CLASSICAL CIPHERS

Nitya Mani and Andy Chen

April 10, 2016

Stanford SPLASH 2016

INTRODUCTION

What is Cryptography?

Cryptography is the study of methods of secure communication between two parties.

Goals of cryptography:

- Two parties (Alice and Bob) want to send messages to each other
- They want to avoid the possibility of a malicious third party (Eve) understanding these messages

Cryptography is the battle between Alice and Bob on one side, and Eve on the other.

What is a Cipher?

Suppose Alice wishes to send HELLO to Bob.

Alice sends something so that Bob knows it means HELLO, but it looks like gibberish to Eve.

In a cipher, Alice modifies some/all of the letters in a pre-specified manner (encode the message), so that Bob knows how to recover the original text.

However, Eve will not know how the text was modified and will have to figure it out by (clever) guessing.

Some Terminology

Plaintext: the message Alice actually wants to send to Bob

Ciphertext: the text message Alice sends to Bob after applying the cipher

Protocol: the type of cipher used to encode a plaintext message into ciphertext

Key: Protocol specific information that determines encoding

THE CAESAR CIPHER

A Motivating Example

Alice wants to securely send Bob the message

```
cryptography is fun
```

To encode her message, Alice replaces every letter by the one that comes 3 places later in the alphabet with wraparound ($a \to D$, $b \to E$, ..., $w \to Z$, $x \to A$, $y \to B$, $z \to C$). Under this scheme, Alice's message becomes

```
FUBSWRJUDSKB LV IXQ
```

Alice can send this encoded message and Eve will be stumped.

However, since Bob knows how Alice encoded the message, Bob can replace each letter in the ciphertext (FUBSWRJUDSKB LV IXQ) with the letter three places *earlier* in the alphabet to find the actual message.

The Caesar Cipher

In a Caesar cipher, every letter is shifted over by *the same amount n*, so that Alice is replacing every letter with the letter that comes *n* letters further along in the alphabet.

Bob can decode a message sent by Alice encrypted using the Caesar cipher by replacing each letter in the ciphertext with the letter that comes *n* letters *earlier* in the alphabet.

Alice and Bob agree on *n* (the key) in advance at a cipher meeting.

Cracking the Caesar Cipher

Even if Eve doesn't know n in advance, she can quickly crack a Caesar cipher. Suppose Eve received the encoded message

```
FUBSWRJUDSKB LV IXQ
```

Eve then guessed (wrongly) that it was the Cæsar cipher obtained by shifting each letter by 6 instead of by 3. After attempting to decode, she would end up with the message

```
zovmqldoxmev fp crk
```

She should suspect that this is wrong, because those don't look like English words. Usually, only the correct shift will give a sensible answer, so she can try until she gets words back.

THE SUBSTITUTION CIPHER

A Better Cipher

Instead of shifting everything over by a fixed amount, we can have every letter stand for a different letter, at random perhaps.

Consider the following set of substitutions:

```
abcde fghij klmno pqrst uvwxy z
XHZRW FGPTY JOCAE ULNKQ IVSMD B
```

Hence our plaintext message

```
cryptography is fun
```

becomes the following ciphertext:

```
ZNDUQEGNXUPD TK FIA
```

Decryption

If Alice sends `ZNDUQEGNXUPD TK FIA`, Bob can decrypt it by looking up each letter in the table and replacing until he has decrypted the whole message.

He may find the task easier if he sorts the table by the ciphertext rather than plaintext character:

```
ABCDE FGHIJ KLMNO PQRST UVWXY Z
nzmyo fgbuk sqxrl htdwi pveaj c
```

In contrast to Bob, Eve has difficulty deciphering the message since she doesn't know which letter substitutions were used.

A New Message

Suppose Alice wishes to send the new message

```
i have a truly marvelous demonstration of this
proposition
```

After Alice encodes it using the substitution table, Alice sends the ciphertext

```
T PXVW X QNIOD CXNVWOEIK RWCEAKQNXQTEA EF QPTK
UNEUEKTQTEA
```

How can Eve try to make progress on cracking this code without the table?

Cracking the Substitution Cipher: 1-Letter Words

There are only two 1-letter words in English: “a” and “I.” So, “T” and “X” must be “a” and “i,” in some order.

Suppose Eve guesses correctly and then makes the substitutions. She is left with:

```
i PaVW a QNIOD CaNVWOEIK RWCEAKQNaQiEA EF QPiK
                        UNEUEKiQiEA
```

Cracking the Substitution Cipher: Bigrams

How about that 2-letter word? (It doesn't contain an "a" or "I.")

There are several reasonable possibilities: "of", "or", "be" ... Suppose she correctly guesses that "EF" is supposed to be "of." Then, making those substitutions, we have

```
i PaVW a QNIOD CaNVWOoIK RWCoAKQNaQioA of QPiK
                          UNoUoKiQioA
```

What next?

Cracking the Substitution Cipher: Suffixes

There are two words that end with “QioA.” What’s a common ending for words of this form? The logical choices are “sion” and “tion.” Both end with “n,” so let’s replace “A” with “n.” Eve gets

```
i PaVW a QNIOD CaNVWOoIK RWConKQNaQion of QPiK
                    UNoUoKiQion
```

That wasn’t a huge help. However, let’s try replacing “Q” with “t” (which we can guess). Then Eve gets

```
i PaVW a tNIOD CaNVWOoIK RWConKtNation of tPiK
                    UNoUoKition
```

Cracking the Substitution Cipher: Uniqueness

We can use a crossword puzzle solver to see that there is exactly one word of the form `??o?o?ition` namely “proposition.’ If this is right, we can fill in lots of stuff!

```
i PaVW a trIOD CarVWOoIs RWConstration of tPis
                        proposition
```

Remember, we can only use each letter once! This is often useful in eliminating possibilities.

Cracking the Substitution Cipher: Word Guessing

From here, we make some guesses: `t?is` is probably “this” and `???onstration` must be “demonstration” since we have filled in so many letters already. These substitutions give

```
i haVe a trIOD marVeOoIs demonstration of this
proposition
```

Can we take it from here and guess words?

Cracking the Substitution Cipher: Context Clues

`marVeOoIs` → “marvelous”

```
i have a trulD marvelous demonstration of this
                        proposition
```

`trulD` → “truly”, `haVe` → “have”

```
i have a truly marvelous demonstration of this
                        proposition
```

And we are done!

Decipherment Demo

What if our message doesn't have spaces?

- Much harder to find specific words, like "I"
- Letter Frequency - look for the cipher symbols that appear the most often in longer messages
  - 'E' comprises of about 12% of English text, followed by 'T', 'O', and 'A'
- N-Gram Analysis - pairs of N letters
  - "THE" contains the bigrams "TH" and "HE"
  - similar frequency analysis of N-grams in longer messages
- Demonstration using machine learning: `https://youtu.be/orcoYJrCorE`

Application to Machine Translation

How do we automatically translate from one language to another?

IBM Alignment Models

- Given one sentence in Spanish and another sentence in English that have the same meaning
- “Align” words that roughly mean the same thing
- Decide which words (like “the” or “la”) that need to be added or deleted
- Example: “la casa verde” and “the green house”

Questions?