BILL, RECORD LECTURE!!!!

BILL RECORD LECTURE!!!

IF YOU DIDN"T GET THE EMAIL

LET ME KNOW.
I send an email to the class on Aug 31 at night asking you to respond. If you DID NOT get that email then see me TODAY after class so I can get the email you want me to use and add you to the list.

GRADESCOPE

Gradescope Code: P56D84

The Shift Cipher (cont)

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

1. $f_E \cdot f_E \sim 0.065$.
2. For $1 \le i \le 25$, f_i is English shifted by i. $f_E \cdot f_i \sim 0.035$.

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

1. $f_E \cdot f_E \sim 0.065$.
2. For $1 \leq i \leq 25$, f_i is English shifted by i. $f_E \cdot f_i \sim 0.035$.
3. Find correct shift i by seeing which $f_E \cdot f_i$ is ~ 0.065.

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

1. $f_E \cdot f_E \sim 0.065$.
2. For $1 \le i \le 25$, f_i is English shifted by i. $f_E \cdot f_i \sim 0.035$.
3. Find correct shift i by seeing which $f_E \cdot f_i$ is ~ 0.065.
4. Only one of the dot products will be close to 0.065.

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

1. $f_E \cdot f_E \sim 0.065$.
2. For $1 \leq i \leq 25$, f_i is English shifted by i. $f_E \cdot f_i \sim 0.035$.
3. Find correct shift i by seeing which $f_E \cdot f_i$ is ~ 0.065.
4. Only one of the dot products will be close to 0.065.

Note For the Shift Cipher we can easily derive the numbers (0.065,0.035). Because of the big gap, shift is crackable.

A Caveat on Cracking The Shift Cipher

We used the following reasoning:

1. $f_E \cdot f_E \sim 0.065$.
2. For $1 \le i \le 25$, f_i is English shifted by i. $f_E \cdot f_i \sim 0.035$.
3. Find correct shift i by seeing which $f_E \cdot f_i$ is ~ 0.065.
4. Only one of the dot products will be close to 0.065.

Note For the Shift Cipher we can easily derive the numbers (0.065,0.035). Because of the big gap, shift is crackable.

For more complicated ciphers we may need more sophisticated IS-ENGLISH programs and the parameters may be harder to fine-tune.

One Way to Find Freqs

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, …, number of z's. τ gets first char of T.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.
   2.1 τ is current character.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.
   2.1 τ is current character.
   2.2 If $\tau = a$ then $N_a \leftarrow N_a + 1$. If $\tau = b$ then $N_b \leftarrow N_b + q$. Etc.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.

    2.1 τ is current character.

    2.2 If $\tau = a$ then $N_a \leftarrow N_a + 1$. If $\tau = b$ then $N_b \leftarrow N_b + q$. Etc.

    2.3 τ gets next char.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.
   2.1 τ is current character.
   2.2 If $\tau = a$ then $N_a \leftarrow N_a + 1$. If $\tau = b$ then $N_b \leftarrow N_b + q$. Etc.
   2.3 τ gets next char.
3. For $\sigma = a$ to z, $f_\sigma = \frac{N_\sigma}{N}$.

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, ..., number of z's. τ gets first char of T.
2. Repeat until end of T.
   - 2.1 τ is current character.
   - 2.2 If $\tau = a$ then $N_a \leftarrow N_a + 1$. If $\tau = b$ then $N_b \leftarrow N_b + q$. Etc.
   - 2.3 τ gets next char.
3. For $\sigma = a$ to z, $f_\sigma = \frac{N_\sigma}{N}$.

Is this a good approach?

One Way to Find Freqs

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N. We have variables $N_a, \ldots, N_z$ which will keep track of the number of a's, …, number of z's. τ gets first char of T.
2. Repeat until end of T.
   2.1 τ is current character.
   2.2 If $\tau = a$ then $N_a \leftarrow N_a + 1$. If $\tau = b$ then $N_b \leftarrow N_b + q$. Etc.
   2.3 τ gets next char.
3. For $\sigma = a$ to z, $f_\sigma = \frac{N_\sigma}{N}$.

Is this a good approach? No– we spend 26 steps on every letter!

A Much Better Way

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.
3. Repeat until end of T.

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.
3. Repeat until end of T.
   3.1 τ is current number.

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.
3. Repeat until end of T.
   - 3.1 τ is current number.
   - 3.2 $N[\tau] = N[\tau] + 1$

A Much Better Way

1. Input T, a text of letters in $\{a, \ldots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.
3. Repeat until end of T.
   3.1 τ is current number.
   3.2 $N[\tau] = N[\tau] + 1$
4. For $i = 0$ to 25, $f[i] = N[i]/N$.

A Much Better Way

1. Input T, a text of letters in $\{a, \dots, z\}$ of length N.
2. Convert T into numbers, a is 0, b is 1, etc. For $0 \leq i \leq 25$, $N[i]$ will be the number of i's.
3. Repeat until end of T.
    3.1 τ is current number.
    3.2 $N[\tau] = N[\tau] + 1$
4. For $i = 0$ to 25, $f[i] = N[i]/N$.

This is much faster.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss

Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss
Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss
Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.
2. For $0 \le i \le 25$ find f_i, the freq vector of the T shifted by i.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss
Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.
2. For $0 \le i \le 25$ find f_i, the freq vector of the T shifted by i.
3. Compute $i_0 = \max_{0 \le i \le 25} f_E \cdot f_i$. i_0 is shift to decode with.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss
Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.
2. For $0 \le i \le 25$ find f_i, the freq vector of the T shifted by i.
3. Compute $i_0 = \max_{0 \le i \le 25} f_E \cdot f_i$. i_0 is shift to decode with.

Note Didn't need the parameters 0.065, 0.035 to do this.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss

Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.
2. For $0 \le i \le 25$ find f_i, the freq vector of the T shifted by i.
3. Compute $i_0 = \max_{0 \le i \le 25} f_E \cdot f_i$. i_0 is shift to decode with.

Note Didn't need the parameters 0.065, 0.035 to do this.

Downside Since we knew the parameters 0.065, 0.035 we knew there was a big gap. We knew there would be no close calls. If we do not know these kind of parameters then we are not as confident.

Can Crack Shift If Did Not Know Parameters 0.065, 0.035?

Discuss
Important point is that $f_E \cdot f_E$ is BIG, $f_E \cdot f_i$ SMALL. Do not need to know HOW BIG, HOW SMALL.

1. Input(T). T is a text that has been coded by the shift cipher.
2. For $0 \le i \le 25$ find f_i, the freq vector of the T shifted by i.
3. Compute $i_0 = \max_{0 \le i \le 25} f_E \cdot f_i$. i_0 is shift to decode with.

Note Didn't need the parameters 0.065, 0.035 to do this.

Downside Since we knew the parameters 0.065, 0.035 we knew there was a big gap. We knew there would be no close calls. If we do not know these kind of parameters then we are not as confident.

But if we have a few candidates for IS-ENGLISH there may be other ways to pick out the real one.

Variants of the Shift Cipher

What About Texts With Numbers?

We have discussed English texts with $\Sigma = \{a, \ldots, z\}$.

What About Texts With Numbers?

We have discussed English texts with $\Sigma = \{a, \ldots, z\}$.

What if the text has numbers in it? Examples:

What About Texts With Numbers?

We have discussed English texts with $\Sigma = \{a, \ldots, z\}$.

What if the text has numbers in it? Examples:

1. Financial Documents. $\Sigma = \{a, b, \ldots, z, 0, \ldots, 9\}$.

What About Texts With Numbers?

We have discussed English texts with $\Sigma = \{a, \ldots, z\}$.

What if the text has numbers in it? Examples:

1. Financial Documents. $\Sigma = \{a, b, \ldots, z, 0, \ldots, 9\}$.
2. Math books such as:
   `https://www.amazon.com/Mathematical-Muffin-Morsels-Problem-Mathematics/dp/9811215979/ref=sr_1_2?dchild=1&keywords=gasarch&qid=1593879329&sr=8-2`

$$\Sigma = \{a, \ldots, z, 0, \ldots, 9, +, \times, -, \div, =, \equiv, <, >, \cap, \cup, \emptyset\}$$

Include other symbols depending on the branch of math. E.g., $\wedge, \vee$ for logic.

What About Texts With Numbers?

We have discussed English texts with $\Sigma = \{a, \ldots, z\}$.

What if the text has numbers in it? Examples:

1. Financial Documents. $\Sigma = \{a, b, \ldots, z, 0, \ldots, 9\}$.
2. Math books such as:
   `https://www.amazon.com/Mathematical-Muffin-Morsels-Problem-Mathematics/dp/9811215979/ref=sr_1_2?dchild=1&keywords=gasarch&qid=1593879329&sr=8-2`

   $$\Sigma = \{a, \ldots, z, 0, \ldots, 9, +, \times, -, \div, =, \equiv, <, >, \cap, \cup, \emptyset\}$$

   Include other symbols depending on the branch of math. E.g., $\wedge, \vee$ for logic.

What to do? Find distribution of alphabet for these types of docs. Write code sim to **Is-English** and try all shifts.

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

It turns out that the frequency of numbers and symbols was around that of q, z so had no real affect. Parameters are about the same.

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

It turns out that the frequency of numbers and symbols was around that of q, z so had no real affect. Parameters are about the same.

Students Darn! We didn't find anything interesting.

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

It turns out that the frequency of numbers and symbols was around that of q, z so had no real affect. Parameters are about the same.

Students Darn! We didn't find anything interesting.

Bill The fact that the numbers didn't affect anything **is** interesting. We didn't know, and now we do!

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

It turns out that the frequency of numbers and symbols was around that of q, z so had no real affect. Parameters are about the same.

Students Darn! We didn't find anything interesting.

Bill The fact that the numbers didn't affect anything **is** interesting. We didn't know, and now we do!

Students You are so wise!

What Happens with Math Texts?

I had some High School students find frequencies of letters and numbers in math texts to see if we still get the big gap.

It turns out that the frequency of numbers and symbols was around that of q, z so had no real affect. Parameters are about the same.

Students Darn! We didn't find anything interesting.

Bill The fact that the numbers didn't affect anything **is** interesting. We didn't know, and now we do!

Students You are so wise! (They didn't really say this.)

Is Shift Cipher Secure if we are Transmitting Just Numbers?

What if Alice sends Bob a credit card number? **Discuss**

Is Shift Cipher Secure if we are Transmitting Just Numbers?

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

Is Shift Cipher Secure if we are Transmitting Just Numbers?

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

1. Visa cards always begin with 4.

Is Shift Cipher Secure if we are Transmitting Just Numbers?

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.

Is Shift Cipher Secure if we are Transmitting Just Numbers?

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.

Is Shift Cipher Secure if we are Transmitting Just Numbers?

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.
4. Parity Checks.

Lessons Learned from Shift Being Crackable

Lessons Learned from Shift Being Crackable

1. Shift is crackable because key space is **small**.

Lessons Learned from Shift Being Crackable

1. Shift is crackable because key space is **small**.
2. **small** and **large** may be technology-dependent.

Lessons Learned from Shift Being Crackable

1. Shift is crackable because key space is **small**.
2. **small** and **large** may be technology-dependent.
3. Needed to use **IS-ENGLISH** program which we will use later as well.

Kerckhoff's Principle

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?
More generally we will always use the following assumption.
Kerckhoff's principle:

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?
More generally we will always use the following assumption.
Kerckhoff's principle:

- Eve knows **The encryption scheme**.

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?
More generally we will always use the following assumption.
Kerckhoff's principle:

- Eve knows **The encryption scheme**.
- Eve knows **the alphabet and the language**.

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?
More generally we will always use the following assumption.
Kerckhoff's principle:

- Eve knows **The encryption scheme**.
- Eve knows **the alphabet and the language**.
- Eve does not know **the key**

Kerckhoff's principle

We looked at the question:
If Eve KNOWS that SHIFT is used then is the cipher crackable?
More generally we will always use the following assumption.
Kerckhoff's principle:

- Eve knows **The encryption scheme**.
- Eve knows **the alphabet and the language**.
- Eve does not know **the key**
- The key is chosen **at random**.

Arguments For And Against Kerckhoff's Principle

Arguments For:

Arguments For And Against Kerckhoff's Principle

Arguments For:

- Easier to keep **key** secret than **cipher**.

Arguments For And Against Kerckhoff's Principle

Arguments For:

- Easier to keep **key** secret than **cipher**.
- Easier to change **key** than to change **cipher**.

Arguments For And Against Kerckhoff's Principle

Arguments For:

- Easier to keep **key** secret than **cipher**.
- Easier to change **key** than to change **cipher**.
- Standardization:

Arguments For And Against Kerckhoff's Principle

Arguments For:

- Easier to keep **key** secret than **cipher**.
- Easier to change **key** than to change **cipher**.
- Standardization:
  - Ease of deployment.
  - Public validation.
- If prove system secure then very strong proof of security since even if Eve knows cipher she can't crack.

Arguments For And Against Kerckhoff's Principle

Arguments For:

- Easier to keep **key** secret than **cipher**.
- Easier to change **key** than to change **cipher**.
- Standardization:
  - Ease of deployment.
  - Public validation.
- If prove system secure then very strong proof of security since even if Eve knows cipher she can't crack.

Arguments Against:

- When initially use a cipher then Eve won't know what the cipher is for a while (months? days? hours?) For that (perhaps short) period of time the secrecy of the cipher will make it hard to crack.

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

Can Two 1-Letter Messages Leak Information?

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

Scenario

Visible to all: **Is Saj a double agent working for the Klingons?**

Can Two 1-Letter Messages Leak Information?

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

Scenario

Visible to all: **Is Saj a double agent working for the Klingons?**
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

Visible to all: **Is Saj a double agent working for the Klingons?**
The answer comes via a shift cipher: **A** (which is either Y or N)

In clear: **Is Saj a double agent working for the Romulans?**

Can Two 1-Letter Messages Leak Information?

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

Scenario

Visible to all: **Is Saj a double agent working for the Klingons?**
The answer comes via a shift cipher: **A** (which is either Y or N)

In clear: **Is Saj 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

Visible to all: **Is Saj a double agent working for the Klingons?**
The answer comes via a shift cipher: **A** (which is either Y or N)

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

Since the answer to both questions was **the same**, namely **A**,
Eve knows Saj is working for either **both** or **neither.**

Eve Can Tell if Two Message Are Same

Issue If Eve sees 2 messages, she knows if same or diff.

Does this leak information Discuss.

Eve Can Tell if Two Message Are Same

Issue If Eve sees 2 messages, she knows if same or diff.

Does this leak information Discuss. Yes.

Eve Can Tell if Two Message Are Same

Issue If Eve sees 2 messages, she knows if same or diff.

Does this leak information Discuss. Yes.

What to do about this? Discuss.

Eve Can Tell if Two Message Are Same

Issue If Eve sees 2 messages, she knows if same or diff.

Does this leak information Discuss. Yes.

What to do about this? Discuss.

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

Eve Can Tell if Two Message Are Same

Issue If Eve sees 2 messages, she knows if same or diff.

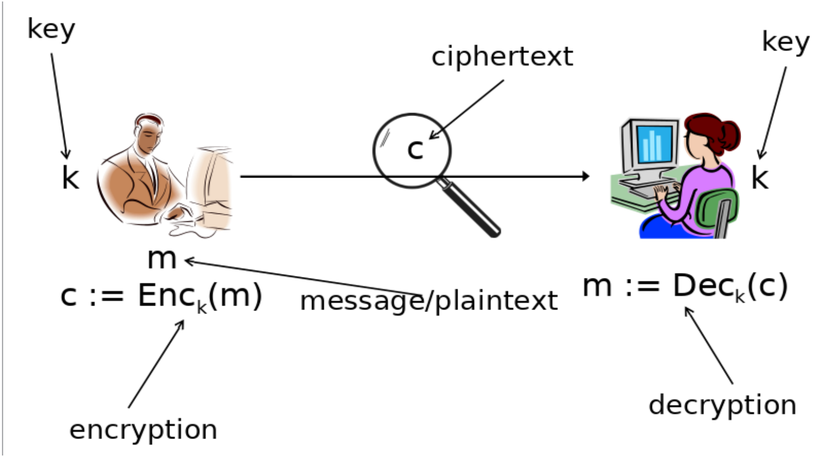
Does this leak information Discuss. Yes.

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

ciphertext

key

k

c

k

m

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

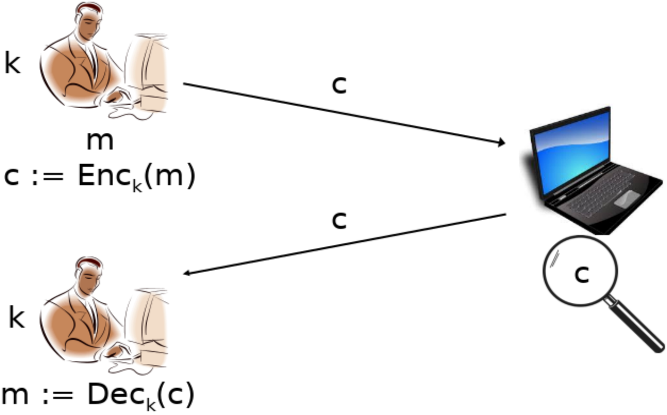
message/plaintext

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

decryption

encryption

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 $k \in \{0, \dots, 25\}$. 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 $\text{Enc}(m_1, \dots, m_n) = (m_1 + k, \dots, m_n + k)$.)
  - **Dec** (decryption algorithm): takes key k and ciphertext c as input; outputs m or "error"

    $$m := Dec_k(c)$$

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

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

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

BILL, STOP RECORDING LECTURE!!!!

BILL STOP RECORD LECTURE!!!