BÁNKI KÖZLEMÉNYEK
V. ÉVFOLYAM 1. SZÁM

Cryptography and cryptographic packets in Python

Una Sredović
University of Donja Gorica, Podgorica, Montenegro
una.sredovic@udg.edu.me

***Abstract* — Nowadays, we usually mean that the information we receive/ send digitally is secure and that only those to whom we in- tended it and the person sending it have access. Cryptography deals with the ques- tions of how, why and by what methods data of any type flows. Algebra, combinatorics and probability are the mathematical branches on which data encryption and de- cryption are based. All information passes through the protocol under some key - sym- metric or asymmetric. Each of them has its own methods, which in turn have advantages or disadvantages, and are still used today, al-though they are constantly being developed and improved. Such an information exchange system is important for security. although at-tacks (or their types) have been detected and classified over time, constant caution againstnew "hacker" methods and variations is always necessary. Recently, the Python pro- gramming language has been widely used, for example in databases, where it is extremely important that data remains anonymous or circulates securely. In this regard, many packages have been developed that have facili- tated their encryption and are regularly up- dated. As artificial intelligence and machine learning develop in parallel, models are pro- duced that facilitate tracking and analysis without constant human monitoring, and onesuch example is the placement of productsby a company from a foreign user to a user, based on some of his data.**

Keywords: Cryptography, Symmetric key, Asymmetric key, Python, Machine learning

1 INTRODUCTION

The modern age in which digital devices, both computers and mobile devices such as phones, tablets or smart watches, have developed, has contributed to thedevelopment of electronic communication. Such com- munication must be secure at all times, since the in- dividual does not have a visible path through which any message passes, but it is an abstract, digital path.Cryptography (a word of Greek origin, derived from the words kryptós- meaning “hidden” and graphein- mean- ing “to write”) is a science that deals with methods for forming and keeping information secret. Throughout history, even today, it has always been important to pass on information, that is, to turn it into a message. As we live in a digital age where artificial intelligence and robotics are evolving rapidly, we can get a machine that would construct or decrypt codes much faster and more efficiently than human. In order for a machine to function independently, machine learning is required. Machine learning is a set of algorithms whose model is used to analyse data based on previously processed information (usually based on images). It was first mentioned in the middle of the 20th century and has been constantly evolving since then. It is closely related to statistics, but also to other mathematical disciplines. As an extremely large amount of data flows in the process of cryptography or cryptanalysis, a person can easily lose and "waste" a lot of time, which for some more complex algorithms requires a machine, but for basic calculations and some already discovered methods, it is extremely easier for computer. Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

2 EVOLUTION

Before analysing cryptography in more detail, we will try to put together the development of encoding and decoding. Although they seem obvious and too sim- ple, the methods and ideas used by the Egyptians and the ancient Greeks are still used today. Some even be- lieve that thanks to them, Hitler and World War II were stopped. Today we distinguish three historical periods in the development of cryptography: ancient period (until 1918) - Caesar’s algorithm and Wisner’s code; technical period (from 1918 to 1975) - Playfair’s algorithm and Enigma; paradoxical period (since 1975)-AES and DES system [1].

Enigma: The forerunner of the Enigma was a device similar to today's rotor, which served for key generation. There was a recess in the machine that would turn and move the position of the letter each time the button was pressed. German scientist Arthur Scherbius set the initial conditions that were later upgraded since it would be repeated after the sixth rotation of the letters, which gave an insufficiently large number of combinations, ie 26*26*26 = 17576. Later it was developed to 10,000,000,000,000,000 combination, so it was accepted in the army and used during World War II rate. Enigma was developed in England, mostly thanks to Alan Turing and Marian Rejewski, who concluded that every German message must be encrypted twice at the beginning - and thus came across the first clue to decryption, after which it was much easier to reach other segments. With the help of 5 rotors, they managed to find out the keys that the Germans would send every day, and change exactly them at midnight.Adjusting the weights between neurons enables the learning ability of an artificial neuron.

Keys in cryptography

The function of copying a message into some incomprehensible text in order to maintain secrecy and the

basic message is called key. Based on the ideas of previous encryption methods, much larger keys are being formed today, for the simple reason that to- day there are more advanced machines that can count much faster than humans. Today, two types of keys are used: symmetric (same key used by both parties) and asymmetric key (two keys where each party uses a different one). An example we can find in everyday life is the activation key for Windows operating system.

If two people have the ability to exchange the key verbally or outside the communication channel, and then continue to exchange messages over the key agreed upon, the symmetric key can work. The main opera- tion used to generate such keys is XOR (exclusive or) which is the addition of two numbers per module 2. Let A be the message and B the key, then the value (A xor B) = M passes through the communication chan- nel, while the other part gets the original message by calculating M xor B. pseudo-random string generators are often used to form deterministic encryption algo- rithms, although such strings are similar to random strings. There are two types of symmetric key:

Block cipher With such code, the message is broken into several blocks/segments with the help of a key and it passes to its recipient. Some of the algorithms that use the symmetric key block code are AES, DES, IDEA, Blowfish, RC6, RC6.

Stream cipher Instead of being stored in memory, such codes are encrypted with a key during the pas- sage of the code through the communication channel. An example of a symmetric key flow code is the RC4 algorithm.

Since Alice and Bob do not always have the ability to exchange a key outside the communication chan- nel in the real world, symmetric keys are rarely used today. The above algorithms are mostly inapplicable, although some of their more advanced forms are used, such as 3DES). We may encounter these algorithms in some transactions, message authentication, or hashing.

Asymmetric keys were first mentioned in *IEEE Transactions on Information Theory* in 1976 as a discovery by Whitfield Diffie and Martin Hellman. There are twokeys to this algorithm - private and public. The pub- lic key is for encryption only, while the private key is used for decryption and is owned by anyone who wants to receive a message. Unlike the symmetric algorithm, there is no danger of "intercepting" the message, because in that case the attacker would have access onlyto the text that is encrypted under public key, and it is available to everyone. Three parameters are importantin this algorithm: The pair (K_P, K_S), where K_P is the public key and K_S is the private key; Enc encryption algorithm; Dec decryption.

Example 1: As an example, we can take prime p =21489151 (although p is not large enough in real life). We can pick an arbitrary g like g = 1609879. Then A picks x = 3916708 at random and com- putes X = gxmodp = 13164781 and sends it to A. Then B picks y = 16766518 at random and computes Y = gymodp = 4109137 and sends it to B. In paral- lel, B computes K = Xymodp = 13275737. A can also compute K = Y xmodp = 13275737 [2]

In order for such a code to be of good quality, it is necessary to form a large enough p, which is not at all simple. That is why the Euler function and the following theorem are important to us:

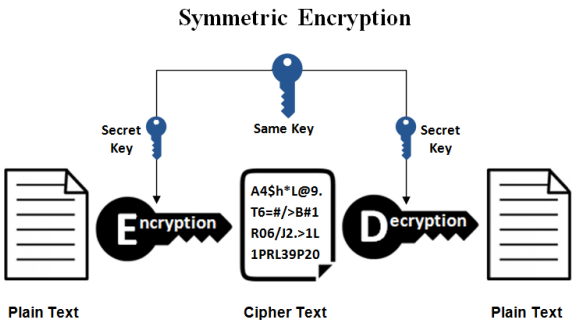
Theorem 1. Let a and b be nonzero integers. Then there exist integers r and s such that gcd(a, b) = ar +bs

and:

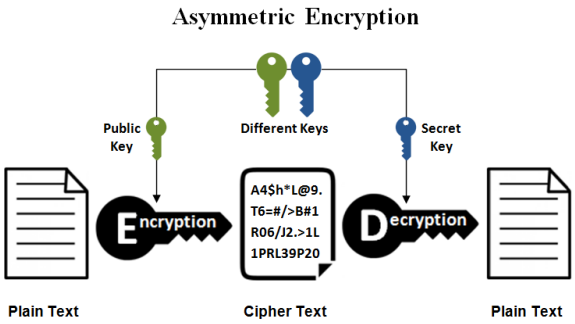
Corollary 1. Let a and b be two integers that are relatively prime. Then there exist integers r and s such that ar+bs = 1.

Definition 1. $\phi(n)$ is the number of non-negative integers less than n that are relatively prime to n. In other words, if n>1 then $\phi(n)$ is the number of elements in Un, and $\phi(1)=1$.

RSA model Let p and q be two prime numbers. Let n = pq and $\phi(n) = m = (p\ 1)(q\ 1)$ where ϕ is the Euler function. We are looking for a number E that is mutually prime with m (we need such an E to satisfy the gcd (E, m) = 1). By Euclidean algorithm we can get the number D, so that DE 1 modulo m. When we have calculated this, we publish the numbers E and n as the public key. In order for Bob to send Alice a mes- sage x, he needs to send a new number/code y = xE; while Alice, in order to decipher the message, needs to solve x = yD where only she knows D.



1. figure: Scheme of symmetric encription[3]



2. figure: Scheme of asymmetric encription[3]

Symmetric or asymmetric key algorithms? Although it is in much wider use asymmetric key, it has some disadvantages compared to symmetric. Namely, symmetrical keys do not require strong hardware due to a simpler algorithm, they need more time (precisely because of the more complex key) but they are more reliable than the symmetrical one keys and work where they can't (if

Alice and Bob can't exchange key outside the communication channel).

3 SECURITY OF CRYPTOGRAPHIC ALGORITHMS

The key is usually represented in a binary system and is formed large enough for the brute-force algorithm to be insignificant when it is discovered. In addition to keeping the message secure, the secrecy of the key is also important.

If the attacker does not have any new knowledge about the basic message after reading the encrypted message, we say that the cryptosystem has perfect security. The system that provides perfect security is called one - time pad and its basic features are:

1. The number of basic messages is less than or equal to the number of keys
2. The key is chosen randomly from the domain of uniform probability.
3. The key is used for one encryption only

Definition 2 (Protocol). A protocol is a set of rules and conventions for sending a message over a network.

It is a service that is provided by a protocol layer of Protocol is safe if the following is provided:

- Data integrity: ensures consistency and accuracy of the message
- Data confidentiality: data protection from unauthorized persons
- Authenticity: confirmation of message originality
- Access control: Checks if the client has legal access to the data
- Nonrepudiation: prevents the recipient from modifying the content (most commonly used for digital signatures)
- Resource availability: procedure in case of cancellation, error or detection of an attack

The higher the number of keys, the less likely an attacker is to hit the right one. However, if the attacker were to find out the length of the key or basic message, security is still at a high level considering the length of the message is insufficient to reveal the context of the same. [2]

SSH protocol: If there are a computer and a server (or two computers) that should connect, but for some reason, they can't do it with a secure communication channel, SSH ("Secure SHell") is a network protocol that allows users exchange of data in such a situation. When the client (computer) should connect to the server, the client forms two keys - private (Ks) and public (Kp). Then sends the public key to the server (which is legitimate, anyone can have the public key), where the server stores that key (Kp) in memory and forms a new key (Kv) and sends it back to the client two (Kp, Kv). Since with the private key (X), the client can erase which is (Kv), he sends such a key back to the server based on which he can confirm the identity of the client with whom he communicates and form a new, secure channel.

No matter how many rules you set, there will always be the idea that any cryptosystem is an "easy target" for revealing trusted or any other data. The knowledge so far about the interception and attack on the cryptosystem is divided into two groups: passive and active attacks.

Passive attacks: Passive attacks are similar to machine learning in that they aim to "learn" what the keys are based on monitoring the message flow. There is no possibility to modify the content but can abuse it. The most common forms of passive attack are:

- Analyzing the flow - Alice and Bob communicate believing they have a secure channel, but the attacker observes the frequency of their messages

- Posting a message - Although he cannot change it, the basic message of the attacker may publish or share

Active attacks: Types of active attacks, except that they can change its content, can also redirect the message to an unwanted location; answer or pretend to the right recipient for a certain period. Such attacks are more difficult to detect than passive ones, so their algorithms are quite more complex.

4 EXAMPLE IN PYTHON

Packages that were used in this example are:

Pandas: Pan(el) da ta s is "open - source" code used to analyse, process and sort data as well as their tables. The data types supported in this package are Data Frame - a data structure formed of types and columns, where each column contains the value of one attribute, and each type represents a set of values of all attributes; Data - an argument that represents data and is mandatory for definition; index, columns, dtype are optional arguments and serve to present the table more accurately. It is important in cryptography because of the simple manipulation of data, which generally has a lot.

Numpy: Mainly used with Pandas, it is also "open - source" and works with arrays. It is useful because it can easily work with arrays and matrices of large dimensions. ndarray is the main data structure in this package, in which all members of the array must be of the same type.

Paillier: One of the packages from the Python Library formed for data encryption. It is used for homomorphic encryption and its main advantage is that encrypted numbers can be multiplied by unencrypted ones. The parameters of this package are

- Publickey - the key based on which the message is encrypted;
- ciphertext - encrypted message in the form of an integer;
- exponent - mostly a negative number, the one that represents the exponent in the encryption equation

JSON: A package that works with JSON (JavaScript Object Notation) data is used to store and share data. It is easy to read because its elements are shown in pairs (name, value).

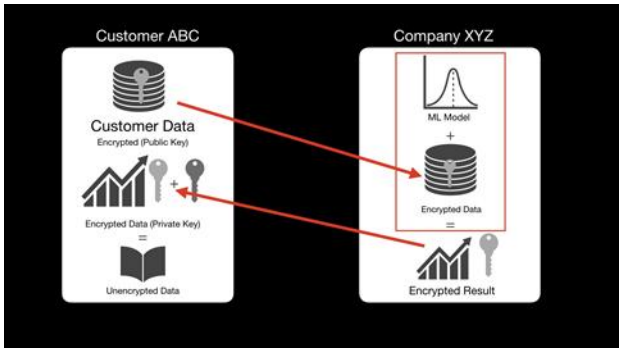
This example is a mini - version of large companies, a model that shows how data passes through the protocol and how it is processed. The idea is to present the product on the client based on his salary, the data from the table we received in csv format. Since privacy is mandatory, all messages and all data that would be useful for the client/company are sent in the form of a code (number). It would be unthinkable for every company to have a human being behind every computer waiting for the codes to arrive, then to analyse them and based on that analysis send back a product that he assumes is interesting/attractive to the client. In order for a machine to do all that, machine

learning is needed. With the help of the packages, the computer analyses, remembers and tries to guess which product to return as feedback and thus make a profit. It is clear that such an analysis requires as many attributes as possible, as well as data, and of course a longer period of time in order for the analysis to be more accurate.

In this example, we have the ABC client and the XYZ firm that get some data encrypted with an asym- metric cryptography model, using a public key. In com- bination with its machine algorithm and the obtained information about the client, it can extract its data (products) which will be returned to the client also with the help of a public key. Since the customer has his private key, he will use it to decrypt the message received from XYZ and thus read the product/message he received. [4]

Process:

1. Machine learning will be based on 4 coefficients: age, gender, lifestyle and healthy diet. The first file is the part where we form the machine learning model. Based on the data from the table, we train a machine that should predict which data will be important to the client. Such a model is used by XYZ, which is based on a linear regression type algorithm.



3. figure: Scheme of coding and decoding data [5]

2. The second file contains improvised data that is encrypted using a public and private key. Such keys are stored in a single JSON file in which all data is encrypted. What XYZ gets is just a dictionary with completely unclear values (large volume). The answer part is used to load products from the company.
3. Since it now has its own model for machine learning and the data it received from the client, the company, with the help of mathematical calculations, extracts the data that it will return to the client. There is a function named "getData" that displays what the com- pany sees as the information it received (it must be the same as in the JSON file). "computeData" will be able to generate the product to be presented to the client with the help of the coefficients from file 1 and the information just obtained. Of course, just as it must be secure, it will be encrypted with a public key and such code will be sent back at the target group. With the private key, that group will be able to decipher the message and thus find out the product it received from the company.

5 Conclusion

Consciously or unconsciously, we use cryptography and its methods every day. Using ATM, online payments and messages via social networks are life situations for which data encryption (in the form of numbers) is necessary. It is believed that the breaking of the German Enigma by Alan Turing, Joan Kankors, Hugh Alexander and Joan Clark stopped the Second World War, which created the Turing machine - the forerunner of today's computers. Thanks to increasingly advanced machines and programming languages, one no longer cares about the daily security of information exchange, although there is always the side that tries to access them illegally. Clearly, as economics, politics, and technology evolve, there will be an increasing need for cryptography development as well.

References

[1] Shyam Nandan Kumar. Review on network security and cryptography. International Transaction of Electrical and Computer Engineers System, 3(1):1–11, 2015.

[2] Bašić, B. D., Čupić, M., Šnajder, J. (2008). Umjetne neuronske mreže. *Zagreb: Fakultet elektrotehnike i računarstva*, 7–15.

[3] David Bishop. Introduction to Cryptography with java Applets. Jones & Bartlett Learning, 2003

[4] Example,https://www.ssl2buy.com/wiki/symmetric-vs-asymmetric-encryption-what-are-differences, Last access 26.03.2023.

[5] Example,howpublished=https://www.youtube.com/watch?v=nlsd2los50, note = last access: 21.04.2022.

[6] Python,howpublished=https://github.com/satssehgal/homomorphicencryption/blob/master/pic.jpg, note = last access: 21.04.20