The Shift Cipher

lecture 01

Classical Cryptography

lecture 01

Motivation

- Allows us to "ease into things. . . ,"
- Shows why unprincipled approaches are dangerous (unprincipled means not-rigorous, not immoral)
- Illustrates why things are more difficult than they may appear

Alice, Bob, and Eve

- Alice sends a message to Bob in code.
- Eve overhears it.
- We want Eve to not be able to decode it.

This can mean one of two things:

- Eve does not have enough information to decode it. So even if Eve had unlimited computing power she could not decode.
- Assuming Eve can't Factor quickly (or some other function) then Eve cannot break the code.

The First Step in Any Cipher-Spaces

I want to encode

Cryptography is an important part of security

Spaces give away information! For example, SHIFT-BY-1 yields:

Dszuphsbqiz jt bo jnqpsubou qbsu pg tfdvsjuz

Without any fancy math Eve knows that the second and third word are two letters long. Thats information she can use!

What to do?

The First Step in Any Cipher-Blocks of Five

I want to encode

Cryptography is an important part of security

Break it up into blocks of 5:

Cryto graph yisan impor tantp artof secur ity

However you code it, spaces will not give anything away.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.
   Get rid of all punctuation.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.
   Get rid of all punctuation.
3. What to do about numbers?

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.
   Get rid of all punctuation.
3. What to do about numbers?
   Just like letters- alphabet is 36 characters

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.
   Get rid of all punctuation.
3. What to do about numbers?
   Just like letters- alphabet is 36 characters
   More generally, set your mod equal to your alphabet size.

The First Step in Any Cipher-Other Issues

I want to encode

Are my TAs for CMSC/MATH 456 awesome? YES!

1. Capital and small letters leak information.
   Map everything to Capitals.
2. Punctuation leaks information.
   Get rid of all punctuation.
3. What to do about numbers?
   Just like letters- alphabet is 36 characters
   More generally, set your mod equal to your alphabet size.

Note: In this class we will use 26-letter English only.

The Shift Cipher

lecture 01

The Shift Cipher

- Consider encrypting English text
- associate 'a' with 0; 'b' with 2; . . . ; 'z' with 25
- $k \in \mathcal{K} = \{0, \ldots, 25\}$ (or could think of $k \in \{a, \ldots, z\}$)
- To encrypt using key k, shift every letter of the plaintext by k positions (with wraparound)
- Decryption just does the reverse

```
 hello world
+22222 22222
=jgnnq yqtnf
```

Modular arithmetic

- $x \equiv y \pmod{N}$ if and only if N divides $x - y$.
- $[x \bmod N]$ = the remainder when x is divided by N.
  - i.e. the unique value $y \in \{0, \ldots, N-1\}$ such that $x \equiv y \pmod{N}$.
- $25 \equiv 35 \pmod{10}$
- $25 \neq [35 \bmod 10]$
- $5 = [35 \bmod 10]$

The Shift Cipher, Formally

- $\mathcal{M} = \{$all texts in lowercase English alphabet$\}$
  All arithmetic mod 26.
- Choose uniform $k \in \{0, \ldots, 25\}$
- Encode $(m_1 \ldots m_t)$ as $(m_1 + k, \ldots m_t + k)$
- Decode $(c_1 \ldots c_t)$ as $(c_1 - k, \ldots c_t - k)$
- Can verify that correctness holds.

Is the Shift Cipher Secure?

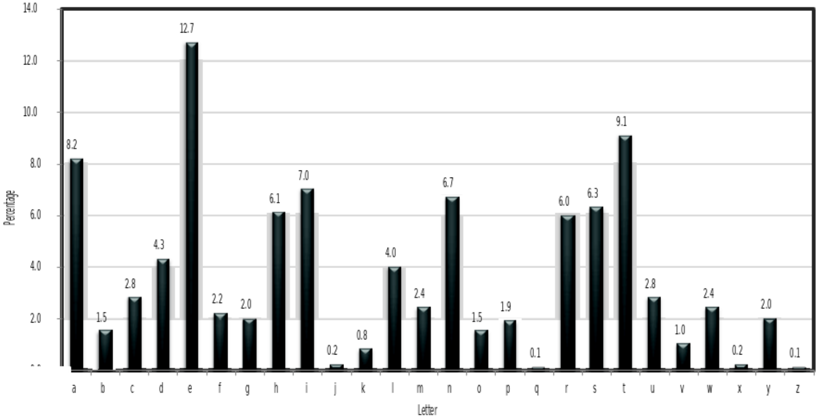
- No – only 26 possible keys!
  - Given a ciphertext, try decrypting with every possible key
  - Only one possibility will “make sense”
- Example of a “brute-force” or “exhaustive-search” attack

Example

- Ciphertext `uryyb jbeyq`
- Try every possible key...
  - tqxxa iadxp
  - spwwz hzcwo
  - ...
  - hello world

Question: We can tell that hello world is correct but how can a computer do that. Can we mechanize the process of picking out the right one?

Letter Frequencies



Use Letter Freqs to Test "Looks Like English"

Let T be a long text of normal English.
Let $\vec{f}$ be the freq vector of English. The components are all between 0 and 1 and add up to 1.
We assume freq vector of T is approx $\vec{f}$.

- One can compute that

$$\vec{f} \cdot \vec{f} \approx 0.065$$

- Let $s \in \{1, \ldots, 25\}$. Let T_s be the text shifted by s. Let $\vec{g}$ be the freq vector for T_s. One can compute that

$$\vec{f} \cdot \vec{g} \leq\approx 0.038$$

Is English

We describe a way to tell if a text Is English that we will use throughout this course.
Let $\vec{f}$ be the freq vector for English.

1. Input(T) a text
2. Compute $\vec{g}$, the freq vector for T
3. Compute $\vec{g} \cdot \vec{f}$. If ≈ 0.065 then output YES, else NO

Cracking Shift Cipher

- Given T a long text that you KNOW was coded by shift.
- For $s = 0$ to 25
  - Create T_s which is T shifted by s.
  - If Is English(T_s)=YES then output T_s and stop. Else try next value of i.

Note: No Near Misses. There will not be two values of s that are both close to 0.065.

Pedagoical Note: Would normally have written Key instead of Note but the word Key is important in crypto so I can't use it to say something is important. Oh Well.

A Note on Cracking Shift Cipher

In the last slide we tried *all* shifts in order.
Can do better:

- Given T a long text that you KNOW was coded by shift.
- Find frequencies of all letters, form vector $\vec{f}$
- Sort vector. So most common letter is σ_1, next is σ_2, etc.
- For $i = 0$ to 25
  - Create T_s which is T shifted as if σ_i maps to e.
  - Compute $\vec{g}$, the freq vector for T_s
  - Compute $\vec{g} \cdot \vec{f}$. If ≈ 0.065 then stop: T_s is your text. Else try next value of s.

Note: Quite likely to succeed in the first try, or at least very early.

Kerckhoffs's principle

We made the comment We KNOW that SHIFT was used. More generally we use this principle.

- *The encryption scheme* is not secret
  - Eve knows the encryption scheme
  - The only secret is the key
  - The key must be chosen at random; kept secret
- Some arguments in favor of this principle
  - Easier to keep *key* secret than *algorithm*
  - Easier to change *key* than to change *algorithm*
  - Standardization
    - Ease of deployment
    - Public validation

Is this cipher secure if we are transmitting numbers?

If Alice sends Bob a Document in English via Byte-Shift then insecure!

What if Alice sends Bob a credit card number? Discuss

Is this cipher secure if we are transmitting numbers?

If Alice sends Bob a Document in English via Byte-Shift then insecure!

What if Alice sends Bob a credit card number? Discuss
Credit Card Numbers also have patterns:

1. Visa cards always begin with 4
2. American Express always begins 34 or 37
3. Mastercard starts with 51 or 52 or 53 or 54.

Upshot: If Eve knows what kind of information is being transmitted (English, Credit Card Numbers, numbers on checks) she can use this to make any cipher with a small key space insecure.