Admin and The Shift Cipher

lecture 01

Crypto Is. . .

- Crypto is amazing
    - Can do things that initially seem impossible
- Crypto is important
    - It impacts us every day
- Crypto is fun!
    - Deep theory
    - Attackers' mindset

Crypto Is Not. . .

Crypto Is Not...

James Bond

Crypto Is Not...

James Bond

- James Bond is Fictional.

Crypto Is Not. . .

James Bond

- James Bond is Fictional.
- James Bond is a drunk
  See article on course website: *License to Swill*.

Crypto Is Not...

James Bond

- James Bond is Fictional.
- James Bond is a drunk
  See article on course website: *License to Swill*.
- James Bond's Villains are stupid
  See video on course website *Goodbye Mr. Bond*.

Crypto Is Not...

James Bond

- James Bond is Fictional.
- James Bond is a drunk
  See article on course website: *License to Swill*.
- James Bond's Villains are stupid
  See video on course website *Goodbye Mr. Bond*.

Seriously: Spying depends a lot more on Math than on Fancy Weapons.

Necessary administrative stuff

- Course webpage: https://www.cs.umd.edu/users/gasarch/COURSES/456/F19/index.html
    - Prerequisites/information posted there
    - Syllabus posted there
    - HWs posted there
    - Announcements posted there
    - Midterm already scheduled- Oct 28 in class.

Necessary administrative stuff

- Gradescope: hw will be posted there.
- Gradescope: hw will be graded there.
- Regrade Requests due within a week of the HW being graded.
- Grades on Elms.
- Piazza is great for asking questions.

TAs

- Nathan Grammel
- Marina Knittel
- Erik Metz
- Justin Hontz

What You Need For This Class

- Mathematical prerequisites
  - Discrete math, probability, modular arithmetic
- Requires mathematical maturity
  - Proofs, abstraction

What You Need For This Class

- CS prerequisites
    - Binary, hex, pseudocode, algorithms, big-O notation
- Programming assignments
    - Hard part should not be the programming, but the thought behind it
    - Flexibility in choice of language

How to Get the Most Out of This Class

1. Read notes and slides before class.
   Note: On Slide Website it says on some line
   WHAT IS BELOW IS STILL A WORK IN PROGRESS.
   Should not read slides that are below that line.

2. Ask questions on Piazza and/or bring questions to class

3. This course will be taped so can catch up or review. Caution:

   3.1 If cut class and DO watch videos in sync, fine.
   3.2 If cut class and INTEND to watch videos in sync, not fine.
   3.3 Tape might not always work.

HWs/exams

- HWs most weeks.
- Due Monday before class begins.
- Dead Cat Policy: Can submit HW Wed before class without penalty
- WARNING: YOU have already been given an extension, HW solutions will be posted on Wed, so NO extensions past that.
- We will keep track of your lateness NOT for grade, but for recommendation letters.

- In-class midterm and final

Textbook

Required None. There will be notes, slides, and recording of lecture online.

Laptops/electronics

- No laptops/electronics policy
  - Distracting to you
  - Distracting to others
- If you feel you need an exception, talk to me

How to contact Prof or TAs

- Prof email: `gasarch@cs.umd.edu`
- Please put “CMSC456” in subject line
- Prof Office hours MW 1-2, 3:30-5:00 or by Apt.
- Prof around a lot outside of office hours, feel free to drop in.
- TA’s - email and office hours will be on syllabus by Aug 29, 2019.
- Piazza

Classical VS Modern cryptography

Classical: (1900 BCE?–1975)

1. More of an art. Not much Mathematics.
2. WW II: They used people good at crossword puzzles (see course website for an article on this).
3. Turing and others brought math into it, but not much math compared compared to Modern

Modern: (1976-today)

1. Lots of Math. Lots of Rigor.
2. The notion of Provably Secure important.

Note: The cutoff of 1975–1976 is approximate.

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
- Simple examples of what will later be advanced concepts.

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. This is Information-Theoretic Security.
- Assuming Eve can't Factor quickly (or some other computational limitation) then Eve cannot break the code. This is Computational-Security.

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:

Dszquphsbiz 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, we will take into account 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, we will take into account alphabet size.

Note: In this class we will mostly use 26-letter English only unless otherwise noted.

The Shift Cipher

lecture 01

The Shift Cipher

- Consider encrypting English text
- associate ‘a’ with 0; ‘b’ with 1; . . . ; ‘z’ with 25
- $s \in \{0, \ldots, 25\}$ (or could think of $s \in \{a, \ldots, z\}$)
- To encrypt using key s, shift every letter of the plaintext by s positions (with wraparound)
- Decryption just does the reverse

$$\begin{aligned} &\texttt{hello world} \\ +&\texttt{22222 22222} \\ =&\texttt{jgnnq yqtnf} \end{aligned}$$

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

- $\mathcal{M} = \{\text{all texts in lowercase English alphabet}\}$
  $\mathcal{M}$ for Message space.
  All arithmetic mod 26.

- Choose uniform $s \in \mathcal{K} = \{0, \ldots, 25\}$. $\mathcal{K}$ for Keyspace.

- Encode $(m_1 \ldots m_t)$ as $(m_1 + s, \ldots m_t + s)$

- Decode $(c_1 \ldots c_t)$ as $(c_1 - s, \ldots c_t - s)$

- 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

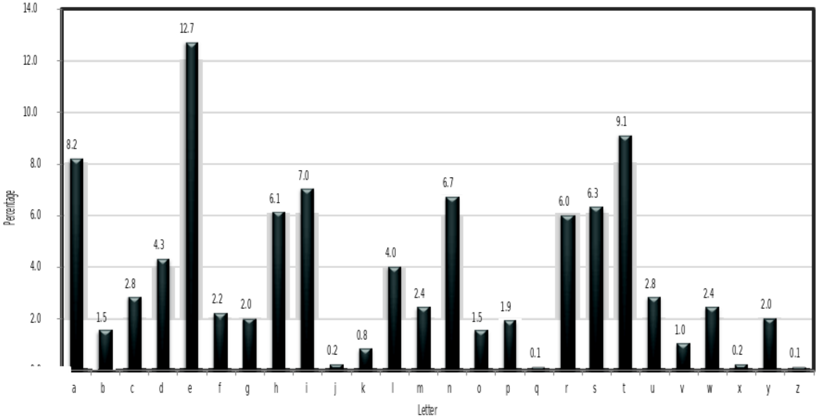
- Ciphertext `uryyb jbeyq`
- Try every possible key. . .
  - tqxxa iadxp
  - spwwz hzcwo
  - . . .
  - hello world

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

Note: What if $\vec{g} \cdot \vec{f} = 0.0630$? If coded using Shift then this will never happen. Other ciphers may need more care.

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 s.

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

Pedagogical 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.

What if only transmit one letter?

Odd Situation: What if message is only one letter long?
Discuss: Can Eve crack a one-letter message?

What if only transmit one letter?

Odd Situation: What if message is only one letter long?
Discuss: Can Eve crack a one-letter message? No

What if only transmit one letter?

Odd Situation: What if message is only one letter long?
Discuss: Can Eve crack a one-letter message? No
Discuss: How to define uncrackable?

What if only transmit one letter?

Odd Situation: What if message is only one letter long?
Discuss: Can Eve crack a one-letter message? No
Discuss: How to define uncrackable?
Message is m. What is seen is c.

Before Eve sees the message what does she know?

What if only transmit one letter?

Odd Situation: What if message is only one letter long?
Discuss: Can Eve crack a one-letter message? No
Discuss: How to define uncrackable?
Message is m. What is seen is c.

Before Eve sees the message what does she know?
Assume Eve knows $p_0, \ldots, p_{25}$ where

$$\Pr(m = 0) = p_0, \Pr(m = 1) = p_1, \ldots, \Pr(m = 25) = p_{25}.$$

If after Eve sees the message she knows

$$\Pr(m = 0) = p_0, \Pr(m = 1) = p_1, \ldots, \Pr(m = 25) = p_{25}$$

then Eve has learned *nothing*.

We Need Conditional Probability

Conditional probability: Probability that one event occurs, *given that some other event occurred*

Notation: $\Pr[A|B]$.

Formal Definition: Notation: $\Pr[A|B] = \frac{\Pr(A \cap B)}{\Pr(B)}$.
Intuition: $\Pr[A|B] = \frac{\Pr(A \cap B)}{\Pr(B)}$ is saying that the entire space is now $\Pr(B)$. Within that space what is the prob of A happening? Its $\Pr(A \cap B)$.

Examples of Conditional Probability

Justin rolls two dice d_1, d_2 and takes the sum s. What is the $\Pr(s = 5)$?

Examples of Conditional Probability

Justin rolls two dice d_1, d_2 and takes the sum s. What is the $\Pr(s = 5)$? $\frac{1}{9}$.

Examples of Conditional Probability

Justin rolls two dice d_1, d_2 and takes the sum s. What is the $\Pr(s = 5)$? $\frac{1}{9}$.

What if you know that d_1?

Examples of Conditional Probability

Justin rolls two dice d_1, d_2 and takes the sum s. What is the $\Pr(s = 5)$? $\frac{1}{9}$.

What if you know that d_1?

$\Pr(s = 5 | d_1 = 1) = \frac{\Pr(s=5 \wedge d_1=1)}{\Pr(d_1=1)} = \frac{1/36}{1/6} = \frac{1}{6}$.

$\Pr(s = 5 | d_1 = 2) = \frac{\Pr(s=5 \wedge d_1=2)}{\Pr(d_1=2)} = \frac{1/36}{1/6} = \frac{1}{6}$.

$\Pr(s = 5 | d_1 = 3) = \frac{\Pr(s=5 \wedge d_1=3)}{\Pr(d_1=3)} = \frac{1/36}{1/6} = \frac{1}{6}$.

$\Pr(s = 5 | d_1 = 4) = \frac{\Pr(s=5 \wedge d_1=4)}{\Pr(d_1=4)} = \frac{1/36}{1/6} = \frac{1}{6}$.

$\Pr(s = 5 | d_1 = 5) = \frac{\Pr(s=5 \wedge d_1=5)}{\Pr(d_1=5)} = \frac{0}{1/6} = 0$.

$\Pr(s = 5 | d_1 = 6) = \frac{\Pr(s=5 \wedge d_1=6)}{\Pr(d_1=6)} = \frac{0}{1/6} = 0$.

Definition of Secure

Assume we have a crypto system. m will be a message and c will be what is sent. If the following holds then the system is *secure*.

$$(\forall m, a, b, c)[\Pr(m = a | c = b) = \Pr(m = a)].$$

So seeing the b does not help Eve at all.

Definition of Secure

Assume we have a crypto system. m will be a message and c will be what is sent. If the following holds then the system is *secure*.

$$(\forall m, a, b, c)[\Pr(m = a | c = b) = \Pr(m = a)].$$

So seeing the b does not help Eve at all.
Is this info-theoretic security or comp-security? Discuss

Definition of Secure

Assume we have a crypto system. m will be a message and c will be what is sent. If the following holds then the system is *secure*.

$$(\forall m, a, b, c)[\mathrm{Pr}(m = a | c = b) = \mathrm{Pr}(m = a)].$$

So seeing the b does not help Eve at all.
Is this info-theoretic security or comp-security? Discuss
Info-Theoretic: If Eve has unlimited computing power she still learns nothing.

Definition of Secure

Assume we have a crypto system. m will be a message and c will be what is sent. If the following holds then the system is *secure*.

$$(\forall m, a, b, c)[\Pr(m = a | c = b) = \Pr(m = a)].$$

So seeing the b does not help Eve at all.
Is this info-theoretic security or comp-security? Discuss
Info-Theoretic: If Eve has unlimited computing power she still learns nothing.
Slides Title Should have Been:

Definition of Info-Theoretic Security

.

One Letter Shift is Uncrackable! Eve's View

Example

1. Before message is send Eve knows $\Pr(m = i) = p_i$.
2. Eve sees that Alice sends Bob the number 12.
3. Lets see what Eve knows.

One Letter Shift is Uncrackable! Final

Before seeing 12 Eve knew that $\Pr(m = 17) = p_{17}$. She sees 12. what is prob that $m = 17$? Let c be what Eve sees.

$$\Pr(m = 17 | c = 12) = \frac{\Pr(m = 17 \wedge c = 12)}{\Pr(c = 12)}$$

$\Pr(m = 17 \wedge c = 12) = Pr(m = 17 \wedge s = 21) = p_{17} \times \frac{1}{26}$.
$\Pr(c = 12) =$

$$p_0 \Pr(s = 12) + \cdots + p_{12} \Pr(s = 0) + p_{13} \Pr(s = 25) + \cdots p_{25} \Pr(s = 13)$$

$$= \frac{1}{26}(p_0 + \cdots + p_{25}) = \frac{1}{26}$$

SO

$$\Pr(m = 17 | c = 12) = \frac{\Pr(m = 17 \wedge c = 12)}{Pr(c = 12)} = p_{17} \times \frac{1}{26} / \frac{1}{26} = p_{17}.$$

Upshot: $\Pr(m = 17 | c = 11) = p_{17}$. So Eve has learned nothing!

One Letter Shift is Uncrackable! Final

Before seeing 12 Eve knew that $\mathrm{Pr}(m = 17) = p_{17}$. She sees 12. what is prob that $m = 17$? Let c be what Eve sees.

$$\mathrm{Pr}(m = 17|c = 12) = \frac{\mathrm{Pr}(m = 17 \wedge c = 12)}{\mathrm{Pr}(c = 12)}$$

$\mathrm{Pr}(m = 17 \wedge c = 12) = Pr(m = 17 \wedge s = 21) = p_{17} \times \frac{1}{26}$.

$\mathrm{Pr}(c = 12) =$

$$p_0\mathrm{Pr}(s = 12) + \cdots + p_{12}\mathrm{Pr}(s = 0) + p_{13}\mathrm{Pr}(s = 25) + \cdots p_{25}\mathrm{Pr}(s = 13)$$

$$= \frac{1}{26}(p_0 + \cdots + p_{25}) = \frac{1}{26}$$

SO

$$\mathrm{Pr}(m = 17|c = 12) = \frac{\mathrm{Pr}(m = 17 \wedge c = 12)}{Pr(c = 12)} = p_{17} \times \frac{1}{26} / \frac{1}{26} = p_{17}.$$

Upshot: $\mathrm{Pr}(m = 17|c = 11) = p_{17}$. So Eve has learned nothing!

Is 2-letter Shift Uncrackable?

Is 2-letter Shift Uncrackable? Discuss.

Is 2-letter Shift Uncrackable?

Is 2-letter Shift Uncrackable? Discuss.
No.
If Eve sees AB then she knows that the original message was one of

$$\{AB, BC, CD, \ldots, YZ, ZA\}$$

So Eve has learned something.

Can Two 1-Letter Messages Leak Information?

Can Two 1-Letter Messages using the same shift Leak Information?

Can Two 1-Letter Messages Leak Information?

Can Two 1-Letter Messages using the same shift Leak Information?
Yes

Scenario:

In clear: Is Erik a double agent working for the Klingons?
The answer comes via a shift cipher: A (which is either Y or N)

In clear: Is Erik a double agent working for the Romulans?
The answer comes via a shift cipher: A (which is either Y or N)

Can Two 1-Letter Messages Leak Information?

Can Two 1-Letter Messages using the same shift Leak Information?
Yes
Scenario:
In clear: Is Erik a double agent working for the Klingons?
The answer comes via a shift cipher: A (which is either Y or N)

In clear: Is Erik a double agent working for the Romulans?
The answer comes via a shift cipher: A (which is either Y or N)

Eve knows Erik is working for either both or neither.

Eve Can Tell if Two Message Are Same or Not

Issue: If Eve sees two message, will know if they are the same or different.

Does this leak information: Discuss

Eve Can Tell if Two Message Are Same or Not

Issue: If Eve sees two message, will know if they are the same or different.

Does this leak information: Discuss

What to do about this? Discuss

Eve Can Tell if Two Message Are Same or Not

Issue: If Eve sees two message, will know if they are the same or different.

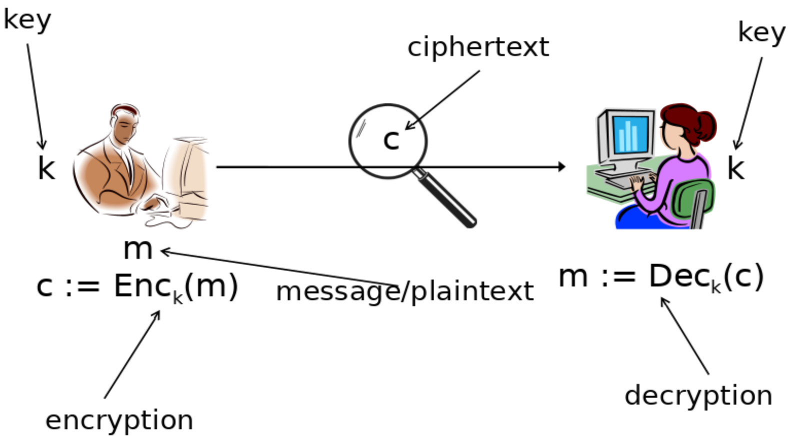
Does this leak information: Discuss

What to do about this? Discuss

For Now Nothing Will come back to this issue after a few more ciphers.

For Now A lesson in how even defining security and leak must be done carefully.

Private-key encryption



key

k

m

$c := \mathrm{Enc}_k(m)$

encryption

ciphertext

c

message/plaintext

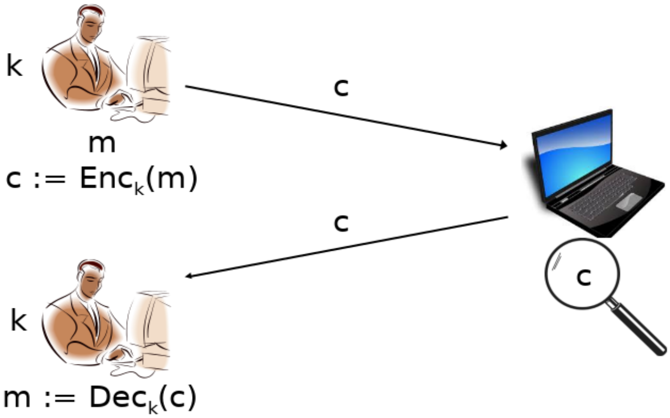
key

k

$m := \mathrm{Dec}_k(c)$

decryption

Private-key encryption



k

m

$c := \mathrm{Enc}_k(m)$

c

c

c

k

$m := \mathrm{Dec}_k(c)$

Private-key encryption

- A *private-key encryption scheme* is defined by a message space $\mathcal{M}$ and algorithms (Gen, Enc, Dec):
  - Gen (key generation algorithm): outputs $k \in \mathcal{K}$
    (For SHIFT this is $k \in \{0, \ldots, 25\}$. Should 0 be included?)
  - Enc (encryption algorithm): takes key k and message $m \in \mathcal{M}$ as input; outputs ciphertext c

    $$c \leftarrow Enc_k(m)$$

    (For SHIFT this is $\mathrm{Enc}(m_1, \ldots, m_n) = (m_1 + k, \ldots, m_n + k)$.)
  - Dec (decryption algorithm): takes key k and ciphertext c as input; outputs m or "error"

    $$m := Dec_k(c)$$

    (For SHIFT this is $\mathrm{Dec}(c_1, \ldots, c_n) = (c_1 - k, \ldots, c_n - k)$.)

    $\forall k$ output by Gen $\forall m \in \mathcal{M}, Dec_k(Enc_k(m)) = m$

(For SHIFT this is $(m + k) - k = m$)