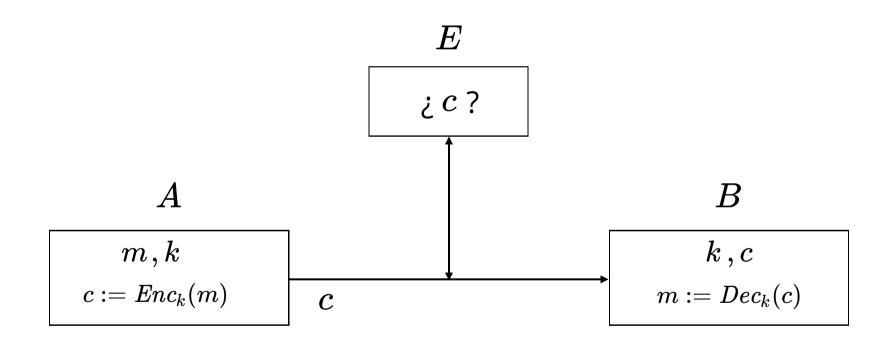
IIC3253

OTP y perfect secrecy

Cifrado (simétrico)



Cifrado del César

A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z X Y Z A B C D E F G H I J K L M N Ñ O P Q R S T U V W

HOLA MUNDO EMIX JRKAM

MANDEN BITCOINS A UCRANIA JXKABK YFQZMFKP X RZOXKFX

¿Problemas?

Cifrado del César + llave

Llave = shift 7

A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z T U V W X Y Z A B C D E F G H I J K L M N Ñ O P O R S

HOLA MUNDO AIET FÑGWI

MANDEN BITCOINS A UCRANIA FTGWXG UBNVIBGM T ÑVLTGBT

¿Problemas?

La probabilidad de que un atacante "seleccione" o "adivine" la llave correcta debe ser muy baja.

⇒ El espacio de llaves posibles debe ser muy (muy) grande

¿Cómo podríamos agrandar el espacio de llaves siguiendo la idea de "sustituir"?

Shift → Permutación

A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z P O O W I E U R Y T L K A J S H D F G Ñ M Z N X B C V

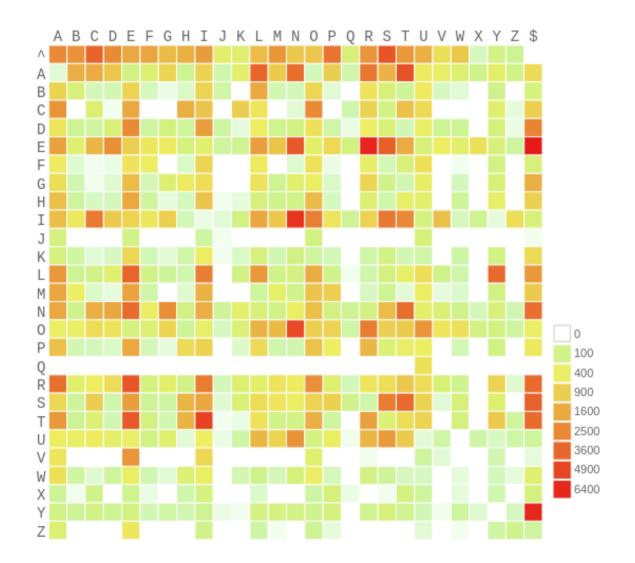
HOLA MUNDO RHKP AZJWH

¿Cuántas llaves posibles?

27! = 10,888,869,450,418,352,160,768,000,000

¿Es este un buen cifrado?





Un espacio de llaves grande es necesario, no suficiente.

ONE-TIME PAD (OTP)

Operación Módulo

(Recordatorio)

Dados $a,n\in\mathbb{Z}$, existe un único par de elementos $(q,r)\in\mathbb{Z}^2$ tal que:

$$0 \leq r < |n|$$
 $a = q \cdot n + r$ Cuociente Resto

Decimos entonces que $a \bmod n = r$ y que $a \equiv r \bmod n$

Operación Módulo

(Ejemplos)

$$10 \mod 3 = 1$$
 $28 \mod 8 = 4$
 $6 \mod -20 = 6$
 $-6 \mod -20 = 14$

Siempre esperaríamos que

$$n \cdot \lfloor \frac{a}{n} \rfloor + (a \bmod n) = a$$

Programando, esto se ve como

$$n * (a / n) + a % n = a$$
División entera

```
1 # Python
2 print("La división entera entre 6 y -20 es:")
3 print(6 // -20)
```

Output: -1

```
1 // C++
2 #include <iostream>
3 using namespace std;
4
5 int main() {
    cout << "La división entera entre 6 y -20 es: " << endl;
7    cout << (6 / -20) << endl;
8    return 0;
9 }</pre>
```

Output: 0



Esperamos que

$$n * (a / n) + a % n = a$$

$$-20 * (6 / -20) + 6 % -20 = 6$$

Python:
$$-20 * -1 + 6 \% -20 = 6 \implies 6 \% -20 = -14$$

C++:
$$-20 * 0 + 6 % -20 = 6$$
 $\Rightarrow 6 % -20 = 6$



Operación Módulo

Dados $a,n\in\mathbb{Z}$, existe un único par de elementos $(q,r)\in\mathbb{Z}^2$ tal que:

$$0 \le r < |n|$$

$$a = q \cdot n + r$$

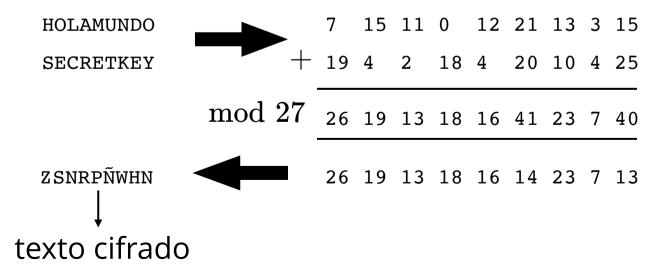
Decimos entonces que $a \bmod n = r$ y que $a \equiv r \bmod n$

ONE-TIME PAD (OTP)

Partimos enumerando las letras

A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

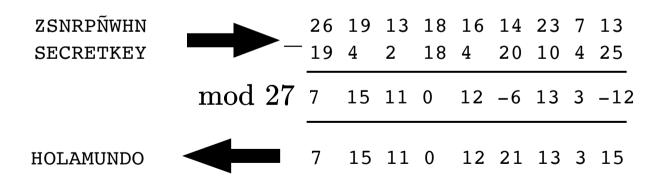
Para enviar un mensaje de largo ℓ necesitaremos una llave de largo ℓ



$Enc_{\mathrm{SECRETKEY}}(\mathrm{HOLAMUNDO}) = \mathrm{ZSNRP ilde{N}WHN}$

¿Cómo decriptar?

 $Dec_{ ext{SECRETKEY}}(ext{ZSNRPNWHN}) = ext{HOLAMUNDO}$



A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26



La clase pasada...

El operador módulo en lenguajes de programación...

$$n * (a (n) + a % n = a)$$
División entera

Si
$$7/-15 = -1$$
 entonces $7\%-15 = -8$

$$Si 7/-15 = 0 entonces 7\%-15 = 7$$

Si nos mantenemos **siempre** en módulo -15 entonces todo funciona bien.

¿Si salimos de ese mundo?

Al menos $1 \equiv 16 \mod -15$

¿Si aplicamos otros módulos?

$$10 \% (7 \% -15) = \frac{-6}{3}$$

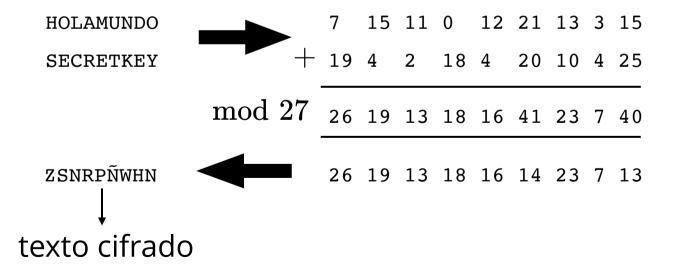
Aquí tenemos:

$$3 \not\equiv -6 \mod -15$$
 $3 \not\equiv -6 \mod 7$ $3 \not\equiv -6 \mod -8$

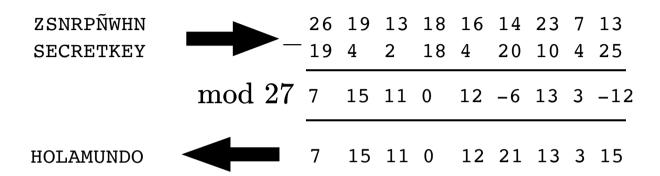
ONE-TIME PAD (OTP)

ABCDEFGHIJK L M N Ñ O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

Para enviar un mensaje de largo ℓ necesitaremos una llave de largo ℓ



¿Cómo decriptar?



Para formalizarlo necesitamos convertir mensajes, llaves y textos cifrados en arreglos de enteros

$$m= ext{HOLAMUNDO}$$
 $ar{m}=(7,15,11,0,12,21,13,3,15)$ $ar{k}= ext{SECRETKEY}$ $ar{k}=(19,4,2,18,4,20,10,4,25)$ $ar{c}= ext{ZSNRP\~NWHN}$ $ar{c}=(26,19,13,18,16,14,23,7,13)$

De la misma forma necesitamos hacer la conversión en la otra dirección

$$a = (4, 9, 4, 12, 16, 11, 15)$$
 $\bar{a} = \texttt{EJEMPLO}$

Naturalmente, siempre se cumple que $\overline{\overline{s}}=s$

Con esto definimos OTP en base a

$$Enc_k(m) = \overline{(ar{m} + ar{k}) \mod 27}$$

$$Dec_k(c) = \overline{(ar{c} - ar{k}) \mod 27}$$

Desde ahora supondremos que nuestros mensajes y llaves **son** arreglos de números

Definiremos OTP simplemente usando

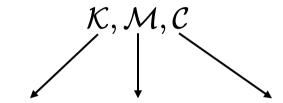
$$Enc_k(m) = (m+k) \mod 27$$

$$Dec_k(c) = (c - k) \mod 27$$

$$Dec_k(Enc_k(m)) = ((m+k) \mod 27 - k) \mod 27$$

$$m = (m + k - k) \mod 27 = m \mod 27 = m$$

Esquema criptográfico



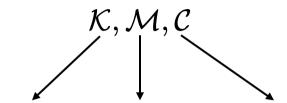
Espacio de llaves, mensajes y textos cifrados

Un esquema es un triple (Gen, Enc, Dec)

Gen es una distribución de probabilidades sobre ${\cal K}$

Es decir, $Gen: \mathcal{K}
ightarrow [0,1]$ tal que $\sum_{k \in \mathcal{K}} \mathit{Gen}(k) = 1$

Esquema criptográfico



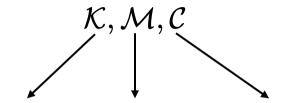
Espacio de llaves, mensajes y textos cifrados

Un esquema es un triple (Gen, Enc, Dec)

 $Enc = \{Enc_k \mid k \in \mathcal{K}\}$ es una familia de algoritmos para encriptar

Para cada $k \in \mathcal{K}$, se tiene que $Enc_k : \mathcal{M} \to \mathcal{C}$

Esquema criptográfico



Espacio de llaves, mensajes y textos cifrados

Un esquema es un triple (Gen, Enc, Dec)

 $Dec = \{Dec_k \mid k \in \mathcal{K}\}$ es una familia de algoritmos para decriptar

Para cada $k \in \mathcal{K}$, se tiene que $Dec_k : \mathcal{C} o \mathcal{M}$

Esperamos que para un esquema criptográfico (Gen, Enc, Dec) se cumpla

$$orall k \in \mathcal{K} \, orall m \in \mathcal{M} : Dec_k(Enc_k(m)) = m$$

En este caso diremos que el esquema es perfectamente correcto

¿Por qué *perfectamente*?

OTP: sobre $\{0,1,\ldots,N-1\}$ y mensajes de largo ℓ (OTP $^{N,\ell}$)

$$\mathcal{K} = \mathcal{M} = \mathcal{C} = \{0,\dots,N-1\}^\ell$$

Gen es la distribución uniforme sobre $\{0,\dots,N-1\}^\ell$

$$Enc_k(m) = (m+k) \mod N$$

$$Dec_k(c) = (c - k) \mod N$$

¿Qué tan bueno es OTP?

¿Qué pasa si veo un mensaje cifrado \emph{c} pasar?

Aquí un ejemplo:

c = YFTGXEIWIWEHAGQGESLPKRVLMYGXSJIQZVIYHVBRJGNTR

m = ESTEMENSAJEESLITERALMENTEIMPOSIBLEDEDECRIPTAR

k = UNACLAVEINADIVINABLEYNISIQUIERAPORFUERZABRUTA

Perfect Secrecy

¿Cuándo decimos que un esquema criptográfico es *perfectamente secreto?*

Pensemos en la idea de que si un atacante ve un texto cifrado *no gana información*.

Podríamos decir algo como lo siguiente:

"Al ver un texto cifrado $c_{\it 0}$ pasar, para el atacante el mensaje original m podría haber sido cualquiera"

¿Cómo formalizamos esto?

Dado un texto cifrado c_0 se cumple que

$$orall m_0 \in \mathcal{M}: egin{array}{c} \Pr_{k \sim Gen} \left[Enc_k(m_0) = c_0
ight] = rac{1}{|\mathcal{M}|} \end{array}$$

$$orall m_0 \in \mathcal{M}: egin{array}{c} \Pr_{k \sim Gen} \left[Enc_k(m_0) = c_0
ight] = rac{1}{|\mathcal{M}|} \end{array}$$

¿Cómo se calcula esta probabilidad?

Es simplemente la probabilidad de haber elegido una llave que encripte m_0 como c_0

$$\sum_{k \in \mathcal{K} \,:\, Enc_k(m_0) = c_0} Gen(k)$$

¿Qué pasa si el atacante tenía información previa sobre el mensaje?

Por ejemplo sabe que el mensaje puede ser "atacar ahora" o "emprender retirada"

Podría incluso estimar que atacarán con probabilidad 1/3

¿Cómo modelamos esto matemáticamente?

Supondremos que el atacante tiene una distribución de probabilidad \mathbb{D} sobre \mathcal{M}

Para cada distribución de probabilidad $\mathbb D$ sobre $\mathcal M$ y cada texto cifrado $c_0 \in \mathcal C$ se cumple que

$$egin{array}{lll} orall m_0 \in \mathcal{M}: & \Pr_{m \sim \mathbb{D}} & [m = m_0 \mid Enc_k(m) = c_0] & = & \Pr_{m \sim \mathbb{D}} [m = m_0] \ k \sim Gen \end{array}$$

Recordemos que
$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$rac{\mathbb{D}(m_0) \sum_{k \in \mathcal{K} \colon Enc_k(m_0) = c_0} Gen(k)}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \sum_{k \in \mathcal{K} \colon Enc_k(m) = c_0} Gen(k)} \quad \stackrel{?}{=} \quad \mathbb{D}(m_0)$$

¿Es $OTP^{N,\ell}$ perfectamente secreto?

- 1. Gen es la distribución uniforme $1/N^\ell$
- 2. Para cada c_0 y cada m_0 existe una única llave k tal que $Enc_k(m_0)=c_0$

Sea $c_0 \in \mathcal{C}$ un texto cifrado y m_0 un mensaje.

$$rac{\mathbb{D}(m_0) \sum_{k \in \mathcal{K} \,:\, Enc_k(m_0) = c_0} \mathit{Gen}(k)}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \sum_{k \in \mathcal{K} \,:\, Enc_k(m) = c_0} \mathit{Gen}(k)}$$

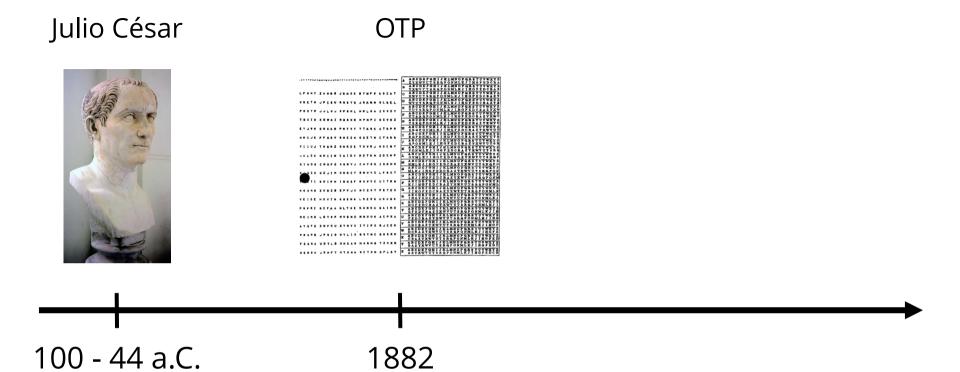
$$1 = \frac{\mathbb{D}(m_0) \cdot 1/N^\ell}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \cdot 1/N^\ell} = \mathbb{D}(m_0)$$

¿Definiciones alternativas?

- 1. La probabilidad de ver cualquier texto cifrado sin conocimiento previo es la misma que la probabilidad de ver dicho texto cifrado conociendo el mensaje de antemano.
- 2. La distribución de probabilidad sobre los mensajes es independiente de la distribución de probabilidad sobre los textos cifrados.

Ejercicio: formalizar estas nociones

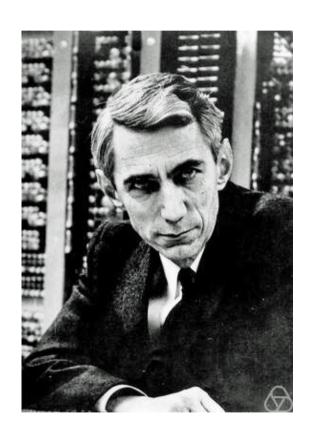
Un poco de historia

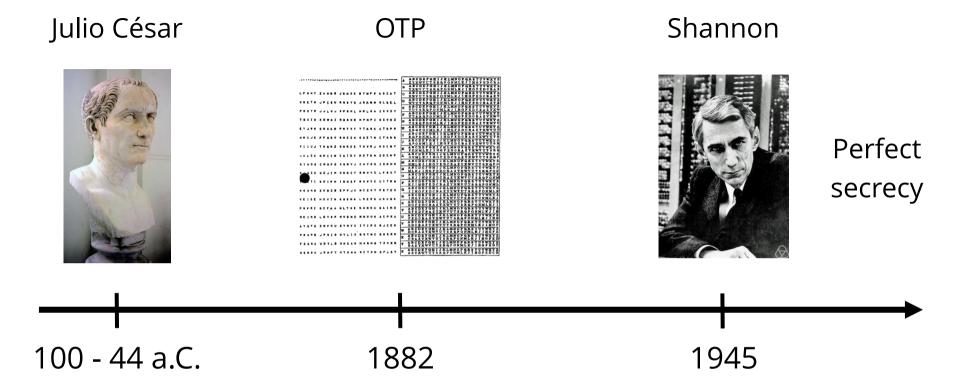


Shannon (1945)

A Mathematical Theory of Cryptography

Primer desarrollo de fundamentos matemáticos de la criptografía





Pero perfect secrecy es una condición muy fuerte

Lamentablemente...

Hemos discutido en clases que pareciera molesto y/o poco razonable que la llave tenga que ser tan larga como el mensaje.

¿Cómo modificamos OTP para tener $|\mathcal{K}| < |\mathcal{M}|$ y seguir teniendo un esquema criptográfico *perfectamente secreto?*

No podemos 🤵

Teorema

Sean $\mathcal{M}, \mathcal{K}, \mathcal{C}$ espacios de mensajes, llaves y textos cifrados, respectivamente.

Si $|\mathcal{K}| < |\mathcal{M}|$, entonces no existe un esquema (Gen, Enc, Dec) que sea perfectamente secreto.

Demostración

Supongamos que $|\mathcal{K}| < |\mathcal{M}| \le |\mathcal{C}|$ y sea $(\mathit{Gen}, \mathit{Enc}, \mathit{Dec})$ un esquema criptográfico

Sea $\mathbb D$ una distribución sobre $\mathcal M$ y $m_0\in \mathcal M$ un mensaje tal que $\mathbb D(m_0)>0$

Como $|\mathcal{K}|<|\mathcal{M}|\leq |\mathcal{C}|$, debe existir $c_0\in\mathcal{C}$ para el cual **ninguna** llave $k\in\mathcal{K}$ satisface $Enc_k(m_0)=c_0$

$$egin{aligned} & \Pr[m = m_0 \mid Enc_k(m) = c_0] \ & k \sim Gen \end{aligned} egin{aligned} & < & \mathbb{D}(m_0) = \Pr_{m \sim \mathbb{D}}[m = m_0] \ & \text{Adiós perfect secrecy...} \end{aligned}$$



Back to reality

OTP y la noción de Perfect Secrecy son fundamentales para entender lo que viene

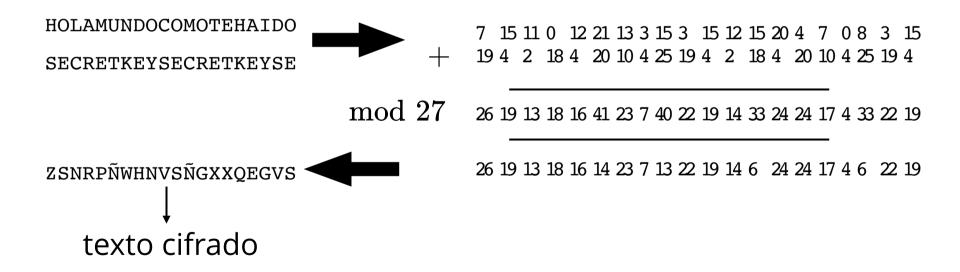
Pero en la práctica vamos a buscar otras propiedades...

¿Bajo qué modelo de ataque es OTP seguro?

- 1. Texto cifrado (sólo veo c_0, c_1, \ldots)
- 2. Texto (yeo $(m_0, c_0), (m_1, c_1), \ldots$)
- 3. Ter o plante gido (mando a encriptar m's)
- 4. Te to circado engido (mando a encriptar m's y a criptar s)

¿Qué pasa si usamos llaves más cortas que el mensaje?

Pensemos por ejemplo en repetir la llave varias veces para encriptar un mensaje más largo que la llave



¿Podemos quebrarlo?