An Exploration of Modern Cryptography

Siddharth
Mahendraker

March 28, 2012

Abstract

Fusce dapibus, tellus ac cursus commodo, tortor mauris condimentum nibh, ut fermentum massa justo sit amet risus. Maecenas sed diam eget risus varius blandit sit amet non magna. Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum. Nulla vitae elit libero, a pharetra augue. Donec id elit non mi porta gravida at eget metus.

Contents

Introduction

Suppose two people, Alice and Bob, wish to communicate by mail and do not want their mailwoman, Eve, to be able to read their messages. Alice and Bob are military personal of the same country, but they have never met each other before. Because Eve is the mailwoman, she will be able to read all of the messages passing between Alice and Bob, but her obligation to the postal service prevents her from tampering with these messages[1].

The question is, is it possible for Alice and Bob to communicate securely in these circumstances? Astoundingly, the answer is yes!

Cryptography is the science of securely sending messages over insecure channels. Using cryptographic techniques, Alice and Bob can be sure that their communications are illegible to Eve.

0.1 Terminology and Basic Concepts

0.1.1 Alice and Bob

Unless specified otherwise, Alice and Bob are two parties attempting to communicate over an insecure channel, and Eve is their adversary trying to read their messages.

0.1.2 Encryption and Decryption

Messages in cryptography are formally called plaintext. When messages are scrambled, or made "illegible", they are encrypted. The encrypted form of these messages is called ciphertext. The reverse process of encryption, decryption, accepts ciphertext as input and returns plaintext. We can describe this mathematically as:

$$E(M) = C$$
$$E'(C) = D(C) = M$$

Where M denotes plaintext, C denotes cipher text, E is the encryption function and D is the decryption function, or the inverse of E. Also note

[1] In reality, Eve would not be bound by such a petty obligation, however, for the sake of simplicity, let us assume this is true.

the following identity:

$$D(E(M)) = M$$
$$E(D(C)) = C$$

0.1.3 Ciphers and Keys

A cryptographic algorithm, or cipher, is a function used for encryption and decryption.

If the workings of a cipher are made public, then the messages of anyone who is known to use the cipher can quickly be comprimised by simply implementing an inverse of the cipher. Therefore, cryptographers introduced a key. A key is a piece of secret, private information upon which the cipher depends. It often takes the form of a number. The total number of possible values a key can take on is called they keyspace. Because ciphers depend on the key to encrypt and decrypt plaintext, encryption and decryption is often denoted:

$$E_K(P) = C$$
$$D_K(C) = P$$

The key is denoted by K and the keyspace is by $\mathcal{K}$. Note that the identity mentioned in 0.1.2 still holds true for ciphers.

0.1.4 Symmetric and Public-Key Ciphers

There are two distinct kinds of ciphers, symmetric ciphers and public-key (or asymmetric) ciphers.

Symmetric ciphers are ciphers for which the key used to decrypt ciphertext and encrypt plaintext is the same. In most symmetric ciphers, this means that Alice and Bob will need to agree on a key before they can begin sending messages. Symmetric ciphers can be further categorized as stream ciphers or block ciphers. Stream ciphers operate on only one bit (or byte) of plaintext at a time, where as block ciphers operate on a large number of bytes at once.

Public-key ciphers are ciphers for which the the key used for encrypting plaintext is different from the the key used for decrypting ciphertext. Further, these keys should be independent of each other, meaning the decryption key can not be calculated[2]from the encryption key. The design of this

cipher is such that the encryption key can be published for anyone to use, but only the owner of the decryption key can decrypt the message. This is why is encryption key is referred to as the public key and the decryption key is referred to as the private key.

0.1.5 Cryptanalysis

Crytanalysis is the study of obtaining the plaintext from encrypted messages without the knowledge of the key. An attempt to cryptanalyse a cipher is called an attack. Successful attacks often reveal either the plaintext, the secret key, or both.

The only assumption made in cryptanalysis is that the only piece of information the users of the cipher, Alice and Bob know that the adversary Eve does not is the secret key. This means that all other information, including communications and the workings of their cryptographic algorithm are available to anyone. This assumption implies that the security of the algorithm rests only in the key, and nothing else.

There are three main cryptanalysis techniques we will be focusing on in this report. Listed in decreasing order of difficulty they are; ciphertext only attacks, known-plaintext attacks and chosen plaintext attacks.

In ciphertext only attacks, the cryptanalyst (or attacker) Eve has access to several different ciphertexts. The attack is considered successful if Eve successfully retrieves the plaintexts corresponding to the ciphertexts or the key used in encryption.

In known-plaintext attacks, Eve has access to the ciphertexts as well as their corresponding plaintexts. The attack is successful if Eve finds the key (or keys) used to encrypt each plaintext.

In chosen plaintext attacks, Eve can not only access the ciphertexts, and their corresponding plaintext, but can also choose which plaintexts are encrypted and which are decrypted. The attack is successful if Eve retrieves the key (or keys) used to encrypt each plaintext.

Note that in all of the cases above, the adversary, Eve, had to know some amount of "information" about the ciphertext, plaintext or the relationship between the two. The only other technique which can yield the key is a brute force attack or exhaustive search attack, in which Eve checks the

[2] In a reasonably amount of time ofcourse.

ciphertext against all possible keys in the keyspace until one of the keys reveals the plaintext.

1 Substitution Ciphers

A substitution cipher is a cipher in which each character or byte in the plaintext is substituted with a character or byte in the cipher text.

This can be seen mathematically as:

$$\begin{aligned} m &= fromChar(i) \\ m + k &= n \\ c &= toChar(n) \end{aligned}$$

Decryption works the other way around. An already shifted character, i, is transformed into it's integer representation, m, from which the key is subtracted to yield the integer n, which is finally transformed back into the plaintext.

$$\begin{aligned} m &= fromChar(i) \\ m - k &= n \\ p &= toChar(n) \end{aligned}$$

This can be more concisely explained in pseudocode:

1.1 Information Theory and Languages

Before we begin a deconstruction of the Caesar cipher, there are a few assumptions we make that must be explained.

Firstly, we must clarify the definition of information we used in section 0.X.X. Information can be rigorously defined as the least number of bits it would take to represent all possible meanings of a message, assuming all messages are equally likely.

For example, suppose we are trying to determine the amount of information in a list of possible sexes:

```
1. Male
2. Female
```

Clearly, this data can be represented using one bit, where the 1 represents male and 0 represents female. Therefore, we can say that there is only one bit of information present in this list.

Now, if we take a look at words used in the English language, we clearly see that English does not represent this information very succintly, e.i there is lots of redundance per character.

For example, the sentence "met u tmrw @ 9" conveys the same information as the sentence "meet you tomorrow at nine", yet does do much more succinctly. Therefore, we could say that many of the character in the latter sentence are redundant or useless.

Although this may not seem related at all to cryptography, it is. This redundancy in languages causes sentences to "leak" more information than they need to. As we shall soon see, this often manifests itself as discrepancies in the frequency and location of certain characters in relation to others, and makes breaking the Caesar cipher a piece of cake.

1.2 Cryptanalysis of the Caesar Cipher

If we take the most nave cryptanalytic approach, a brute force attack, the Caesar cipher appears quite strong. Indeed the keyspace of the cipher is $26!$ or approximately 4×10^{26}! This means that even if we were to check even a million keys per second, it would still take us around 1.27×10^{13} years to check every possible key! That longer than the estimated age of the universe!!

However, we know from our understanding of redundancy in language that there is information being leaked here.

Because the output of this algorithm merely "switches" one letter with another, the letters in any particular ciphertext will continue to follow the known statistic rules regarding English text. Particularly, they will maintain certain distributions of characters, bi-grams and tri-grams over the message.

Letter Frequency (%)			
E	13.11	M	2.55
T	10.47	U	2.45
A	8.15	G	1.95
O	8.05	Y	1.95
N	7.15	P	1.96
R	6.85	W	1.55
I	6.35	B	1.45
S	6.15	V	0.95
H	5.25	K	0.45
D	3.75	X	0.15
L	3.35	J	0.15
F	2.95	Q	0.15
C	2.75	Z	0.05

Figure 1: General frequency of English characters in decreasing order

Therefore, if we are given the following ciphertext:

```
ofobiyxocryevnkvcyexnobcdkxndrovswsdkdsyxcypmbizdyqbkz
rikckdyyvgroxeconsxmyxtexmdsyxgsdrcyvsnzbyqbkwwsxqzbkm
dsmockxnpsbwwkdrowkdsmkvmyxtomdebocsdmkxlorsqrvioppomd
sforygofobspwscecondrobowkilonsckcdobyecmyxcoaeoxmocsx
mvensxqwkccnkdkdropdybcobfobrsqrtkmusxq
```

We would first construct a table of the character present in the text and their respective frequencies, as so:

Character	o	s	c	k	d	x	y	b	m	r	n	e	w	v	q	p	i	f	z	g	t
Frequency	28	21	19	19	19	18	17	15	14	13	10	9	9	8	7	6	5	4	4	3	3

Figure 2: Frequency of English characters in the ciphertext

Then we would map the most frequent characters to each other. Clearly, the character "o" appears to represent the character "e". Based on our knowledge of the algorithm, we know the key is used as a shift for the integer values of each character. A quick glance at the ASCII character table reveals that e = 101 and o = 111 as integers. Therefore, we can conclude that the key used to encrypt this plaintext is the number 10.

A quick check reveals that we were indeed correct.

```
everyoneshouldalsounderstandthelimitationsofcryptograp
hyasatoolwhenusedinconjunctionwithsolidprogrammingprac
ticesandfirmmathematicalconjecturesitcanbehighlyeffect
ivehoweverifmisusedtheremaybedisasterousconsequencesin
cludingmassdatatheftorserverhighjacking
```

With proper punctuation and capitalization the plaintext becomes:

```
Everyone should also understand the limitations of cry
ptography as a tool. When used in conjunction with sol
id programming practices and firm mathematical conject
ures, it can be highly effective. However, if misused,
there may be disasterous consequences, including mass
data theft or server highjacking.
```

1.3 Advantages and Disadvantages of the Caesar Cipher

At this point, you may be asking yourself why the Caesar cipher would ever be considered a viable method of encrypting data, considering we have been able to break it quite easily using a ciphertext only attack.

Although this cipher is very deeply flawed, it still has certain advantages which make it practical in certain situations.

For example, if speed and memory are your main concerns, and your plaintext only has to be superficially secure, then this cipher is one of your best options. The Caesar cipher has a time complexity of $\Theta(1)$ and a memory complexity of $\Theta(1)$. This means that the cipher's speed and memory usage stay within a constant range, and do not grow (or shrink) in relation to the size of the cipher's input. This is because the cipher operates on only one character (or byte) at a time, and performs the same constant time operation on each byte.

Furthermore, the Caesar cipher may also be practical in situations where only a small amount of plaintext is being encrypted. The statistical analysis which was used is only relevant to plaintexts of sufficient length. It has been shown that highly competent cryptanalysts can break the Caesar cipher using only 25 English characters of plaintext. Therefore, the Caesar cipher might be practical for messages shorter than 25 characters.

2 Block Ciphers

3 ECC - Elliptic Curve Cryptography