turing's days) or extremely fast computers (today), the study of cryptography is of interest to everyone and is becoming increasingly accessible to a wider public. Formally, the purpose of cryptography is to accomplish one or more of the following objectives:

- *confidentiality,* or secrecy of given data;
- *integrity,* or assurance that data has not been tampered with;
- *non–repudiation,* or definitive proof that data was exchanged between parties;
- *authentication,* or proof of the origin of given data.

David Salomon's recent textbook [6], ventures to survey classical cryptography and steganography in an accessible manner. While there already exist several volumes covering these topics e.g. [7, 8, 9], Salomon's book is a practical and readable reference that has much to commend it. Interestingly, it is one of few books to treat steganography on a par with classical cryptographic techniques. This review contains a brief summary of the book's chapters.

2 Coverage

Salomon starts with a fascinating account of the Zimmermann telegram, which famously contained a plot to discourage the United States from entering the First World War. The telegram was decrypted in England and this, as the author points out, changed the course of history. This highlights the significance of cryptanalysis, and is followed in the book by a definition of all the relevant terminology. Some basic terms of interest are:

Code: A code is a direct transformation of a word, a phrase, or even an entire message.

Cipher: A cipher is a transformation defined over each symbol in a particular *alphabet.*

Nomenclator: A nomenclator is a combination of code and cipher.

Encryption: Encryption is the process of applying a code or cipher to a message, called the *plaintext.*

Decryption: Decryption is the process of recovering the plaintext from an encrypted message, known as the *ciphertext.*

The book's introduction goes on to discuss some simple ciphers, including the *Caesar cipher* and the *one–time pad.* The Caesar cipher is no more than letter shifting; a message is encrypted by replacing each letter with the letter n positions ahead of it. A fashionable version of this cipher is known as **ROT13**, and has $n = 13$, i.e. half the length of the English alphabet. For example, **ROT13** transforms the message `SIGACT NEWS` into `FVTNPG ARJF`.

The one–time pad is a rare example of a *perfect cryptosystem;* a perfect, or unconditionally secure cryptosystem, cannot be broken even if the enemy has unlimited time and computational power. To encrypt a message m with the one–time pad, one must generate a key, k, which is at least as long as m. The same key must be used to encrypt and decrypt the message; the ciphertext is the exclusive-or of k and m. As long as a different key is used for every message, this system provides perfect secrecy — in other words, an enemy cannot obtain any information about the key given only the ciphertext.

The one–time pad cryptosystem suffers from the need to distribute the key to all legitimate receivers of a message; the key itself must be exchanged in secrecy. This so–called *key distribution problem* is addressed in public–key cryptography, and also by using quantum key distribution techniques, described later.

2.1 Chapters 1–3: Substitution and Transposition Ciphers

The first three chapters of the book deal with all the traditional ciphers. In a *substitution cipher,* each letter in a message is replaced with another letter from one or more alphabets. When all letters are drawn from a single alphabet, we have a *monoalphabetic* substitution cipher; these are covered in Chapter 1. When the letters in a message are replaced with letters from several alphabets, we have a *polyalphabetic* substitution cipher, and these are discussed in Chapter 3. *Transposition ciphers* replace a message by a permutation of itself, as explained in Chapter 2 of Salomon's book.

The ciphers discussed in **Chapter 1** include Polybius' cipher, the Playfair cipher, fractionation, and homophonic ciphers. Of these, we will discuss only the Playfair cipher here. In the Playfair cipher, all messages are encrypted with the help of a 5×5 square, which contains all the letters of the alphabet except `J`, which is rarely used in messages anyway. The square serves as an encryption key, and is constructed by choosing a long word with relatively few or no repeating letters. The unique letters of this word are placed, in sequence, into the square, followed by all the other letters in the alphabet. Take, for example, the word `COMPUTATION`; the corresponding square becomes:

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

Now, suppose we wish to encrypt the plaintext `FOLLOW ME EARLY`. To do this, we divide the plaintext into pairs of letters (we remove duplicate letters, and pad out with an `X`): `FO`, `LO`, `WM`, `EA`, `RL`, `YX`. Then, for each pair of letters (x, y), we locate x and y in the square and draw the rectangle which has x and y as opposite corners. Next x and y are replaced by the letters in the other two corners of this rectangle. When x and y do not form a rectangle, they are replaced with the letters immediately below (that's when x and y are in the same column) or immediately to the right (that's when x and y are in the same row). Using these rules, we obtain for the pairs in our example plaintext: `EM`, `WA`, `XO`, `WL`, `SQ`, `VZ`. If you would like to know why `YX` maps to `VZ`, and what happens in more complicated cases, you should buy the book!

Chapter 2 deals with transposition ciphers: the turning template, the columnar transposition cipher and the variations due to Myzkowsky and Scott. The author explains how such ciphers can be decrypted, as he does also in Chapter 1. Breaking both substitution and transposition ciphers is actually done by taking advantage of the relative frequency of letters in the English alphabet. Did you know that the probability of finding an 'E' in Shakespeare's plays is 0.1196? It should be added that Matlab code for some of these ciphers is provided in the text.

Chapter 3 discusses a great variety of ciphers, including those due to Beaufort, Trithemius, Vigenère, Gronsfeld, Eyraud, Hill and Jefferson. Let's consider the Hill cipher briefly. In the Hill cipher, all the letters in the alphabet are numbered 0 to 25; a number $n < 26$ is chosen, and the key is formed by generating a $n \times n$ matrix K whose elements are integers in the range 0 to 25. The first n letters of a given plaintext are converted to a column vector, P, and the ciphertext is obtained by computing

$$C = K \cdot P \mod m$$

Decryption in the Hill cipher consists of computing a matrix inverse modulo an integer; in particular, $P = K^{-1} \cdot C \mod m$. Unfortunately, the Hill cipher has limited power and can be subverted using a so–called *chosen–plaintext attack*, in which the enemy knows a finite number of ciphertexts, C_i, and their corresponding plaintexts, P_i.

2.2 Chapters 4 and 5: Random Numbers and The Enigma

Chapter 4 of *Data Privacy and Security* is devoted to random numbers, which are of primordial importance in cryptography. In practice, pseudorandom number generators are used, and algorithms for this purpose are discussed in the text. Also, the author describes the main statistical tests that can be performed to gauge the degree of randomness of a given number sequence. These various topics are covered *in extensis* in Don Knuth's seminal work [2], which is aimed at a more advanced reader.

One of the most attractive features of this book is the material in **Chapter 5,** which details the Enigma machine, used by the Germans in World War II. The historical background is discussed, and the workings of the machine are carefully explained by means of numerous diagrams and pictures.

2.3 Chapters 6–8: Stream Ciphers, Block Ciphers, Public Key Cryptography

Chapters 6–8 of the book cover the more fashionable aspects of cryptography, though no differently from most of the other books [7, 8, 9] on the subject. *Stream ciphers* (Chapter 6) and *block ciphers* (Chapter 7) are the two principal classes of cipher used on modern–day computers. Since all messages are now ultimately reduced to strings of zeros and ones, secure ciphers have to be based on the manipulation of bits. Stream ciphers encrypt a string of bits by treating each bit individually, while block ciphers divide a bit string into blocks and transform each block.

Chapter 6 points out the distinction between symmetric–key and public–key cryptosystems. Symmetric key cryptosystems use the same key for encryption and decryption, while public–key systems use two distinct keys. Linear and nonlinear shift registers are discussed, along with cellular automata, SEAL and the RC4 cipher.

Chapter 7 describes substitution–permutation ciphers, Lucifer and the Data Encryption Standard (DES). This leads on to a presentation of Blowfish, IDEA, RC5 and Rijndael. Rijndael is also known as the Advanced Encryption Standard, or NIST standard FIPS-197. Rijndael involves several rounds and consists of the following operations: byte substitution, row shifting, column mixing, and adding a subkey (or 'round key'). It is not yet known how secure Rijndael is; Salomon states that its security 'can be demonstrated only with time.'

Chapter 8 presents Diffie–Hellman–Merkle key exchange, RSA (the Rivest–Shamir–Adleman public–key cryptosystem), Rabin's system and the El–Gamal scheme. All of these are discussed briefly, with an emphasis on RSA. Threshold schemes and authentication are then covered. Elliptic curve cryptography, now quite *en vogue,* is then described at length. We cannot do justice to the many topics covered, in the framework of this brief review; let us at least reproduce, from page 200 of the book, a two–line implementation of RSA in Perl:

```
print pack"C*",split/\D+/,‘echo "16iII*o\U@{$/=$z;[(pop,pop,unpack"H*",<>
)]}\EsMsKsN0[lN*1lK[d2%Sa2/d0<X+d*lMLa^*lN%0]dsXx++lMlN/dsM0<J]dsJxp"|dc‘
```

2.4 Chapter 9: Quantum Cryptography

Of the cryptographic techniques described in this book, none is more exciting than *quantum cryptography,* which relies for its security on the laws of quantum physics; the BB84 protocol for *quantum key distribution* is presented in **Chapter 9**. It has been proven that this protocol is unconditionally secure against all possible attacks and therefore solves, at least in principle, the age–old problem of key distribution. Combined with an unconditionally secure cryptosystem, such as the one–time pad, quantum key distribution paves the way for truly unbreakable cryptography.

While public–key cryptosystems, such as RSA, resolve the problem of distributing keys in a mathematical way, their security remains largely dependent on the complexity of certain computational problems, such as prime factoring, which can be performed efficiently on a quantum computer. Quantum computers are still mostly objects of theoretical speculation, but small–scale ones have been built in experimental physics labs. Peter Shor famously devised efficient quantum algorithms for factoring and the discrete logarithm; a full–scale quantum computer could run these algorithms and efficiently break several cryptosystems in current use. The security of quantum cryptography, or more specifically, quantum key distribution, is not threatened by the computational power of quantum computers.

Gilles Brassard ran a 'Cryptology' column in *SIGACT News* for years; before I describe quantum cryptography in more detail, let me quote his words on the tie between this newsletter and the subject:

> "The fates of SIGACT News and Quantum Cryptography are inseparably entangled. The exact date of Stephen Wiesner's invention of 'conjugate coding' is unknown but it cannot be far from April 1969, when the premier issue of SIGACT News [...] came out. Much later, it was in SIGACT News that Wiesner's paper finally appeared [*Vol. 15, No. 1, 1983*] in the wake of the first author's [*Gilles Brassard's*] early collaboration with Charles Bennett [...]. It was also in SIGACT News that the original experimental demonstration for quantum key distribution was announced for the first time [*Vol. 20, No.4, 1989*] and that a thorough bibliography was published [*Vol. 24, No. 3, 1993*]. Finally, it was in SIGACT News that Doug Wiedemann chose to publish his discovery when he reinvented quantum key distribution in 1987, unaware of all previous work but Wiesner's [*Vol. 18, No. 2, 1987*].

Quantum key distribution protocol BB84 allows two parties, 'Alice' and 'Bob,' to establish a secret binary key. The idea is to represent each bit in a given string by either a rectilinearly or diagonally polarized photon. In the *rectilinear basis,* a `0` is represented by a photon polarized at 0°, and a `1` by a 90°–polarized photon. In the *diagonal basis,* a 45°–polarized photon stands for `0` and a 135°–polarized photon stands for `1`. In brief, the protocol proceeds as follows. Alice generates a random bit string, and a random string of encoding bases. Each bit is mapped to a polarized photon using the corresponding basis and then transmitted to Bob. Bob does not know which basis has been used to encode each photon, and so chooses one of the two bases at random in order to make a measurement. Due to the laws of quantum physics, the result of Bob's measurement is only guaranteed to be correct for a given photon if his choice of measurement basis matches the one used by Alice for encoding. When the entire binary string has been transmitted in this way, Bob will have obtained a subset of Alice's bit string. Using a classical —possibly even public— communications channel, Alice tells Bob which of his basis choices were correct.

There is much more to quantum cryptography than we can say here, and the interested reader is referred to [4] for a recent introduction to the subject. It should be added that commercial quantum cryptographic devices do exist already (see `http://www.magiqtech.com` and `http://www.idquantique.com`). Quantum protocols such as BB84 are of special interest to computer scientists today; for example, [3] discusses modelling and analysing the security of these schemes using automated tools.

Salomon's presentation of quantum cryptography in *Data Privacy and Security* is very readable, and is innovative in the sense that not many crypto textbooks deal with this subject. On a lighter note, Schneier [7] exclaims: "this would still be on the lunatic fringe of cryptography, but Bennett and Brassard actually went and built a working model of the thing [...]."

2.5 Chapters 10–12: Steganography

The three final chapters of the book are devoted to *steganography,* or data hiding. The goal of steganography is to hide a message in another item of data, known as the *cover.* If the original message is to be embedded in a text, referred to as *covertext,* the product of the steganographic procedure is termed a 'stegotext.' Messages can also be embedded in images, sounds and video, leading respectively to the use of the terms 'coverimage' and 'stegoimage,' 'coveraudio' and 'stegoaudio,' 'covervideo' and 'stegovideo.' Word-smithing will never go out of style!

The basic ingredients of a data hiding system are: (1) the data to be embedded; (2) the cover, in which embedding will occur; (3) a stego–key; (4) an embedding algorithm; and (5) a decoder. The embedding algorithm produces a *stego–cover* given the first three items above. To recover the data hidden in it, the stego–cover is fed into a decoder along with the stego–key. **Chapter 10** of the book elaborates on these fundamentals and focuses on data hiding *in text.* The principal characteristics of a data hiding algorithm are identified and explained; these are embedding capacity (how much data can be hidden in a given cover), invisibility (how much distortion is caused to the cover), undetectability (a statistical measure of the distortion), robustness (the degree of immunity of the stego–cover to subsequent alterations, such as compression), tamper resistance (the degree of immunity of the stego–cover to direct tampering) and the signal–to–noise ratio. Watermarking is introduced. As an example of data hiding in text, that merely modifying the spaces in a text file is a means of conveying a message. This particular idea is extended in an interesting manner on page 256:

> "The TEX typesetting software permits very fine control over the interword spaces and the spaces following certain punctuation marks. The smallest dimension that TEX can use is called a scaled point (sp). One inch equals 72.27 printers' points (pt), and one pt equals 65,536 scaled points. Thus, the value of a sp is about the wavelength of visible light, and changing the normal interword space by 1 sp is invisible.

> TEX can also list the precise values of all the components of text [...] Because of these features, TEX may be an ideal tool for hiding data in spaces, although it was originally designed as a high–quality typesetting system for the production of books."

This quotation reminds me of David Salomon's equally commendable TEX volume [5].

Chapter 11 of *Data Privacy and Security* focuses on data hiding in images, both in the spatial domain and the transform domain. Topics covered include LSB encoding, bit–plane complexity segmentation, spread–spectrum steganography and wavelet–based watermarking. Several methods of watermarking are covered, and variations are explained clearly.

The last chapter is concerned with data hiding in audio and video, and goes as far as to describe the remarkable 'steganographic file system' [1]. This file system hides a given set of files f_i in several ordinary files, in such a way that an unauthorized user cannot even determine whether the f_i exist.

The book's appendices deal with important mathematical background associated with the subject, namely convolution, hashing, cyclic redundancy codes and Galois fields. Answers to all the exercises in the text are given, and there is also a 10–page historical timeline of cryptography. This is supplemented by a glossary, a detailed bibliography and a comprehensive index.

3 Opinion and Conclusion

Overall, *Data Privacy and Security* is a handy and concise volume on cryptography and steganography. While the more advanced reader will find alternative references — such as [7] — more satisfying, this book is clearly written and self–contained. All the important cryptosystems are covered, and the discussion of recent developments, including Rijndael and quantum cryptography, is welcome.

References

[1] Anderson, R., Needham, R., and Shamir, A. The steganographic file system. In *2nd Information Hiding Workshop* (1998).

[2] Knuth, D. E. *Seminumerical Algorithms*. Addison-Wesley, 1981.

[3] Papanikolaou, N. Techniques for design and validation of quantum protocols. Master's thesis, Department of Computer Science, University of Warwick, 2005.

[4] Papanikolaou, N. Introduction to quantum cryptography. *ACM Crossroads Magazine 11.3* (Spring 2005 Issue). To appear.

[5] Salomon, D. *The Advanced TeXbook*. Springer Verlag, 1995.

[6] Salomon, D. *Data Privacy and Security*. Springer–Verlag New York, Inc., 2003.

[7] Schneier, B. *Applied Cryptography*, 2nd ed. Wiley, 1996.

[8] Smart, N. *Cryptography: An Introduction*. McGraw–Hill Education (UK), 2003.

[9] Welsh, D. *Codes and Cryptography*. Clarendon Press, 1998.