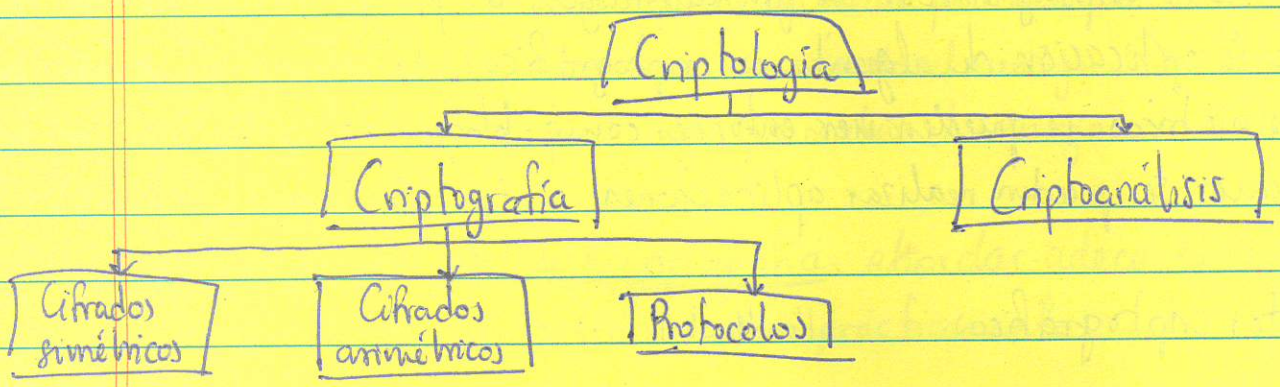
Cap. 1. Introducción a la Criptografía y Seguridad de Datos



→ El término criptología es el más general. Se divide en dos grandes ramas:

-) La criptografía es la ciencia de la escritura secreta con el propósito de esconder el significado de un mensaje.

-) El criptoanálisis es la ciencia (y a veces arte) de romper criptosistemas. Es la única forma de asegurar que un criptosistema es seguro.

→ La criptografía se divide en tres ramas:

-) Algoritmos simétricos (llave privada): dos entidades tienen un método de cifrado y descifrado para el cual comparten una llave secreta. Toda la criptografía desde la antigüedad hasta 1976 era exclusivamente de llave privada.

-) Algoritmos asimétricos (llave pública): un usuario tiene una llave secreta pero también una llave pública.

-) Protocolos criptográficos. A grandes rasgos, los protocolos criptográficos tratan de la aplicación de algoritmos criptográficos. Los algoritmos simétricos y asimétricos se pueden ver entonces como bloques de construcción con los cuales se pueden realizar aplicaciones. Ej: TLS (Transport Layer Security) → navegadores web

→ Propósitos criptográficos: (Handbook)

① Confidencialidad = Mantener en secreto el contenido de la información de todos excepto de aquellos autorizados a tenerla.
Garantizar que la información no sea entendida por otras personas

② Integridad de los datos = Se ocupa de la alteración no autorizada de los datos
El receptor del mensaje debe poder verificar si el mensaje fue ~~alterado~~ modificado durante su envío

③ Autenticación → Relacionado a identificación
↳ aplica a ~~cuentas~~ tanto entidades como a la información

* Las dos partes que se comunican deben poder identificarse mutuamente al inicio de la "conversación"

* La información transmitida debe ser autenticada
↳ Origen ↳ Fecha de ~~pro~~ origen ↳ Contenido ↳ Hora de envío

Se subdivide en autenticación de entidades y autenticación de origen de los datos
implícitamente ←
da integridad de los datos
(si el mensaje se modifica, el origen ha cambiado)

④ No repudio- Impide a una de las entidades negar acciones anteriores
↳ Es necesario un medio para resolver la situación.

Ej. El emisor no puede negar que mandó el mensaje
El receptor no puede negar que recibió el mensaje

Objetivo fundamental de la criptografía: abordar adecuadamente estas cuatro áreas, tanto teórica como prácticamente

→ Terminología básica y conceptos (Handbook, 1.4)

* Dominios y codominios de cifrado

- $\mathcal{A}$ denota un conjunto finito llamado alfabeto de definición.
  Ej. $\mathcal{A} = \{0, 1\}$, el alfabeto binario, es el alfabeto de definición más usado frecuentemente.

- $\mathcal{M}$ denota un conjunto llamado espacio de mensajes. $\mathcal{M}$ consiste de cadenas de símbolos de un alfabeto de definición.
  Un elemento de $\mathcal{M}$ es llamado mensaje en texto plano o simplemente texto plano.

  Ej. $\mathcal{M}$ = cadenas binarias / textos en español / inglés / códigos de computadora

- $\mathcal{C}$ denota un conjunto llamado espacio de texto cifrado. $\mathcal{C}$ consiste de cadenas de símbolos de un alfabeto de definición, el cual puede diferir del alfabeto de definición para $\mathcal{M}$.
  Un elemento de $\mathcal{C}$ es llamado un texto cifrado.

* Transformaciones de cifrado y descifrado

- K denota un conjunto llamado espacio de llaves.
  Un elemento de K es llamado llave.

- Cada elemento $e \in K$ determina de forma única una ~~biyección~~ inyección de $\mathcal{M}$ a $\mathcal{C}$ denotada por E_e, la cual es llamada función o transformación de cifrado.

  Obs E_e es una biyección $E_e: \mathcal{M} \longrightarrow E_e(\mathcal{M}) \subseteq \mathcal{C}$

  Nos interesa que E_e sea biyección si queremos revertir el proceso y recuperar un único texto plano de cada texto cifrado distinto.

- Para cada $d \in K$, D_d denota una biyección de $\mathcal{C}$ a $\mathcal{M}$ y es llamada función o transformación de descifrado.

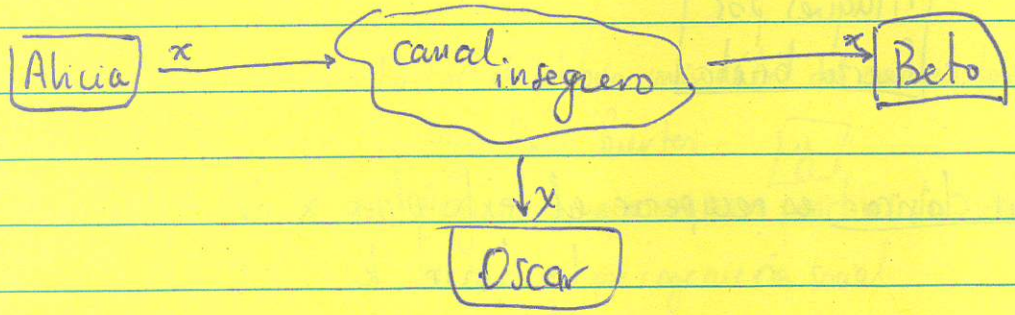
→ Principios de Kerckhoffs (1883) (Handbook)

1. El sistema debe ser, si bien no teóricamente, sí prácticamente irrompible
2. Comprometer el sistema en cuanto a detalles no debe ser un inconveniente para su efectividad
3. La llave debe poder recordarse sin notas y debe poder cambiarse fácilmente
4. El sistema debe poder aplicarse a la correspondencia telegráfica (arrojar resultados alfanuméricos)
5. El aparato de cifrado debe ser portable y operable por una sola persona
6. El sistema debe ser fácil de usar, sin una lista larga de reglas o condiciones

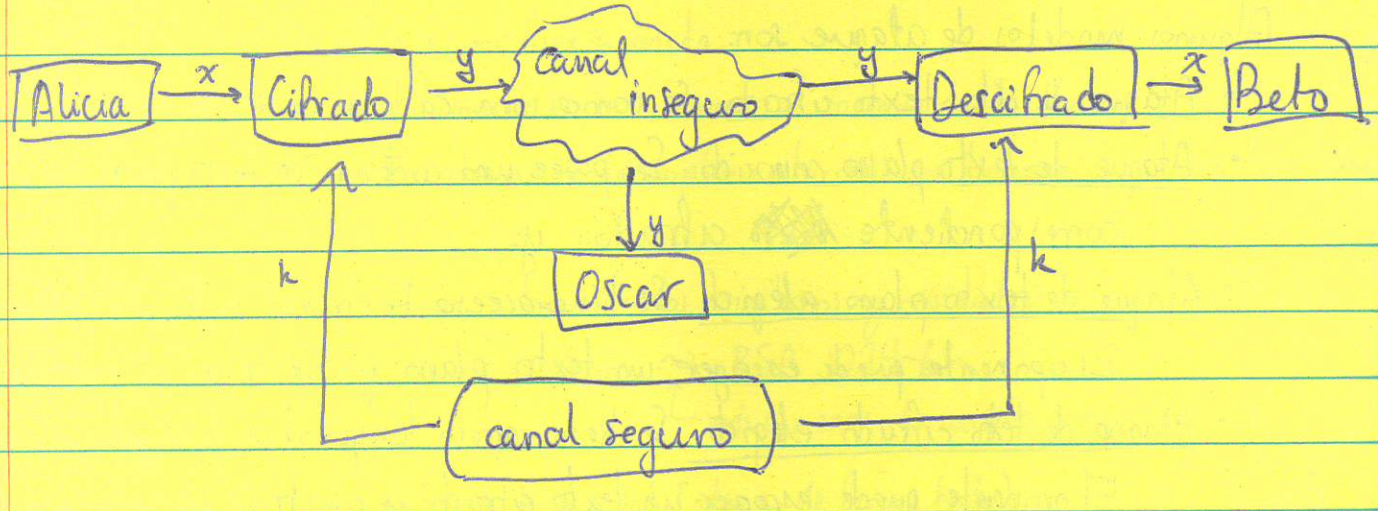
El punto 2 permite que las funciones de cifrado puedan ser públicas y que la seguridad del sistema resida únicamente en la llave elegida.

1.2 Criptografía simétrica

La criptografía simétrica se puede introducir mejor con un problema fácil de entender: hay dos usuarios, Alicia y Beto, que se quieren comunicar mediante un canal inseguro. El problema empieza con Oscar, que tiene acceso a este canal (espionaje).

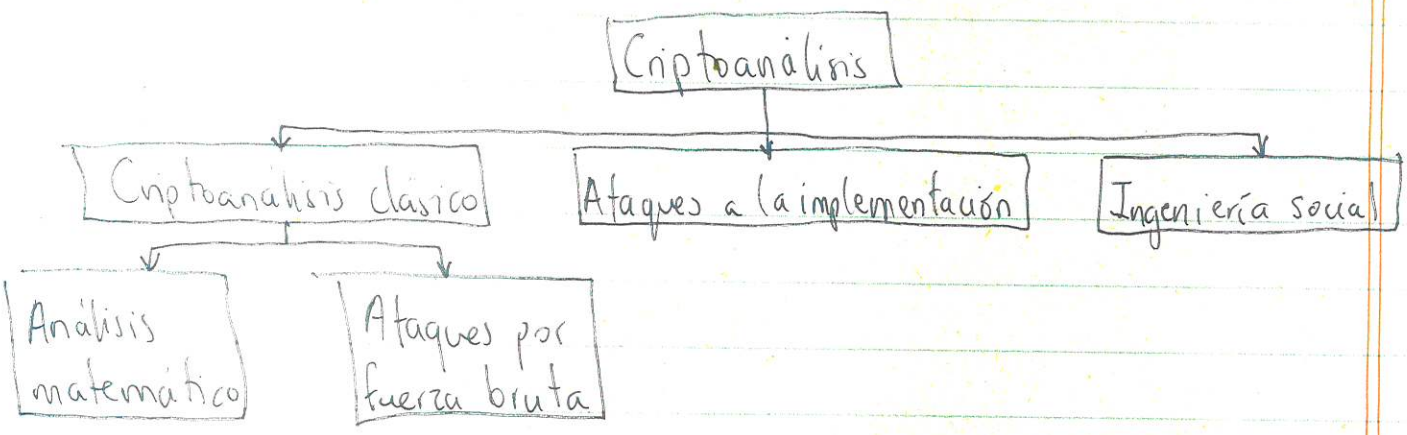


En este caso, la criptografía simétrica ofrece una solución: Alicia cifra su mensaje usando un algoritmo simétrico y obtiene el texto cifrado y. Bob recibe el texto cifrado y descifra el mensaje para obtener x.



El sistema necesita un canal seguro para distribuir la llave entre Alicia y Beto, lo cual puede resultar engorroso.

1.3 Criptoanálisis



→ Criptoanálisis clásico: es recuperar el texto plano x del texto cifrado y la llave k " " " "

↳ Análisis matemático : explorar la estructura interna del método de cifrado

↳ Ataques por fuerza bruta: tratar al algoritmo como una caja negra y probar todas las posibles llaves.

Algunos modelos de ataque son:

- Ataque de sólo texto cifrado. Se conoce una cadena de texto cifrado, y.
- Ataque de texto plano conocido. Se posee una cadena de texto plano, x, y su correspondiente cifrado, y.
- Ataque de texto plano elegido. Se tiene acceso temporal a la maquinaria de cifrado. El oponente puede escoger un texto plano y construir el correspondiente cifrado y
- Ataque de texto cifrado elegido. Se tiene acceso temporal a la maquinaria de descifrado. El oponente puede escoger un texto cifrado y construir el correspondiente descifrado

En cada caso, el objetivo es determinar la llave k usada.

→ Ataques a la implementación.

El ataque de canal lateral (side-channel attack) se puede usar para obtener una llave, por ejemplo, midiendo el consumo de energía eléctrica de un procesador que opera con dicha llave.

↳ Son más útiles cuando tenemos acceso físico al dispositivo

→ Ingeniería social.

¡¡ Un atacante siempre va a buscar el eslabón más débil del criptosistema !!

Por eso hay que buscar: ① Algoritmos fuertes y

② Asegurarnos que no sean prácticos los ataques por implementación ni por ingeniería social.

deberían seguir a Kerckhoffs

→ ¿Cuántos bits debe tener la llave? (Necesario pero no suficiente)

↳ simétrico

① Para cifrados de llave simétrica esta discusión sólo es relevante si el único ataque plausible es un ataque por fuerza bruta.

↳ Análisis de frecuencias → no sirve conocer el tamaño de la llave

↳ Ingeniería social o ataques a la implementación → tampoco ayuda una llave larga.

② La longitud de las llaves entre cifrados simétricos y asimétricos es muy diferente.

Ej: Llave de 80 bits de un simétrico ~ RSA 1024-bits

↳ más o menos la misma seguridad

simétrico

Tamaño de la llave	Estimado de seguridad (ataques de fuerza bruta)
56 - 64 bits	Horas o días
112 - 128 bits	Décadas en ausencia de computadoras cuánticas

2.3 – 2.5 Sandy

EJEMPLOS:

① **Corrimiento** Definido sobre $\mathbb{Z}/26\mathbb{Z}$ (o $\mathbb{Z}/27\mathbb{Z}$ si consideramos la ñ) pero puede ser definido sobre $\mathbb{Z}/m\mathbb{Z}$ para cualquier m.

Tomamos $\mathcal{M} = \mathcal{C} = \mathcal{K} = \mathbb{Z}/26\mathbb{Z}$. Para $0 \leq K \leq 25$, definimos

$$e_K(x) = (x+K) \bmod 26, \quad d_K(y) = (y-K) \bmod 26$$

con $x, y \in \mathbb{Z}/26\mathbb{Z}$.

Una tabla quedaría de la siguiente forma:

A	B	C	D	E	F	G	H	I	J	K	L	M	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

Ej: $K = 11$, x = ~~[illegible]~~ wewillmeetatmidnight

y = HPHTWWXPPELEXTOYTRSE

Forma de criptoanálisis 1: búsqueda exhaustiva (fuerza bruta) de la llave.

Hay 26 posibles llaves, es "fácil" buscar de una por una.

En promedio, se puede descifrar un texto plano en 13 iteraciones ($13 = 26/2$)

② **Sustitución.** Tomamos $\mathcal{M} = \mathcal{C} = \mathbb{Z}/26\mathbb{Z}$. $\mathcal{K}$ consiste de todas las posibles permutaciones de los 26 símbolos, i.e., $\mathcal{K} = S_{26}$.

Para cada $\pi \in \mathcal{K}$, definimos

$$e_\pi(x) = \pi(x) \quad \text{y} \quad d_\pi(y) = \pi^{-1}(y)$$

con $x, y \in \mathbb{Z}/26\mathbb{Z}$.

Note que $|\mathcal{K}| = 26!$, aún que una búsqueda exhaustiva no es computacionalmente plausible.

Forma de criptoanálisis 2: probabilidad de ocurrencia de las letras en el idioma que estemos usando.

↳ estamos asumiendo idioma ordinario sin caracteres especiales ni signos de puntuación ni espacios

↳ el tipo más débil de criptoanálisis.

→ Checar Stinson para un ejemplo: p. 7, p. 28

(3) Hill: sistema polialfabético: puede cifrar m caracteres al mismo tiempo.
Sea $m \in \mathbb{Z}^+$, $\mathcal{M} = \mathcal{C} = (\mathbb{Z}/26\mathbb{Z})^m$.

La idea es tomar m combinaciones lineales de m caracteres de un texto plano, por lo que se producen m caracteres cifrados. Ent. $y = xK$ con

$$(y_1, \dots, y_m) = (x_1, \dots, x_m) \begin{pmatrix} k_{11} & \cdots & k_{1m} \\ \vdots & & \vdots \\ k_{m1} & \cdots & k_{mm} \end{pmatrix} \in K$$
$$\in \mathcal{M}$$

Ent. "$K = M_m(\mathbb{Z}/26\mathbb{Z})$". Decimos que el texto cifrado se obtiene del texto plano por una transformación lineal.

Note que, para descifrar, necesitaríamos K^{-1}.

Rec. (1) Sup. $K = (k_{ij}) \in M_m(\mathbb{Z}/n\mathbb{Z})$ con $\det K$ invertible en $\mathbb{Z}/n\mathbb{Z}$. Ent. $K^{-1} = \det(K)^{-1} K^*$ con K^* la adjunta de K.

Más precisamente, sea $m \geq 2$ un entero. Sean $\mathcal{M} = \mathcal{C} = (\mathbb{Z}/26\mathbb{Z})^m$. Sea $K = GL(m, \mathbb{Z}/26\mathbb{Z})$. Para $K \in K$, definimos

$$e_K(x) = xK \quad \text{y} \quad d_K(y) = yK^{-1}$$

Este cifrado puede ser difícil de descifrar por un ataque de sólo texto cifrado, pero cae fácilmente con un ataque de texto plano conocido.

Supongamos que el oponente ha determinado el valor m que se usó. Supongamos que él tiene al menos m pares de textos cifrados y planos, i.e., conoce

$x_j = (x_{1j}, x_{2j}, \dots, x_{mj})$ y $y_j = (y_{1j}, y_{2j}, \dots, y_{mj})$

para $1 \leq j \leq m$ tales que $y_j = e_K(x_j)$, $1 \leq j \leq m$.

Si definimos $X = (x_{ij}) = (x_1 \cdots x_m)$ y $Y = (y_{ij})$, tenemos la ecuación matricial $Y = XK$, con $K \in M_m(\mathbb{Z}/26\mathbb{Z})$ la matriz que buscamos. Como X es invertible ($X = d_k e_k(X)$), el oponente puede calcular $K = X^{-1}Y$. Si X no es invertible, habrá que probar otro conjunto de m pares de textos plano-cifrados.

Si el oponente no conoce m, (y suponiendo que m no es tan grande), él simplemente podría probar con $m = 2, 3, \dots$ hasta encontrar la llave.
Esto es plausible si no hay límite de tiempo.