Cryptography I: An Introduction

Noah Singer

Montgomery Blair High School Cybersecurity Team

November 11, 2016

Overview

What is cryptography?

Definition (Cryptography)

The study of systematic methods for secure communication.

Definition (Cryptosystem)

A method or system used to securely transmit information.

Definition (Cryptoanalysis)

The theoretical analysis of cryptosystems and their vulnerabilities.

Applications of cryptography

- **Encryption**: Encoding a message so that it can only be read by those the sender wants to read it.
- **Digital signatures**: Demonstrating the authenticity of a message to its recipient.
- **Secret sharing**: Splitting a "secret" into shares such that certain numbers and types of shares are necessary to reconstruct the secret.
- **Zero-knowledge proofs**: Proving a fact without revealing any information about that fact.
- and much more!

Some names to start

We'll use these names throughout our protocols to refer to the parties involved (they're pretty standard).

- Alice: The first party in a cryptosystem.
- Bob: The second party in a cryptosystem.
- Eve: An eavesdropper on the conversation.

Encryption

Definition (Encryption)

Encoding a message so that it can only be understood by authorized parties.

The message to be encoded is the **plaintext**, and the message that is the result of encryption is the **ciphertext**. The opposite of encryption is **decryption**, which is decoding an encrypted message.

Keys

Definition (Key)

A piece of information that specifies how to transform plaintext into ciphertext or ciphertext into plaintext.

The key is a "parameter" of the encryption algorithm that determines what each plaintext is mapped to and vice versa. In **symmetric encryption**, the keys for encryption and decryption are the same. The set of all possible keys is known as the **keyspace**.

Kirchoff's principle

THE SECURITY OF THE PLAINTEXT SHOULD DEPEND ONLY UPON THE SECRECY OF THE KEY.

... and not on anything else, like hiding the algorithm's inner workings (aka **security through obscurity**).

Attacks

Definition (Attack model)

A classification of an attack on an encryption algorithm based on the level of access that the attacker has to the system.

- We can characterize attacks by how strong they are (how much information they require). From weakest to strongest, some common types of attack models are:
  1. **Ciphertext-only**: The attacker only has access to some set of encrypted ciphertexts.
  2. **Known-plaintext**: The attacker has access to a set of encrypted ciphertexts and their corresponding plaintexts.
  3. **Chosen-plaintext**: The attacker has access to the corresponding ciphertexts for plaintext of their choosing.

Attacks

We can also classify attacks based on how much information we recover (this is a continuous spectrum, not a discrete list):

1. **Total break**: The attacker recovers the secret key.
2. **Partial break:** The attacker is able to decrypt ciphertexts without knowing the full key.
3. **Informational break:** The attacker can deduce some information about the key or the plaintext corresponding to a ciphertext but not all of it (e.g. certain bits).

Side-channel attacks exploit a cryptosystem's implementation rather than the cryptosystem itself, generally by carefully observing quantities such as:

- The amount of power a processor is drawing.
- The noise a computer makes.
- The amount of time a step of the cryptographic algorithm takes to complete.

Caesar cipher

Now, let's consider a simple type of cipher. Let the key k be in $K = \{0 \ldots 25\}$. We encrypt character-by-character, treating each character A through Z as some integer x between 0 and 25.

$$E(x, k) := (x + k) \pmod{26}$$
$$D(x, k) := (x - k) \pmod{26}$$

Example

Let's decrypt the string `AXEEHPHKEW` with key $K = 19$.

$\texttt{A} \to D(0, 19) = 0 - 19 \pmod{26} = 7 \to \texttt{H}$
$\texttt{X} \to D(23, 19) = 23 - 19 \pmod{26} = 4 \to \texttt{E}$
$\texttt{E} \to D(4, 19) = 4 - 19 \pmod{26} = 11 \to \texttt{L}$
$\texttt{E} \to D(4, 19) = 4 - 19 \pmod{26} = 11 \to \texttt{L}$
$\texttt{H} \to D(7, 19) = 7 - 19 \pmod{26} = 14 \to \texttt{O}$
$\texttt{P} \to D(15, 19) = 15 - 19 \pmod{26} = 22 \to \texttt{W}$
$\texttt{H} \to D(7, 19) = 7 - 19 \pmod{26} = 14 \to \texttt{O}$
$\texttt{K} \to D(10, 19) = 10 - 19 \pmod{26} = 17 \to \texttt{R}$
$\texttt{E} \to D(4, 19) = 4 - 19 \pmod{26} = 11 \to \texttt{L}$
$\texttt{W} \to D(22, 19) = 22 - 19 \pmod{26} = 3 \to \texttt{D}$

So, we get `HELLOWORLD`!

Trivial attack

Ciphertext: `GJBFWJYMJNIJXTKRFWHM`
Plaintext: `BEWARETHEIDESOFMARCH`

Proposition

There exists a known-plaintext attack to get a total break on the Caesar cipher in constant time.

Trivial attack

Ciphertext: `GJBFWJYMJNIJXTKRFWHM`
Plaintext: `BEWARETHEIDESOFMARCH`

Proposition

There exists a known-plaintext attack to get a total break on the Caesar cipher in constant time.

Simply take the first letter of ciphertext and subtract the first letter of plaintext to get the key.

$k = \mathtt{G} - \mathtt{B} = 6 - 1 = 5.$

We can do better!

Brute force

Let's take our ciphertext again: `GJBFWJYMJNIJXTKRFWHM`.

Proposition

There exists a ciphertext-only attack to get a total break on the Caesar cipher in time linear in $|K|$.

Brute force

Let's take our ciphertext again: `GJBFWJYMJNIJXTKRFWHM`.

Proposition

There exists a ciphertext-only attack to get a total break on the Caesar cipher in time linear in $|K|$.

Try every key! This way, we don't need the plaintext at all. This is called **brute force**.

Caveat: in any decent cipher, the number of possible keys $|K|$ is likely too big (for popular modern ciphers at least on the order of $2^{128} \approx 10^{38}$) for us to try everything. Still, it's a good baseline to compare other attacks to.

Brute force

k	**Ciphertext value**
0	GJBFWJYMJNIJXTKRFWHM
1	HKCGXKZNKOJKYULSGXIN
2	ILDHYLAOLPKLZVMTHYJO
3	JMEIZMBPMQLMAWNUIZKP
4	KNFJANCQNRMNBXOVJALQ
5	LOGKBODROSNOCYPWKBMR
⋮	⋮
20	ADVZQDSGDHCDRNELZQBG
21	$\boxed{\text{BEWARETHEIDESOFMARCH}}$
22	CFXBSFUIFJEFTPGNBSDI
⋮	⋮

Frequency Analysis

Ciphertext:

```
CWIOFXVYQYFFGIPYXCZCQYLYUMSIOCZCWIOFXJLUSNIGIPYJL
USYLMQIOFXGIPYGYVONCUGWIHMNUHNUMNBYHILNBYLHMNULIZ
QBIMYNLOYZCRXUHXLYMNCHAKOUFCNSNBYLYCMHIZYFFIQCHNB
YZCLGUGYHNNBYMECYMULYJUCHNYXQCNBOHHOGVYLXMJULEMNB
YSULYUFFZCLYUHXYPYLSIHYXINBMBCHYVONNBYLYMVONIHYCH
UFFXINBBIFXBCMJFUWYMICHNBYQILFXNCMZOLHCMBXQYFFQCN
BGYHUHXGYHULYZFYMBUHXVFIIXUHXUJJLYBYHMCPYSYNCHNBY
HOGVYLCXIEHIQVONIHYNBUNOHUMMUCFUVFYBIFXMIHBCMLUHE
OHMBUEYXIZGINCIHUHXNBUNCUGBYFYNGYUFCNNFYMBIQCNYPY
HCHNBCMNBUNCQUMWIHMNUHNWCGVYLMBIOFXVYVUHCMBXUHXWI
HMNUHNXILYGUCHNIEYYJBCGMI
```

Frequency analysis

Proposition

There exists a ciphertext-only attack to get a total break on the Caesar cipher in constant time.

Frequency analysis

Proposition

There exists a ciphertext-only attack to get a total break on the Caesar cipher in constant time.

1. The most common letter in the English language is `E`.
2. The most common letter in the ciphertext is `Y`. It occurs 61 times, and the next most common letter, `N`, only occurs 45 times.
3. Assuming that `E` corresponds to `Y`, then our key is $K = 20$.

Frequency analysis is especially powerful because it can be applied to any type of **substitution cipher**, not just the Caesar cipher (more on this later).

Frequency

Voilà!

ICOULDBEWELLMOVEDIFIWEREASYOUIFICOULDPRAYTOMOVEPR
AYERSWOULDMOVEMEBUTIAMCONSTANTASTHENORTHERNSTAROF
WHOSETRUEFIXDANDRESTINGQUALITYTHEREISNOFELLOWINTH
EFIRMAMENTTHESKIESAREPAINTEDWITHUNNUMBERDSPARKSTH
EYAREALLFIREANDEVERYONEDOTHSHINEBUTTHERESBUTONEIN
ALLDOTHHOLDHISPLACESOINTHEWORLDTISFURNISHDWELLWIT
HMENANDMENAREFLESHANDBLOODANDAPPREHENSIVEYETINTHE
NUMBERIDOKNOWBUTONETHATUNASSAILABLEHOLDSONHISRANK
UNSHAKEDOFMOTIONANDTHATIAMHELETMEALITTLESHOWITEVE
NINTHISTHATIWASCONSTANTCIMBERSHOULDBEBANISHDANDCO
NSTANTDOREMAINTOKEEPHIMSO