

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

APUNTE IIC3253

Criptografía y Seguridad Computacional

Autor
Cristóbal ROJAS

En base a apuntes de
Prof. Marcelo ARENAS
Prof. Martín UGARTE

En base al texto
Introduction to Modern Cryptography
J. KATZ, Y. LINDELL

10 de abril de 2023



Índice

1. Introducción	2
1.1. Dos problemas fundamentales: cifrado y autenticación	2
1.2. Sintaxis de encriptación	3
1.3. Principio de Kerckhoffs	3
1.4. Principios de la criptografía moderna	4
1.5. Seguridad, noción de adversario y tipos de ataque	4
2. Criptografía simétrica	6
2.1. Cifrado del César	6
2.2. Una primera versión de One-Time Pad (OTP)	6
2.3. Perfect secrecy	8
2.4. Ahora sí, One-Time Pad	9
2.5. Permutaciones pseudo-aleatorias (PRP)	10

1. Introducción

1.1. Dos problemas fundamentales: cifrado y autenticación

La criptografía moderna implica *el estudio de técnicas matemáticas para proteger la información digital, los sistemas y los cálculos distribuidos contra ataques de adversarios*.

La seguridad de los **esquemas criptográficos modernos** (hablaremos de ellos más adelante) recae en una **llave secreta** compartida con anterioridad entre el emisor (A, de *Alice*) y receptor del mensaje (B, de *Bob*), que debe ser desconocida para algún adversario (E, de *Eavesdropper*). Este escenario, en donde las partes que se comunican comparten alguna información secreta por adelantado, es conocida como **encriptación de llave privada** (*private-key encryption*).

En este contexto, Alice y Bob comparten una llave privada cuando se quieren comunicar en secreto. Uno de ellos, por ejemplo, Alice, puede mandar un **mensaje** (*plain text*, texto plano) hacia el otro usando la llave para **encriptar** el mensaje y así obtener un **texto cifrado** o *ciphertext* que es transmitido al receptor. Bob, el receptor, usa la misma llave para **decriptar** el texto cifrado y recuperar el mensaje original.

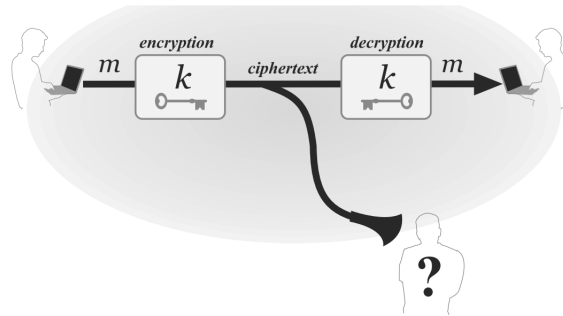


Figura 1: Alice manda un mensaje m a Bob usando una llave privada k que compartieron anteriormente

La **autenticación**, en cambio, es el proceso en el que nos interesa garantizar la **integridad del mensaje** contra un adversario que podría modificar o mandar otros mensajes por el canal de comunicación.

Por ejemplo, Alice y Bob, que desean comunicarse de forma autenticada, empiezan por generar y compartir una llave secreta k antes de su comunicación. Cuando una de las partes desea enviar un mensaje m a la otra, calcula una etiqueta (*tag*) t basada en el mensaje y la llave compartida, y envía el mensaje m junto con t a la otra parte. El *tag* se calcula utilizando un *algoritmo de generación de tags* denotado por **Mac**; así, reformulando lo que acabamos de decir, el emisor de un mensaje m calcula $t \leftarrow \text{Mac}_k(m)$ y transmite (m, t) al receptor. Al recibir (m, t) , la segunda parte verifica si t es un *tag* válido en el mensaje m (con respecto a la llave compartida) o no. Esto se hace ejecutando un *algoritmo de verificación* **Verify** que toma como entrada la llave compartida, así como un mensaje m y un *tag* t , e indica si el *tag* dado es válido. Profundizaremos más sobre este tema en el capítulo X.

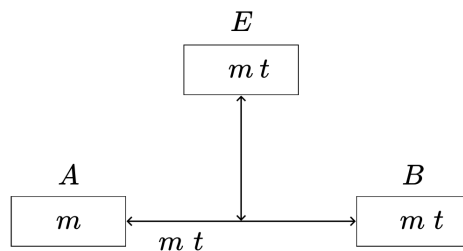


Figura 2: Alice manda (m, t) a Bob, generando t con **Mac** y Bob debe verificar t con **Verify**

1.2. Sintaxis de encriptación

Formalmente, un esquema de encriptación de llave privada es definido por un **espacio de mensajes** \mathcal{M} , un **espacio de llaves** \mathcal{K} y un **espacio de textos cifrados** \mathcal{C} , junto con tres funciones: una para generar llaves (**Gen**), una para encriptar (**Enc**) y otra para decriptar (**Dec**). El espacio de mensajes \mathcal{M} define el conjunto “legal” de mensajes, es decir, los soportados por el esquema. Las funciones del esquema tienen la siguiente funcionalidad:

- ♦ **Gen** es un algoritmo de probabilidades que retorna una llave k elegida acorde a una distribución. Es decir, $\text{Gen} : \mathcal{K} \rightarrow [0, 1]$ tal que

$$\sum_{k \in \mathcal{K}} \text{Gen}(k) = 1$$

- ♦ **Enc** es una familia de algoritmos de encriptación que toma como input la llave k y el mensaje m , y retorna un texto cifrado c . Denotamos $\text{Enc}_k(m)$ la encriptación del texto plano m usando la llave k .

$$\text{Enc} = \{\text{Enc}_k \mid k \in \mathcal{K}\}, \quad \text{con } \text{Enc}_k : \mathcal{M} \rightarrow \mathcal{C}, \forall k \in \mathcal{K}$$

- ♦ **Dec** es una familia de algoritmos de decriptación que toma como input la llave k y el texto cifrado c , y retorna un texto plano m . Denotamos $\text{Dec}_k(c)$ la decriptación del texto cifrado c usando la llave k .

$$\text{Dec} = \{\text{Dec}_k \mid k \in \mathcal{K}\}, \quad \text{con } \text{Dec}_k : \mathcal{C} \rightarrow \mathcal{M}, \forall k \in \mathcal{K}$$

Un esquema de encriptación debe satisfacer el siguiente requisito de correctitud: para cada llave k generada por **Gen** y para cada mensaje $m \in \mathcal{M}$, se tiene que

$$\text{Dec}_k(\text{Enc}_k(m)) = m$$

En otras palabras, encriptar un mensaje y luego decriptar el texto cifrado resultante usando la misma llave devuelve el mensaje original.

1.3. Principio de Kerckhoffs

Si un adversario conoce el algoritmo **Dec** y la llave k compartida por las dos partes comunicantes, podrá descifrar cualquier texto cifrado transmitido por dichas partes. Por este motivo, las partes comunicantes deben compartir la llave k de forma segura y mantener k completamente oculta para los demás. ¿Quizás también deberían mantener en secreto el algoritmo de descifrado **Dec**? ¿No sería mejor que mantuvieran en secreto todos los detalles del esquema de cifrado?

Lo anterior condujo a Auguste Kerckhoffs en 1838 a declarar el siguiente principio:

*“La seguridad de un sistema criptográfico **no** debe depender de que los algoritmos de cifrado y descifrado sean secretos, solo debe depender de que las claves sean secretas”.*

Es decir, un esquema de encriptación debe diseñarse para que sea seguro *incluso* si un espía conoce todos los detalles del esquema, siempre que el atacante no conozca la clave utilizada. Dicho de otro modo, la seguridad no debe depender de que el esquema de cifrado sea secreto; en cambio, el principio de Kerckhoffs exige que la seguridad dependa únicamente del secreto de la clave.

Seguimos este principio por tres simples motivos:

- ♦ Es más fácil mantener la privacidad de una llave que la de un algoritmo.
- ♦ Si la seguridad se ve comprometida es más fácil cambiar una llave que un algoritmo.
- ♦ Es mejor usar algoritmos públicos que hayan sido ampliamente verificados.

1.4. Principios de la criptografía moderna

Podemos mencionar tres grandes principios para la criptografía moderna:

1. **Definiciones formales:** Es importante definir formalmente los sistemas criptográficos y nociones de seguridad usados.

“Si no sabes lo que quieres conseguir, ¿cómo vas a saber si lo has conseguido?”

Las definiciones formales facilitan esa comprensión al ofrecer una descripción clara de las amenazas que entran en juego y de las garantías de seguridad deseadas. Como tales, las definiciones pueden ayudar a guiar el diseño de esquemas criptográficos. De hecho, es mucho mejor formalizar lo que se necesita *antes* de que comience el proceso de diseño, en lugar de llegar a una definición *post facto* —posterior a los hechos— una vez que el diseño se ha completado.

2. **Supuestos precisos:** Es importante que los supuestos detrás del funcionamiento de un sistema criptográfico tengan una **formulación precisa** y sean **conocidos**.

La mayoría de los esquemas criptográficos modernos no pueden demostrarse seguros de forma incondicional; para ello habría que resolver cuestiones de la teoría de la complejidad computacional que hoy en día parecen lejos de tener respuesta. El resultado de esta desafortunada situación es que las pruebas de seguridad suelen basarse en suposiciones. La criptografía moderna exige que estas suposiciones sean explícitas y precisas desde el punto de vista matemático. En el nivel más básico, esto se debe a que las pruebas de seguridad así lo exigen.

3. **Pruebas de seguridad:** Es importante construir **demostraciones formales de seguridad** (basadas en las definiciones y supuestos).

Los dos principios anteriores que acabamos de describir nos permiten alcanzar nuestro objetivo de proporcionar pruebas rigurosas de que una construcción satisface una definición dada bajo ciertos supuestos. Estas pruebas son especialmente importantes en el contexto de la criptografía, donde hay un atacante que intenta activamente “romper” algún esquema. Las pruebas de seguridad ofrecen una garantía irrefutable —relativa a la definición y los supuestos— de que ningún atacante tendrá éxito; esto es mucho mejor que adoptar un enfoque sin principios o heurístico del problema. Sin una prueba de que ningún adversario con los recursos especificados puede romper un esquema, sólo nos queda nuestra intuición de que es así. La experiencia ha demostrado que la intuición en criptografía y seguridad informática es desastrosa. Hay innumerables ejemplos de esquemas no probados que se rompieron, a veces inmediatamente y otras años después de ser desarrollados.

1.5. Seguridad, noción de adversario y tipos de ataque

En general, una definición de seguridad tiene dos componentes: una **garantía de seguridad** (o, desde el punto de vista del adversario, lo que constituye un ataque exitoso) y un **modelo de amenaza** (tipos de ataque). La garantía de seguridad define lo que se pretende que el esquema impida al atacante hacer, mientras que el modelo de amenaza describe el poder del adversario, es decir, qué acciones se supone que el atacante puede llevar a cabo.

Los siguientes puntos engloban lo que un esquema de encriptación debería **garantizar**:

- ♦ *Debe ser imposible para un adversario recuperar la llave:* aunque la seguridad criptográfica depende de que el adversario no pueda encontrar la llave, este requisito por sí solo no es suficiente para garantizar la seguridad del esquema.
- ♦ *Debe ser imposible para un adversario recuperar el mensaje (texto plano) a partir del texto cifrado:* este requisito es fundamental para la seguridad del cifrado. Si un adversario puede recuperar el mensaje original a partir del texto cifrado, entonces el esquema de cifrado no es seguro.

- ♦ *Debe ser imposible para un adversario recuperar cualquier carácter del mensaje a partir del texto cifrado:* este requisito es aún más riguroso, ya que no sólo busca proteger el mensaje completo, sino también la información contenida en él.
- ♦ *La “respuesta correcta”:* independientemente de la información que tenga el adversario, el texto cifrado no debe filtrar información adicional sobre el mensaje. Esta definición amplia garantiza que el esquema de cifrado protege la información sin importar cuánta información tenga el adversario. El objetivo del cifrado es evitar que el adversario obtenga cualquier información adicional sobre el mensaje, lo que se logra si el texto cifrado no contiene información adicional más allá de lo que ya se sabe.

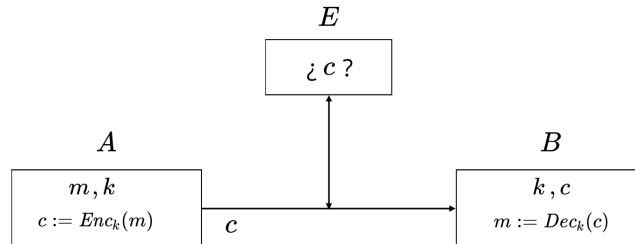
El **modelo de amenaza** (los tipos de ataque) especifica qué “poder” se supone que tiene el atacante, pero no impone ninguna restricción en la *estrategia* del adversario. Esta es una distinción importante: especificamos lo que asumimos sobre las habilidades del adversario, pero **no** asumimos nada acerca de cómo usa esas habilidades. Es imposible prever qué estrategias podrían ser utilizadas en un ataque, y la historia ha demostrado que los intentos de hacerlo están destinados al fracaso. Algunos de los tipos de ataque son:

- ♦ **Ciphertext-only attack:** es el ataque más básico, en el que el adversario solo observa el texto cifrado y trata de determinar información sobre el texto plano subyacente. Este es el modelo de amenaza que hemos estado asumiendo implícitamente al discutir el esquema de encriptación definido anteriormente.
- ♦ **Known-plaintext attack:** aquí, el adversario puede aprender uno o más pares de texto plano/cifrado generados utilizando alguna clave. El objetivo del adversario es deducir información sobre el texto plano subyacente de otro texto cifrado producido utilizando la misma clave.
- ♦ **Chosen-plaintext attack:** en este ataque, el adversario puede obtener pares de texto plano/cifrado para textos planos de su elección.
- ♦ **Chosen-ciphertext attack:** el último tipo de ataque es aquel en el que el adversario es capaz de obtener (alguna información sobre) el descifrado de textos cifrados de su elección. El objetivo del adversario, una vez más, es aprender información sobre el texto plano subyacente de algún otro texto cifrado (cuyo descifrado el adversario no puede obtener directamente) generado utilizando la misma clave.

Aunque los modelos de amenaza están listados en orden de creciente fuerza, ninguno es inherentemente mejor que otro; el adecuado a utilizar depende del entorno en el que se despliega un esquema de encriptación.

2. Criptografía simétrica o de clave privada

Hasta ahora, un ejemplo de criptografía simétrica se ve así:



Alice desea mandar un mensaje m a Bob, por lo que usa Enc_k para encriptar el mensaje con la llave k que compartieron anteriormente, enviando así el texto cifrado c por un canal de comunicación. Bob, por su lado, puede descifrar c usando Dec_k con la llave k y así obtener el mensaje m . Frente a lo anterior, puede existir un adversario E que quiera conocer la información del mensaje m capturando el texto cifrado c .

2.1. Cifrado del César

Julio César cifró desplazando (*shift*) las letras del alfabeto 3 lugares hacia delante: A se sustituyó por D, B por E, y así sucesivamente. Al final del alfabeto, las letras se envuelven alrededor y así Z fue reemplazado con C, y con B, y X con A. Por ejemplo, el cifrado del mensaje BEGIN THE ATTACK NOW, con espacios eliminados, da:

EHJLQWKHDWDFNQZR

Un problema inmediato de este cifrado es que el método de encriptación es fijo; **no hay llave**. Así, **cualquiera que supiera cómo encriptó César sus mensajes podría descifrarlos sin esfuerzo**.

A pesar de que es poco seguro, este cifrado nos servirá como base para entender OTP, un esquema de encriptación *perfectamente secreto* (ya veremos que significa esto).

2.2. Una primera versión de One-Time Pad (OTP)

Veamos una construcción de OTP basada en el cifrado del César. Partimos enumerando las 27 letras del abecedario:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	N	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 . Por ejemplo, probemos mandar el mensaje HOLAMUNDO con la llave SECRETKEY:

HOLAMUNDO	→	7	15	11	0	12	21	13	3	15	
SECRETKEY	→	+	19	4	2	18	4	20	10	4	25
			<hr/>								
		mod 27	26	19	13	18	16	41	23	7	40
			<hr/>								
ZSNRPÑWHN	←	26	19	13	18	16	14	23	7	13	
↓											
texto cifrado											

En el ejemplo anterior, a cada letra del mensaje y de la llave, se convierten a un número según la tabla del abecedario, luego se suman y les aplica la operación módulo 27 (ya que son 27 letras). Así, se obtiene un número que representa el cifrado de la letra correspondiente. Es decir:

$$\text{Enc}_{\text{SECRETKEY}}(\text{HOLAMUNDO}) = \text{ZSNRPÑWHN}$$

Para descryptar, hacemos el mismo proceso, solo que en vez de sumar, restamos.

$$\begin{array}{rcl}
 \text{ZSNRPÑWHN} & \xrightarrow{\quad} & 26 \ 19 \ 13 \ 18 \ 16 \ 14 \ 23 \ 7 \ 13 \\
 \text{SECRETKEY} & \xrightarrow{\quad} & 19 \ 4 \ 2 \ 18 \ 4 \ 20 \ 10 \ 4 \ 25 \\
 & & \hline
 & \text{mod } 27 & 7 \ 15 \ 11 \ 0 \ 12 \ -6 \ 13 \ 3 \ -12 \\
 & & \hline
 \text{HOLAMUNDO} & \xleftarrow{\quad} & 7 \ 15 \ 11 \ 0 \ 12 \ 21 \ 13 \ 3 \ 15
 \end{array}$$

Y tenemos entonces que

$$\text{Dec}_{\text{SECRETKEY}}(\text{ZSNRPÑWHN}) = \text{HOLAMUNDO}$$

Para **formalizar** este algoritmo, necesitamos convertir mensajes, llaves y textos cifrados en arreglos de enteros:

$$\begin{aligned}
 m &= \text{HOLAMUNDO} & \bar{m} &= (7, 15, 11, 0, 12, 21, 13, 3, 15) \\
 k &= \text{SECRETKEY} & \bar{k} &= (19, 4, 2, 18, 4, 20, 10, 4, 25) \\
 c &= \text{ZSNRPÑWHN} & \bar{c} &= (26, 19, 13, 18, 16, 14, 23, 7, 13)
 \end{aligned}$$

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

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

Naturalmente, vemos que siempre se cumple que $\bar{\bar{s}} = s$. Con esto, definimos OTP en base a

$$\begin{aligned}
 \text{Enc}_k(m) &= (\bar{m} + \bar{k}) \bmod 27 \\
 \text{Dec}_k(m) &= (\bar{c} - \bar{k}) \bmod 27
 \end{aligned}$$

Desde ahora supondremos que nuestros mensajes y llaves **son arreglos de números**. Con esta suposición en mente, definiremos OTP simplemente usando:

$$\begin{aligned}
 \text{Enc}_k(m) &= (m + k) \bmod 27 \\
 \text{Dec}_k(m) &= (c - k) \bmod 27
 \end{aligned}$$

Con esta definición, vemos entonces que

$$\text{Dec}_k(\text{Enc}_k(m)) = ((m + k) \bmod 27 - k) \bmod 27 = (m + k - k) \bmod 27 = m \bmod 27 = m$$

Generalizando, podemos describir $\text{OTP}^{N,\ell}$ sobre $\{0, 1, \dots, N-1\}$ y mensajes de largo ℓ , con:

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

Gen es la **distribución de probabilidad uniforme** sobre $\{0, 1, \dots, N-1\}$, y

$$\begin{aligned}
 \text{Enc}_k(m) &= (m + k) \bmod N \\
 \text{Dec}_k(c) &= (c - k) \bmod N
 \end{aligned}$$

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

$$\forall k \in \mathcal{K}. \forall m \in \mathcal{M} : \text{Dec}_k(\text{Enc}_k(m)) = m$$

En este caso diremos que el esquema es *perfectamente correcto*.

2.3. Perfect secrecy

Recordemos que un esquema criptográfico está compuesto por una tupla $(\text{Gen}, \text{Enc}, \text{Dec})$, con un espacio de mensajes \mathcal{M} , un espacio de llaves \mathcal{K} y un espacio de textos cifrados \mathcal{C} . ¿Cuándo podemos decir que un esquema así es *perfectamente secreto*? Podríamos decir algo como lo siguiente: “Al ver un texto cifrado c_0 pasar, para el atacante el mensaje original m podría haber sido cualquiera”.

¿Cómo formalizamos lo anterior? Dado un texto cifrado c_0 , se cumple que:

$$\forall m_0 \in \mathcal{M}. \quad \Pr_{k \sim \text{Gen}} [\text{Enc}_k(m_0) = c_0] = \frac{1}{|\mathcal{M}|}$$

La probabilidad anterior se calcula viendo la probabilidad de haber elegido una llave que encripte m_0 como c_0 :

$$\sum_{k \in \mathcal{K}: \text{Enc}_k(m_0) = c_0} \text{Gen}(k)$$

Ahora, ¿qué pasa si el atacante tiene información sobre el mensaje? En otras palabras, el atacante posee una distribución de probabilidad \mathbb{D} sobre \mathcal{M} (conoce la probabilidad de un mensaje). Para cada distribución de probabilidad \mathbb{D} sobre \mathcal{M} y cada texto cifrado $c_0 \in \mathcal{C}$ se cumple que:

$$\forall m_0 \in \mathcal{M}: \quad \Pr_{\substack{m \sim \mathbb{D} \\ k \sim \text{Gen}}} [m = m_0 \mid \text{Enc}_k(m) = c_0] = \Pr_{m \sim \mathbb{D}} [m = m_0]$$

Recordemos que $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$. Luego, ¿se cumple la siguiente ecuación?

$$\frac{\mathbb{D}(m_0) \sum_{k \in \mathcal{K}: \text{Enc}_k(m_0) = c_0} \text{Gen}(k)}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \sum_{k \in \mathcal{K}: \text{Enc}_k(m) = c_0} \text{Gen}(k)} \stackrel{?}{=} \mathbb{D}(m_0)$$

Para responder la pregunta anterior, volvamos a $\text{OTP}^{N, \ell}$ y analicemos si es *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 $\text{Enc}_k(m_0) = c_0$

Sea $c_0 \in \mathcal{C}$ un texto cifrado y m_0 un mensaje, entonces:

$$\frac{\mathbb{D}(m_0) \sum_{k \in \mathcal{K}: \text{Enc}_k(m_0) = c_0} \text{Gen}(k)}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \sum_{k \in \mathcal{K}: \text{Enc}_k(m) = c_0} \text{Gen}(k)} = \frac{\mathbb{D}(m_0) \cdot \frac{1}{N^\ell}}{\sum_{m \in \mathcal{M}} \mathbb{D}(m) \cdot \frac{1}{N^\ell}} = \mathbb{D}(m_0)$$

Por lo tanto, **OTP es un esquema perfectamente secreto** y luego, la ecuación anterior sí se cumple.

Como hemos podido ver, *perfect secrecy* es una condición muy fuerte. Lamentablemente, pareciera ser molesto y/o poco razonable que la llave tenga que ser tan larga como el mensaje. ¿Hay alguna forma de modificar OTP para tener $|\mathcal{K}| < |\mathcal{M}|$ y seguir teniendo un esquema criptográfico *perfectamente secreto*? Spoiler, no hay tal forma :(

Teorema 1

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 $(\text{Gen}, \text{Enc}, \text{Dec})$ que sea *perfectamente secreto*.

De hecho, en el **único tipo de ataque en que OTP es seguro es texto cifrado** (solo veo c_0, c_1, \dots). Esto se cumple solo si la llave **no se repite** cada vez que se quiera encriptar el mensaje. En los demás tipos de ataque que vimos anteriormente, OTP deja de ser seguro.

Demostración teorema. Supongamos que $|\mathcal{K}| < |\mathcal{M}| \leq \mathcal{C}$ y sea $(\text{Gen}, \text{Enc}, \text{Dec})$ 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 $\text{Enc}_k(m_0) = c_0$, por tanto

$$\Pr_{\substack{m \sim \mathbb{D} \\ k \sim \text{Gen}}} [m = m_0 \mid \text{Enc}_k(m_0) = c_0] \overset{0}{<} \mathbb{D}(m_0) = \Pr_{m \sim \mathbb{D}} [m = m_0]$$

■

Con todo lo anterior, podemos formalizar la definición de *perfect secrecy* para cualquier esquema criptográfico desde otra perspectiva.

Definición formal de perfect secrecy. Un esquema criptográfico $(\text{Gen}, \text{Enc}, \text{Dec})$ con un espacio de mensajes \mathcal{M} es **perfectamente secreto** si para cada distribución de probabilidad para M , para cada mensaje $m_0 \in \mathcal{M}$ y para cada texto cifrado $c \in \mathcal{C}$ para el cual $\Pr[C = c_0] > 0$:

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

con M y C las variables aleatorias para el mensaje y el texto cifrado.

Lema. Un esquema de encriptación $(\text{Gen}, \text{Enc}, \text{Dec})$ con un espacio de mensajes \mathcal{M} es perfectamente secreto si, y sólo si se cumple la siguiente ecuación para todo $m, m' \in \mathcal{M}$ y para todo $c \in \mathcal{C}$:

$$\Pr[\text{Enc}_k(m) = c] = \Pr[\text{Enc}_k(m') = c]$$

En otras palabras, la distribución de probabilidad del texto cifrado cuando ciframos m debe ser idéntica a la distribución de probabilidad del texto cifrado cuando se cifra m' .

2.4. Ahora sí, One-Time Pad

Primero, recordemos que $a \oplus b$ denota el OR-exclusivo (XOR) de dos *strings* binarios de igual longitud a y b (es decir, si $a = a_1 \cdots a_\ell$ y $b = b_1 \cdots b_\ell$ son *strings* de ℓ bits, entonces $a \oplus b$ es un *string* de ℓ bits dado por $(a_1 \oplus b_1) \cdots (a_\ell \oplus b_\ell)$). En OTP, la **llave** es un *string* uniforme de la **misma longitud** que el mensaje, y el texto cifrado se calcula simplemente haciendo XORing entre la llave y el mensaje.

Antes de hablar de la seguridad de OTP, verifiquemos su **corrección**: Para cada clave k y cada mensaje m se cumple que $\text{Dec}_k(\text{Enc}_k(m)) = \text{Dec}_k(k \oplus m) = k \oplus k \oplus m = m$, por lo que OTP constituye un esquema de cifrado válido.

Construcción del esquema. Fijemos un entero $\ell > 0$. El espacio de mensajes \mathcal{M} , el espacio de llaves \mathcal{K} y el espacio de textos cifrados \mathcal{C} son todos equivalentes a $\{0, 1\}^\ell$ (el conjunto de todos los *strings* binarios de largo ℓ).

1. **Gen**: el algoritmo de generación de llaves elige una desde $\mathcal{K} = \{0, 1\}^\ell$ de acuerdo a la distribución de probabilidad **uniforme** (es decir, cada uno de los 2^ℓ *strings* en el espacio es elegido como llave con probabilidad $1/2^\ell$).
2. **Enc**: dada una llave $k \in \{0, 1\}^\ell$ y un mensaje $m \in \{0, 1\}^\ell$, el algoritmo de encriptación retorna el texto cifrado $c = k \oplus m$.
3. **Dec**: dada una llave $k \in \{0, 1\}^\ell$ y un texto cifrado $c \in \{0, 1\}^\ell$, el algoritmo de decriptación retorna un mensaje $m = k \oplus c$.

Demostración de la construcción. Calculamos $\Pr[C = c \mid M = m]$ para un $c \in \mathcal{C}$ arbitrario y un $m \in \mathcal{M}$ con $\Pr[M = m] > 0$. Por OTP, tenemos que

$$\Pr[C = c \mid M = m] = \Pr[K \oplus m = c \mid M = m] = \Pr[K = m \oplus c \mid M = m] = \frac{1}{2^\ell}$$

donde M , K y C son las variables aleatorias para el mensaje, la llave y el texto cifrado, que son **independientes** entre sí. Aquí, la primera ecuación es por definición del esquema y el hecho que condicionamos el evento de que $M = m$, y la última ecuación es válida porque la llave K es un ℓ -bit uniforme que es independiente de M . Fijemos cualquier distribución de probabilidad sobre \mathcal{M} . Usando el resultado de arriba, vemos que para cualquier $c \in \mathcal{C}$ tenemos:

$$\Pr[C = c] = \sum_{m \in \mathcal{M}} \Pr[C = c \mid M = m] \cdot \Pr[M = m] = \frac{1}{2^\ell} \cdot \sum_{m \in \mathcal{M}} \Pr[M = m] = \frac{1}{2^\ell}$$

donde la suma es sobre $m \in \mathcal{M}$ con $\Pr[M = m] \neq 0$. Por el Teorema de Bayes:

$$\Pr[M = m \mid C = c] = \frac{\Pr[C = c \mid M = m] \cdot \Pr[M = m]}{\Pr[C = c]} = \frac{2^{-\ell} \cdot \Pr[M = m]}{2^{-\ell}} = \Pr[M = m]$$

Concluimos entonces que OTP es *perfectamente secreto*. ■

2.5. Permutaciones pseudo-aleatorias (PRP)

Recordemos que una noción de seguridad debe incluir:

- ♦ Un modelo de amenaza (tipos de ataque), que define las capacidades de un **adversario**.
- ♦ Una garantía de seguridad, lo cual normalmente se traduce en definir qué significa que el adversario no tenga éxito en su ataque.

Vamos a definir una primera noción de seguridad que nos va a permitir mostrar que OTP no es seguro si la llave es **reutilizada**, como mencionamos anteriormente.

Consideremos un largo n fijo tal que los espacios de mensajes, llaves y textos cifrados cumplen que $\mathcal{M} = \mathcal{K} = \mathcal{C} = \{0, 1\}^n$ y $(\text{Gen}, \text{Enc}, \text{Dec})$ es un esquema criptográfico. Veamos un **juego** para definir una **pseudo-random permutation** (PRP):

1. Un **verificador** elige $b \in \{0, 1\}$ con distribución uniforme (por ejemplo, tirar una moneda).
 - 1.1. Si $b = 0$, entonces elige una clave $k \in \mathcal{K}$ según la distribución **Gen** y define $f(x) = \text{Enc}_k(x)$.
 - 1.2. Si $b = 1$, entonces elige una permutación π con distribución uniforme y define $f(x) = \pi(x)$.
2. El **adversario** elige una palabra $y \in \{0, 1\}^n$ y el verificador responde con $f(y)$.
3. El paso 2. es repetido q veces.
4. El adversario indica si $b = 0$ o $b = 1$, y gana si su elección es correcta.

La probabilidad de que el adversario gane depende de la cantidad de rondas q . Si el adversario “tira una moneda” en el paso 4., entonces su probabilidad de ganar es $\frac{1}{2}$.

Definición de PRP. Decimos que $(\text{Gen}, \text{Enc}, \text{Dec})$ es una permutación pseudo-aleatoria si **no existe** un adversario que pueda ganar el juego con una probabilidad significativamente mayor a $\frac{1}{2}$.

No imponemos restricciones en las capacidades computacionales del adversario, podríamos tener una noción más débil donde el adversario puede realizar una cantidad de operaciones que es polinomial en n . Por ejemplo, solo puede realizar n^2 operaciones. Debemos entonces definir qué significa que *la probabilidad de ganar el juego sea significativamente mayor a $\frac{1}{2}$* (podría ser $\frac{3}{4}$, por ejemplo). También, tenemos que indicar cuál es el número de rondas q .

Un primer ejemplo. Considere un esquema $(\text{Gen}, \text{Enc}, \text{Dec})$ tal que:

- ♦ $\text{Gen}(k_0) = 1$ para una llave fija $k_0 \in \mathcal{K}$.
- ♦ $\text{Gen}(k) = 0$ para todo $k \in \mathcal{K}$ tal que $k \neq k_0$.

Vamos a demostrar que este esquema criptográfico no es una PRP con una ronda ($q = 1$). La estrategia del adversario es la siguiente:

- ♦ En el paso 2. toma $y = 0^n$ y recibe $f(y)$ como respuesta del verificador.
- ♦ Si $f(y) = \text{Enc}_{k_0}(y)$ entonces indica que $b = 0$, si no, indica que $b = 1$.

¿Puede el adversario equivocarse al indicar el valor de b ? ¿Cuál es la probabilidad de que gane el juego? Veamos. Definamos A como el evento “Adversario gana el juego”, entonces:

$$\begin{aligned}\Pr[A] &= \Pr[A \mid b = 0] \cdot \Pr[b = 0] + \Pr[A \mid b = 1] \cdot \Pr[b = 1] \\ &= \Pr[A \mid b = 0] \cdot \frac{1}{2} + \Pr[A \mid b = 1] \cdot \frac{1}{2}\end{aligned}$$

Vemos que:

- ♦ $\Pr[A \mid b = 0] = 1$
- ♦ $\Pr[A \mid b = 1] = \Pr[\pi(y) \neq \text{Enc}_{k_0}(y)]$

En este caso el verificador elige una permutación $\pi : \{0, 1\}^n \rightarrow \{0, 1\}^n$ con distribución uniforme. Si $\pi(y) = \text{Enc}_{k_0}(y)$, entonces el adversario da la respuesta equivocada. Luego,

$$\begin{aligned}\Pr[\pi(y) \neq \text{Enc}_{k_0}(y)] &= 1 - \Pr[\pi(y) = \text{Enc}_{k_0}(y)] \\ &= 1 - \frac{\text{casos favorables}}{\text{casos totales}} \\ &= 1 - \frac{(2^n - 1)!}{(2^n)!} \\ &= 1 - \frac{1}{2^n}\end{aligned}$$

De la ecuación anterior:

- ♦ **Casos totales:** número de permutaciones $\pi : \{0, 1\}^n \rightarrow \{0, 1\}^n$.
- ♦ **Casos favorables:** número de permutaciones $\pi : \{0, 1\}^n \rightarrow \{0, 1\}^n$ tales que $\pi(0^n)$ es igual al valor fijo $\text{Enc}_{k_0}(0^n)$.

Finalmente, tenemos que

$$\begin{aligned}\Pr[A] &= \Pr[A \mid b = 0] \cdot \frac{1}{2} + \Pr[A \mid b = 1] \cdot \frac{1}{2} \\ &= 1 \cdot \frac{1}{2} + \left(1 - \frac{1}{2^n}\right) \cdot \frac{1}{2} \\ &= 1 - \frac{1}{2^{n+1}}\end{aligned}$$

Por ende, el adversario **sí** gana el juego con probabilidad significativamente mayor a $\frac{1}{2}$. Puede verificar este hecho tomando $n = 10$ o $n = 100$.