Encryption for Organizations and Individuals

Basics of Contemporary and Quantum Cryptography

Robert Ciesla

Apress®

Encryption for Organizations and Individuals

Basics of Contemporary and Quantum Cryptography

Robert Ciesla

Apress®

Encryption for Organizations and Individuals: Basics of Contemporary and Quantum Cryptography

Robert Ciesla
HELSINKI, Finland

ISBN-13 (pbk): 978-1-4842-6055-5 ISBN-13 (electronic): 978-1-4842-6056-2
https://doi.org/10.1007/978-1-4842-6056-2

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: Rita Fernando
Coordinating Editor: Divya Modi

Cover designed by eStudioCalamar

Cover image designed by Pixabay

Author photo © 2018 by A.C.

Distributed to the book trade worldwide by Springer Science Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-6055-5. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

Dedicated to curious laypeople everywhere.

Table of Contents

About the Author

Robert Ciesla is a freelance writer from Helsinki, Finland. He has worked on many video games on several platforms. He is the author of *Game Development with Ren'Py* (2019) and *Mostly Codeless Game Development* (2017). Ever since finishing *A Brief History of Time* by Stephen Hawking in middle school, Robert has been fascinated by the world of quantum mechanics. Robert's bachelor's thesis in journalism took on some questions on how to popularize the core concepts of quantum physics and related fields. He has devoured most relevant books in the field since and continues to explore this area of reality.

About the Technical Reviewers

Paul Love is the Chief Information Security and Privacy Officer at a financial services organization and has been in the information security field for almost 30 years. He has held information security positions at many major organizations including Federal Home Loan Mortgage Corporation (Freddie Mac), Ernst & Young, Microsoft, Schlumberger, Ally Financial, and Fifth Third Bank. Paul started his information security career when he joined the United States Marine Corps, where he served for eight years, eventually achieving the rank of Sergeant.

Paul holds a Master of Science in Network Security, has authored/co-authored nine books on Information Security and Unix/Linux, and has been the technical editor of ten books on Linux and Unix. Paul holds multiple information security and privacy certifications, including Certified Information Systems Security Professional (CISSP), Certified Information System Auditor (CISA), and Certified Information Security Privacy Manager (CISM); multiple privacy certifications including Certified Information Privacy Professional/United States/Europe/Canada (CIPP/US, CIPP/E, CIPP/C); as well as other technical and professional certifications.

Sai Matam is a software architect with over 20 years of diverse experience in software. He has a Bachelor of Engineering from Osmania University. His interests include low latency, highly scalable systems, algorithms, Java, cloud, and Go programming language.

Astha Keshariya, PhD, MSC (Honors), MBA, has consulted and contributed to several commercial and academic organizations in the field of applied cryptography and information security for over 16 years.

Introduction

Cryptography may or may not sound like the sexiest of topics. However, it's essential to nearly everyone plugged into the planetwide community of the Internet. Whether you're working for Area 52 with a Top Secret clearance or shopping online for some swanky items, many elements of cryptography will be present. They not only take the form of virtually unbreakable databases but also (barely noticeable) digital certificates, passwords, PIN codes, and secured email.

As impressive as current-day cryptography is in its security and computational effectiveness, what's behind the corner is even more so. Quantum computing is well on its way. We can expect our world to be profoundly impacted by this paradigm on several levels.

Ultimately, encryption and secrecy are not new phenomena. They have been with us since the earliest days of recorded history, only in more primitive ways. The continuing need for concealing information tells us something about the world at large. I hope this book offers you an understanding of just how big of a deal cryptography actually is.

Encryption for Organizations and Individuals is for the curious layperson. Equations are therefore kept to a minimum. This book is roughly divided into two parts: first, we explore contemporary cryptography, and then we probe into its quantum sibling. I hope my book equips you with the tools you need to take on the quantum computing revolution with some confidence.

CHAPTER 1

The First Era of Digital Encryption

You're probably used to entering passwords into devices by now; it's a part of everyday life, like locking and unlocking one's front door. From email services to mobile devices, we all guard our privacy to a varying extent in the digital realm. And that's exactly how it should be. In this chapter, we'll take a quick look at modern-era digital encryption. But first, we'll revisit some of the most game-changing moments in the historical context of all things cryptographic, as you may not be familiar with the incredibly long history of the science.

Classical Cryptography

Let's first define our main term. The word *cryptography* refers to the science of transmitting messages which remain undecipherable to often malicious third parties. It comes from the ancient Greek words of *kryptos*, which stands for hidden, and *graphein*, which means "to write." Cryptography is valued by warring tribes, governments, and individuals alike; as long as there remains the need for any kind of political action or activism, cryptography will continue to thrive.

R. Ciesla, *Encryption for Organizations and Individuals*,
https://doi.org/10.1007/978-1-4842-6056-2_1

There are two other terms of relevance you should become familiar with at this point: *plaintext* and *ciphertext*. The former refers simply to an unencrypted message (e.g., "Hello! Apress is the best publisher!"), while the latter covers encrypted messages, which appear nonsensical to those not in possession of the decryption key(s).

Now, the first recorded instance of hidden messages dates back to ancient Egypt, 1900 BC. A series of nonstandard hieroglyphs (i.e., characters in the Egyptian writing system) were discovered carved into the walls of a tomb. Experts still argue whether these messages contain any pertinent information or not; they may have been created with the intention to amuse or confuse.

Clay tablets from Mesopotamia (its area corresponding with most of modern Iraq, Kuwait, and some parts of Syria) indicate attempts at concealing more "serious" information around 1500 BC. Many of the tablets were found to be encrypted cooking notes. These are clearly important state secrets and should never fall in the wrong hands: empires have been known to collapse for less!

The mighty Romans of ancient times, too, were known to utilize cryptography, creating a device called *Caesar's cipher*. It simply involves shifting the alphabet to a degree as agreed upon by two parties (e.g., using a right shift of two letters so that A becomes C and C becomes E). Although hardly representing the state of the art in encryption in 2020, many a private communique was dispatched between Julius Caesar (100 BC–44 BC) and his allies using this technique. It didn't hurt, of course, that most of his enemies were illiterate.

In medieval times, the state of the art in cryptography was to be found among the Arab people. A grammarian from Basra, Iraq, *Al-Khalil (717–786 AD)*, wrote a seminal work on hidden messages, entitled *The Book of Cryptographic Messages*. His book is famous for its use of permutations and combinations to list all possible Arabic words with and without vowels.

Al-Khalil's work inspired another monumental book in the field, *The Manuscript for the Deciphering of Cryptographic Messages* written by one *Al-Kindi (801–873 AD)*, a mathematician and astronomer from Kufa, Iraq. His work, released around the year 800, detailed most likely for the first time ever the concept of *frequency analysis*, which is still an important concept in cryptography. We will learn more of the basics of Al-Kindi's work in the next section.

The Basics of Frequency Analysis

Frequency analysis is the study of letters contained in an encrypted message in order to reveal at least parts of the plaintext message. The rest should be subject to common sense and basic grammar. Now, most languages have certain letters appearing at a specific frequency. For example, in the English language, the most common letters are E, T, and A. In contrast, Q, X, and Z are not found in English sentences very often. In a historical context, the inventor of Morse code, *Samuel Morse (1791–1872)*, did his part to discover which letters of the alphabet are the most common in English in order to assign to them the most simple codes.

Let's assume we are to decrypt a message which, we're told, only contains a short English sentence. Knowing this, we may statistically determine some parts of the message and deduce the rest, if we're lucky. The first step is count the times a letter appears in an encrypted message. Now, take a look at the ciphertext we are to decrypt:

```
KZ GZK ZKGR KKR
```

Which is the most frequent letter in the example? That would be K with five occurrences. The most common letter was E, right? Changing the K's to E's results in the following:

```
EZ GZE ZEGR EER
```

Not much help you may think. However, let's keep at it. The second most frequent letter here is Z with three occurrences. As for the English language, the second used letter is T. Let's go with that.

ET GTE TEGR EER

The third most frequent letters here are G and R, both with two occurrences of each. As for the English language, the third used letter is usually A, I, N, O, or S. Let's go with S first and replace the message's G's with it.

ET STE TESR EER

Our intuition speaks: that can't be right. After careful consideration (and possibly trying all other statistically significant choices of O, N, and I, which got us nowhere), we decided to replace the G's with A's instead.

ET ATE TEAR EER

Now we see something vaguely resembling English. Let's use our incredible powers of deduction and take a wild guess. What if R equals L?

ET ATE TEAL EEL

Finally, some proper English. Oh, those pesky extraterrestrials and their hunger for our majestic (and, let's face it, delicious) Anguilliformes! This has been a simple demonstration of frequency analysis. Using a combination of statistical evaluation and grammatical sense, especially with intelligence concerning the message's language issued beforehand, one may be able to decrypt some of the simpler ciphertexts, all on paper.

Mind you, we could've also deciphered the message even more easily using Caesar's cipher mentioned earlier in the chapter. By switching the alphabet six times to the right (i.e., A equals G, C equals I), we would've achieved the same result.

It should be noted that an Egyptian mathematician *Al-Qalqashandi (1355–1418)* first described the *polyalphabetic system* which greatly undermined the effectiveness of classical frequency analysis. The polyalphabetic system refers to the method of using multiple letters/symbols per alphabet in a plaintext message to cause further confusion during the decryption process.

The Wonders of Steganography

Steganography is the technique of hiding a message or image within another message or image. Again, the word steganography comes from Greek, consisting of *steganos*, meaning concealed, and *graphe*, meaning writing. Although the term was first used by Johannes Trithemius (1462–1516), an early German cryptographer and Benedictine Abbot, it's very likely steganography has been around for much longer. Written in 1499, Trithemius' seminal three-volume work *Steganographia* was released much later in 1606. While on the surface it seemed to deal with magic and spirits, it was possibly written to conceal and demonstrate the use of cryptographic methods. Scholars still hold differing views on the matter.

Interestingly, British philosopher and statesman Francis Bacon (1561–1626) developed a robust steganographic system all the way back in 1605; it's known as the *Baconian cipher*. This consists of hiding messages not via the content of text, but through its presentation (i.e., typefaces). Bacon visually detailed his steganographic method in his monumental 1623 philosophical work *De Augmentis Scientiarum*.

In practice, classical steganography consists of methods such as invisible ink and the correct interpretation of typefaces to deliver messages to those aware of such content. Modern methods include hiding messages in image files and practically any type of file; digital devices and formats lend themselves well to these techniques. One could utilize audio and

video as well in this context. Digital steganography took off in the mid-1980s and won't be an abandoned practice anytime soon. State secrets and classified military intelligence will continue to be distributed using this method for the unforeseeable future.

European Developments in Cryptography

Europe, too, made great contributions to the science of cryptography. *Leon Battista Alberti (1404–1472)* was an Italian architect and author who devised a cryptographic tool of his own, known as the *Alberti disk*. The device uses polyalphabetics, as originally introduced by Al-Qalqashandi, in the form of two connected disks each divided into 24 cells. The disk was impossible to break without knowing its inner workings. At the time, it was a revolutionary piece of applied cryptography.

Professor Auguste Kerckhoffs (1835–1903) from the Netherlands published two articles in 1883 that are considered classics in the field. His work entitled "Military Cryptography" was featured in the *Journal of Military Science* in France. Kerckhoffs's articles detailed six principles of practical cipher design, which are still quite relevant today. They are as follows:

1. The system should be, if not theoretically unbreakable, unbreakable in practice.
2. The design of a system should not require secrecy, and compromise of the system should not inconvenience the correspondents.
3. The key should be memorable without notes and should be easily changeable.
4. The cryptograms should be transmittable by telegraph.

5. The apparatus or documents should be portable and operable by a single person.
6. The system should be easy, neither requiring knowledge of a long list of rules nor involving mental strain.

Principle number 2 is of particular relevance; it's also referred to as *Kerckhoffs's law* or *Kerckhoffs's axiom*. It states that a cryptosystem should never be vulnerable even if all facets about said system, apart from the decryption key, are public. Although still popular among some government agencies, *security by obscurity (STO)* is mostly an obsolete approach among cryptographers of today. Obscurity shouldn't be considered a factor at all when designing secure systems. If government civil servants defect, for example, your system eventually becomes compromised. STO provides, at best, a layer of pseudo-security.

Some of Kerckhoffs's principles are no longer valid, as computers have become advanced enough to handle complex calculations in mere milliseconds. Also, not many people use telegraphs as of 2020 (although a company called iTelegram has been founded).

At the End of Classical Cryptography

The era and techniques described in the previous sections form a concept called *classical cryptography*. As you probably noticed, it was mostly based on various aspects of linguistics and physical/visual methods. The type of information classical cryptography has an effect on is limited. But we're now moving on to the modern era of all things cryptographic. This is where it gets somewhat complicated – and exciting.

The Digital Cryptographic Revolution

Like many other areas of modern life, computers revolutionized cryptography. In fact, they offered unforeseen possibilities that offered completely secure messaging, a feat almost impossible using traditional methods. Eventually, cryptography was combined with the cutting edge of sciences, including quantum physics. But we're not going there yet; let's have a review of what got us there first.

In 1943 during World War II, British cryptography experts (sometimes called "codebreakers") created the first programmable digital computer, the *Colossus*. It was primarily devised to intercept German military intelligence, but it also helped usher in a new era in electronics. Its German counterpart was the *Lorenz cipher*, a fearsome piece of machinery with a near-perfect track record of encrypting intelligence. However, due to human error, the Lorenz cipher's way of operating was ultimately figured out by the British without ever actually getting their hands on one. Colossus itself was a state secret up until the mid-1970s, with many of the units having been destroyed in the previous decade by the British government.

Strangely, it was only in the 1970s when academia started taking cryptography seriously en masse. Corporations soon picked up the trend; IBM was among the first major companies to develop cryptographic systems and techniques. Their work had a big impact on the US government's data protection policies, for one.

Digital Encryption 101

Encryption in the digital realm consists of basically three things: *an encryption method*, *an encryption key*, and a *decryption key*. An encryption method is the mathematical means of how a message or file is scrambled to appear completely random to a third party. Only the party with a decryption key (i.e., a password) can access the plaintext contents of the file.

Now, there are two widely used encryption approaches in the world today (not to be confused with encryption algorithms, which are a separate concept): *symmetric* and *asymmetric (i.e., public-key cryptography)*. The former uses a single key for both encryption and decryption of the data. The latter uses two separate keys: one public and one private. With this asymmetric approach, the public key is used to encrypt data, while the private key is used for decryption. In a classic example, *Bob* uses *Alice's* public key to encrypt some data. Upon receiving it, *Alice* then uses her private key to decrypt the contents.

Under most circumstances, it's impossible to discover the private key using the public key. Symmetric cryptography is known to be speedier if dealing with large quantities of data. However, the asymmetric approach provides additional security.

Now, a bit is the smallest unit of measurement in data sciences, being represented by either one or zero. The strength of an encryption standard is usually apparent in the amount of bits it carries. There are encryption standards ranging from 40 to 256 bits and more. A couple of these will be discussed next with more elaboration on them coming up later in the book.

The Diffie–Hellman Key Exchange

One of the earliest public-key encryption protocols was the Diffie–Hellman, named after cryptologists *Whitfield Diffie* and *Martin Hellman*. This protocol allows for two parties without any prior knowledge of one another to create a shared secret key/password. The process is done over an insecure channel. Once the shared key is formed, any communications can be secured with it in a separate encryption method. The Diffie–Hellman approach was published in 1976.

Let's go through a simplified Diffie–Hellman exchange. In real-life situations, the numbers would have to be much larger to provide an acceptable degree of security. First, Alice and Bob decide on a modulus (p) and the base (g). Usually the modulus (p) is a large prime number, while the base (g) is kept small to keep things simple.

We'll pick **19** for the modulus and **6** for the base. Next Alice and Bob choose their secret numbers. Let's say Alice (a) picks **5** and Bob (b) picks **2**. The capital A represents the result Alice will send to Bob. Note: *Modulo* is simply the operation of finding the remainder after division of a number by another one.

a = 5 (Alice's choice)

A = ga mod p

= 65 mod 19 = 5

b = 2 (Bob's choice)

B = gb mod p

= 62 mod 19 = 17

Now we calculate Alice's and Bob's secret keys in public without a care in the world:

secretkeya = B a mod p = 175 mod 19 = **6**

secretkeyb = A B mod p = 52 mod 19 = **6**

If both secret keys turn out identical, and they do, the key exchange has been successful. The shared secret number in our example turns out to be 6, which is also the base number. This is not always the case. Rather it's due to us using such small numbers in our example.

The Data Encryption Standard (DES)

The fruit of the interest of IBM in cryptography turned out to be the *Data Encryption Standard (DES)*. This is a flawed but influential encryption symmetric method/algorithm. Released in 1977, the standard initially provided adequate encryption of data and protection against cryptographic attacks, such as *brute-force attacks* which consist of a system/actor trying to enter every single password possible. However, as of 1998, DES was compromised within three days using a computer network created by the *Electronic Frontier Foundation (EFF)*. As of late, due to the increase in computing power, systems encrypted using DES can be compromised within 23 hours. Therefore, the standard is obsolete. You may have noticed an option in some older routers/modems, for example, to secure your wireless Internet connection using DES. Please do not.

DES, despite all the brouhaha back in the day, was only 56-bit long. So it offers 72,057,594,037,927,936 (i.e., 2^{56}) permutations of passwords, which is also known as *key space*. By today's standards, that's not impressive. Any 128-bit (2^{128}) encryption method, on the other hand, provides a key space of 300,000,000,000,000,000,000,000,000,000,000,000 (or 300 decillion) permutations. Now we're talking.

There were corrective measures applied on DES, however. In 1984, a standard called *DESX* was introduced by MIT Professor *Ron Rivest*. His standard added two auxiliary keys to the single 56-bit one found in the original DES, each 64-bit wide. In theory this results in a key space of 184 bits; in practice it's somewhere between 88 and 119 bits. DESX never really took off.

In 1995, an algorithm called *Triple DES* (also stylized as *3DES*) was released. It consists of three rounds of encryption applied to each data block, hence its name. Despite a theoretical key space of 168 bits, which sounds rather impressive, the standard has an effective key space of 112 bits. While better than the original algorithm, Triple DES is still subpar and best avoided. Also, the method is quite slow compared to some of