Data Security

Semester 2

Dr. Noor Alkazaz

Why is data security important?

Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle. It's a concept that encompasses every aspect of information security from the physical security of hardware and storage devices to administrative and access controls, as well as the logical security of software applications. It also includes organizational policies and procedures.

When properly implemented, robust data security strategies will protect an organization's information assets against cybercriminal activities, but they also guard against insider threats and human error, which remains among the leading causes of data breaches today. Data security involves deploying tools and technologies that enhance the organization's visibility into where its critical data resides and how it is used. Ideally, these tools should be able to apply protections like encryption, data masking, and redaction of sensitive files, and should automate reporting to streamline audits and adhering to regulatory requirements.

Introduction to Cryptology

Cryptology is crucial to ensure the protection of information. The main goal of this science is to provide secure confidential systems that guarantee data integrity and privacy. This is achieved by providing well-established cryptographic algorithms that meet the objectives of securing the information system.

The subject of cryptology is the study of security. This science has always been branched into two major lines of study: *cryptography* and *cryptanalysis*. *Cryptography* is the art and science of designing and implementing security algorithms that serve as primitives to provide certain security services such as integrity, confidentiality and authentication. *Cryptanalysis* is the art and science of analysing the security algorithms and defeating their security claims.

Both a cryptographer and a cryptanalyst attempt to translate coded material back to its original form. Normally, a cryptographer works on behalf of a legitimate sender or receiver, whereas a cryptanalyst works on behalf of an unauthorized interceptor.

Cryptology: Where Did it Begin?

Cryptology is the science that has contributed to a variety of disciplines and cultures. The existence of this science can be traced back to the late fourth millennium BC where it was detected in one of the earliest writing systems, Sumerian scripts, in the

form of logographic and syllabic units. Cryptology was also encountered around 1900 BC in Egyptian hieroglyphs. The hieroglyphic writing system combines logographic and alphabetic components that were essentially used to describe the religious literature of that era. For a long period of time, these two writing systems were considered to be an approach used for hidden communication between two different civilizations.

A crucial role was played by the Greeks in constructing ancient cryptology. The use of ciphers or secret codes for the transmission of sensitive messages, whether military, diplomatic or even personal, was reported as early as the era of ancient Greece.

The Caesar substitution cipher was used by the Roman general, Julius Caesar, to communicate secretly with his army during times of war. This cipher is one of the most widely known encryption techniques with each plaintext letter replaced by a letter using a certain shift from its original position in the alphabet. This cipher is referred to as a monoalphabetic substitution cipher as a fixed substitution is applied to the entire plaintext.

Techniques for cryptanalysis developed in parallel to the evolution of cryptography. Statistical analyses were the most effective forms of attack against classical ciphers. In the 800s AD, the Arabic scientist Al-Kindi introduced the first cryptanalysis method against monoalphabetic substitution ciphers using the frequency analysis method. In these types of attacks, statistical analysis is performed on the number of occurrences of specific letter/word combinations. A correlation of ciphertext frequencies with plaintext frequencies and letter distributions helps the opponent to guess the original message.

In order to increase the cryptographic security of ciphers and to overcome the limitations of single substitution, an evolution of this concept was introduced. Several substitution alphabets were used in the newly developed ciphers. Homophonic and polyalphabetic substitution ciphers are examples of these ciphers. In the homophonic substitution cipher, plaintext letters are mapped to more than one ciphertext letter. The number of potential substitutes is proportional to the frequency of the letter (the highest frequency plaintext letters are given more equivalents than lower frequency letters).

The polyalphabetic substitution cipher uses multiple substitution alphabets. The best known example of this cipher is the Vigenère cipher, invented by Blaise de Vigenère in the 16th century. In this cipher, the letters are shifted by different amounts, usually by using a phrase or word as the encryption key. In the 19th century, the first general method of decrypting this cipher was published. Statistical analyses of digrams or trigrams in addition to the examination of repetitions were included in this attack.

More developed versions of substitution ciphers were introduced by armies during times of war and for diplomatic communications, such as the Playfair cipher and new transposition ciphers. In the columnar transposition cipher, digram or trigram statistics were hidden by moving around the plaintext's letters. During World War I (WWI), the use of combined substitution-transposition ciphers was introduced to hide the language statistics.
Cryptanalysis of the new cryptosystems has proven to be more challenging and has required more sophisticated statistical techniques. In most cases, cryptanalysis of these ciphers is extremely difficult especially for short messages with no depth.

In order to overcome the limitations of manual ciphers, several electromechanical cipher machines were introduced from the 1920s through to the early 1960s. These machines, known as rotor cipher machines, are essentially polyalphabetic substitutions which change for each letter encoded. The most famous rotor machine is undoubtedly the Enigma machine used extensively by Nazi Germany during WWII. British Typex and Japanese Red and Purple machines are other famous examples.

The encryption machines introduced in the 1920s and 1930s also led to the invention of other cryptanalysis machines hat were used against the encryption that they produced. These machines played a major role in cryptanalysis while many cryptosystems were still decrypted by hand. The Polish bomba and the Turing bombe are the two main examples of these machines. A team of Polish cryptographers broke a simplified version of the Enigma machine. The attack was extended to the full Enigma machine by British cryptanalysts, including Alan Turing, working at United Kingdom (UK).

In the 1960s and 1970s, fully electronic encryption devices and computer-based encryption were introduced. The invention of computers drove the faster cryptanalysis of the classical cryptosystems. This invention, at the same time, opened the door to the emergence of much more complex encryption systems which were

impossible to implement by hand. The advent of the Data Encryption Standard (DES) is the most noteworthy event. These developments, in addition to the introduction of public key cryptography, marked the end of the era of classical cryptography. In general, even though the principles of classical ciphers form the basis for many of the modern cryptosystems, however, modern ciphers are outside the scope of this research.

The objectives of cryptography

The primary objectives of cryptography are:

1. Confidentiality: This refers to the protection of data from unauthorized disclosure or access. Cryptography can be used to encrypt data so that only authorized parties can access it.
2. Integrity: This refers to the protection of data from unauthorized modification. Cryptography can be used to ensure that data has not been tampered with or altered in transit.
3. Authentication: This refers to the process of verifying the identity of a user or system. Cryptography can be used to provide secure authentication mechanisms that prevent unauthorized access.
4. Non-repudiation: This refers to the prevention of a user from denying that they have performed a particular action. Cryptography can be used to create digital signatures that provide strong evidence of the identity of the signer and the integrity of the signed data.

Terminology

The message (or data) which, prior to ***encoding***, contains intelligible information is called ***plaintext***. The output of ***encoding*** or ***encryption*** after being transformed to a ``secret'' unreadable message is known as ***ciphertext*** or a ***cryptogram***. The set of functions which maps plaintext to ciphertext is called ***encryption***. The key and the reverse algorithm, which generally refer to the secret information, are known as the ***decryption*** process. Decrypting the ciphertext restores the plaintext. An algorithm for performing ***encryption*** or ***decryption*** is known as a ***cipher***. In many systems, the encryption and decryption keys are the same. Such systems are called ***symmetric***; otherwise, the system is ***asymmetric***.

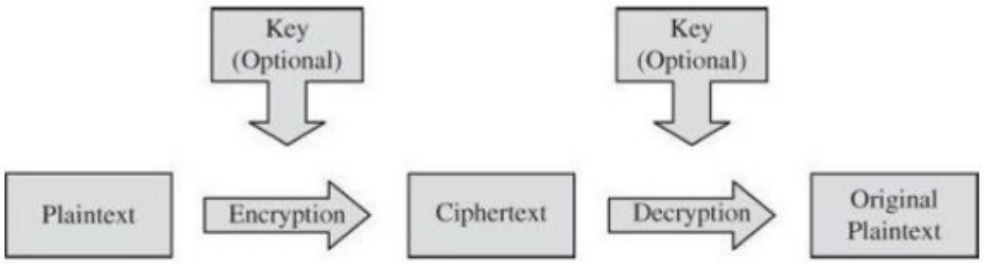

Figure 1: encryption system

We use a formal notation to describe the transformations between plaintext and ciphertext. For example, we write C = E(P) and P = D(C), where C represents the ciphertext, E is the encryption rule, P is the plaintext, and D is the decryption rule. What we seek is a cryptosystem for which P = D(E(P)). In other words, we want to be able to convert the plaintext message to ciphertext to protect it from an intruder, but we also want to be able to get the original message back so that the receiver can read it properly.

Encryption is very old. For example, Caesar's shift cipher 2 was introduced more than 2000 years ago. Every encryption method provides an encryption algorithm E and a decryption algorithm D. In classical encryption schemes, both algorithms depend on the same secret key k. This key k is used for both encryption and decryption. These encryption methods are therefore called symmetric. Another examples of Symmetric encryption and the important examples DES (data encryption standard) and AES (advanced encryption standard).

In 1976, W. Diffe and M.E. Hellman published their famous paper, New Directions in Cryptography. There they introduced the revolutionary concept of public-key cryptography. They provided a solution to the long standing problem of key exchange and pointed the way to digital signatures. The public-key encryption methods are asymmetric. For example, If Alice wants to send a message m to Bob, she encrypts m by use of Bob's publicly known encryption key pk. Bob decrypts the ciphertext by use of his decryption key sk, which is known only to him.

Encryption Keys

A cryptosystem involves a set of rules for how to encrypt the plaintext and decrypt the ciphertext. The encryption and decryption rules, called algorithms, often use a device called a key, denoted by K, so that the resulting ciphertext depends on the original plaintext message, the algorithm, and the key value. We write this

dependence as C = E(K, P). Essentially, E is a set of encryption algorithms, and the key K selects one specific algorithm from the set.

This process is similar to using mass-produced locks in houses. As a homeowner, you would pay dearly to contract with someone to invent and make a lock just for your house. In addition, you would not know whether a particular inventor's lock was really solid or how it compared with those of other inventors. A better solution is to have a few well known, well-respected companies producing standard locks that differ according to the (physical) key. Then, you and your neighbor might have the same brand and style of lock, but your key will open only your lock. In the same way, it is useful to have a few well examined encryption algorithms for everyone to use, but differing keys would prevent someone from breaking into data you are trying to protect.

Sometimes the encryption and decryption keys are the same, so P = D(K,E(K,P)), meaning that the same key, K, is used both to encrypt a message and later to decrypt it. This form is called symmetric or single-key or secret key encryption because D and E are mirror-image processes. As a trivial example, the encryption algorithm might be to shift each plaintext letter forward n positions in the alphabet. For n = 1, A is changed to b, B to c, ... P to q, ... and Z to a, so we say the key value is n, movingn positions forward for encryption and backward for decryption. (You might notice that we have written the plaintext in uppercase letters and the corresponding ciphertext in lowercase; cryptographers sometimes use that convention to help them distinguish the two.)

"Symmetric encryption: one key encrypts and decrypts."

At other times, encryption and decryption keys come in pairs. Then, a decryption key, KD, inverts the encryption of key KE, so that P = D(KD, E(KE,P)). Encryption algorithms of this form are called asymmetric or public key because converting C back to P involves a series of steps and a key that are different from the steps and key of E. The difference between symmetric and asymmetric encryption is shown in Figure 2.

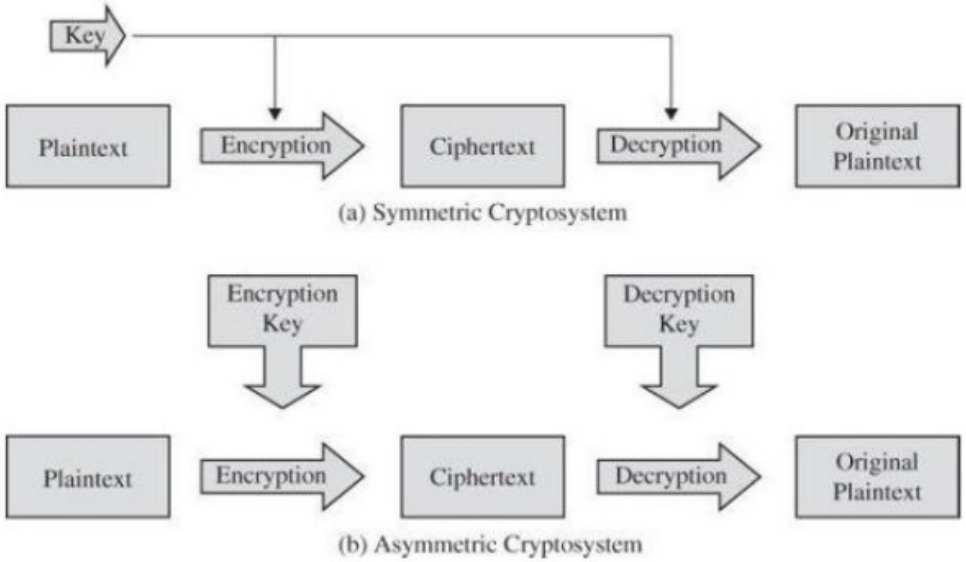

Figure 2: symmetric and asymmetric encryption

"Asymmetric encryption: one key encrypts, a different key decrypts."

A key gives us flexibility in using an encryption scheme. We can create different encryptions of one plaintext message just by changing the key. Moreover, using a key provides additional security. If the encryption algorithm should fall into the interceptor's hands, future messages can still be kept secret because the interceptor will not know the key value. An encryption scheme that does not require the use of a key is called a keyless cipher.

Cryptanalysis

A cryptanalyst's chore is to break an encryption. That is, the cryptanalyst attempts to deduce the original meaning of a ciphertext message. Better yet, the cryptanalyst hopes to determine which decrypting algorithm, and ideally which key, matches the encrypting algorithm to be able to break other messages encoded in the same way.
For instance, suppose two countries are at war and the first country has intercepted encrypted messages of the second. Cryptanalysts of the first country want to decipher a particular message so as to anticipate the movements and resources of the second. But even better is to discover the actual decryption method; then the first country can penetrate the encryption of all messages sent by the second country.

Thus, a cryptanalyst can attempt to do any or all of six different things:

- Break a single message

- Recognize patterns in encrypted messages, to be able to break subsequent ones by applying a straightforward decryption algorithm
- Conclude some meaning without even breaking the encryption, such as noticing an unusual frequency of communication or determining something by whether the communication was short or long.
- Deduce the key, to break subsequent messages easily
- Find weaknesses in the implementation or environment of use of encryption
- Find general weaknesses in an encryption algorithm, without necessarily having intercepted any messages

An analyst works with a variety of information: encrypted messages, known encryption algorithms, intercepted plaintext, data items known or suspected to be in a ciphertext message, mathematical or statistical tools and techniques, and properties of languages, as well as plenty of ingenuity and luck. Each piece of evidence can provide a clue, and the analyst puts the clues together to try to form a larger picture of a message's meaning in the context of how the encryption is done. Remember that there are no rules. An interceptor can use any means available to tease out the meaning of the message.

Work Factor

An encryption algorithm is called breakable when, given enough time and data, an analyst can determine the algorithm. However, an algorithm that is theoretically breakable may in fact be impractical to try to break. To see why, consider a 25-character message that is expressed in just uppercase letters. A given cipher scheme may have 2625 (approximately 1035) possible decipherments, so the task is to select the right one out of the 2625. If your computer could perform on the order of 1010 operations per second, finding this decipherment would require on the order of 1025 seconds, or roughly 1017 years. In this case, although we know that theoretically we could generate the solution, determining the deciphering algorithm by examining all possibilities can be ignored as infeasible with current technology.

The difficulty of breaking an encryption is called its work factor. Again, an analogy to physical locks may prove helpful. As you know, physical keys have notches or other irregularities, and the notches cause pins to move inside a lock, allowing the lock to open. Some simple locks, such as those sold with suitcases, have only one notch, so these locks can often be opened with just a piece of bent wire; worse yet, some manufacturers produce only a few (and sometimes just one!) distinct internal pin

designs; you might be able to open any such lock with a ring of just a few keys. Clearly these locks are cosmetic only.

Work factor: amount of effort needed to break an encryption (or mount a successful attack)

Two other important issues must be addressed when considering the breakability of encryption algorithms. First, the cryptanalyst cannot be expected to try only the hard, long way. In the example just presented, the obvious decryption might require 2625 machine operations, but a more ingenious approach might require only 1015 operations. At the speed of 1010 operations per second, 1015 operations take slightly more than one day. The ingenious approach is certainly feasible. In fact, newspapers sometimes print cryptogram puzzles that humans solve with pen and paper alone, so there is clearly a shortcut to our computer machine time estimate of years or even one day of effort. The newspaper games give hints about word lengths and repeated characters, so humans are solving an easier problem. As we said, however, cryptanalysts also use every piece of information at their disposal.
Second, estimates of breakability are based on current technology. An enormous advance in computing technology has occurred since 1950. Things that were infeasible in 1940 became possible by the 1950s, and every succeeding decade has brought greater improvements. A conjecture known as “Moore’s Law” asserts that the speed of processors doubles every 1.5 years, and this conjecture has been true for over three decades. We dare not pronounce an algorithm secure just because it cannot be broken with current technology, or worse, that it has not been broken yet.