
	<p align="center">UNIVERSIDAD NACIONAL DE SAN AGUSTIN FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMA</p>	
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		
Aprobación: 2022/03/01	Código: GUIA-PRLD-001	Página: 1

GUÍA DE LABORATORIO 01

INFORMACIÓN BÁSICA					
ASIGNATURA:	<i>Escribir nombre del curso</i>				
TÍTULO DE LA PRÁCTICA:	<i>FUNCIONES ELEMENTALES DE LA CRIPTOGRAFÍA</i>				
NÚMERO DE PRÁCTICA:	<i>01</i>	AÑO LECTIVO:	<i>2023</i>	NRO. SEMESTRE:	<i>A</i>
TIPO DE PRÁCTICA:	INDIVIDUAL				
	GRUPAL	MÍNIMO DE ESTUDIANTES	<i>1</i>	MÁXIMO DE ESTUDIANTES	<i>2</i>
FECHA INICIO:	<i>09/06/2023</i>	FECHA FIN:	<i>15/06/2023</i>	DURACIÓN:	<i>7 días</i>
RECURSOS Y EQUIPOS A UTILIZAR: <i>Por Ejemplo: (VSCode, AndroidStudio, smartphone, PC, ..etc)</i>					
DOCENTE(s): <i>Juan Carlos Zuñiga</i>					

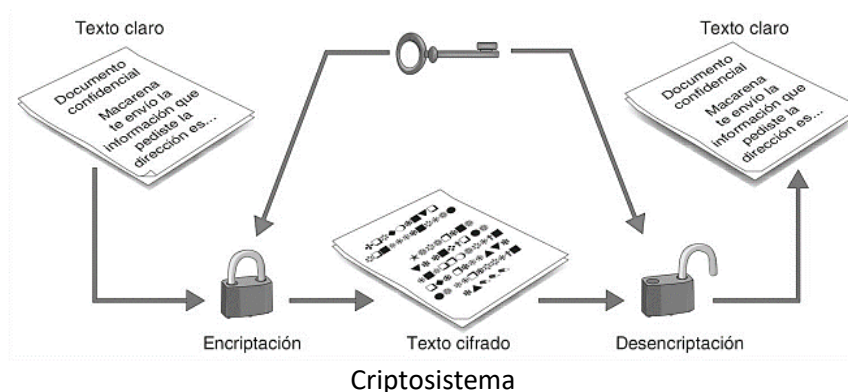
OBJETIVOS/TEMAS Y COMPETENCIAS	
OBJETIVOS:	<ul style="list-style-type: none"> <i>Implementar funciones básicas sobre las técnicas criptográficas, consideradas en la etapa de preprocesamiento, utilizando herramientas de desarrollo de libre elección</i>
TEMAS:	<ul style="list-style-type: none"> <i>Preprocesamiento</i> <i>Sustitución de caracteres</i> <i>Conversión de caracteres</i> <i>Eliminación de caracteres no significativos</i> <i>Padding o relleno</i> <i>Sustitución numérica</i>
COMPETENCIAS A ALCANZAR	<p><i>Mantiene responsablemente, software para que se adecue a las necesidades cambiantes del usuario, cliente o sociedad mediante la aplicación de técnicas y procedimientos establecidos que siguen estándares de calidad destinados a implementar la seguridad informática. (referencia C.n)</i></p> <p><i>Asegura la calidad del software mediante la aplicación de pruebas, validaciones y estándares de seguridad para garantizar el correcto funcionamiento del producto, en el marco de la seguridad informática, considerando el impacto productivo y social. (referencia C.o)</i></p> <p><i>Diseña soluciones informáticas apropiadas para proveer seguridad informática, utilizando los principios de ingeniería que integran consideraciones éticas, sociales, legales y económicas entiendo las fortalezas y limitaciones del contexto (referencia C.q)</i></p>

CONTENIDO DE LA GUÍA

I. MARCO CONCEPTUAL

1. Cifrado

Se entiende el cifrado como el proceso que convierte cualquier mensaje o dato(s) en otro ilegible, el sistema que implementa este proceso recibe el nombre de criptosistema y tiene la siguiente estructura:





En todo proceso de cifrado se identifica:

- El texto claro o texto plano (*plain text*) es el mensaje legible u original que debe protegerse mediante el cifrado.
- El criptograma o texto cifrado (*cypher*) es el mensaje ilegible que se obtiene al aplicar el proceso de cifrado
- La encriptación o cifrado es la aplicación de un algoritmo de cifrado o cifra sobre el texto claro, este requiere un dato protegido denominado clave o llave de cifrado, un criptosistema es seguro al ocultar no el algoritmo, sino la clave
- La desencriptación o descifrado es el proceso de convertir el texto cifrado en el texto plano original a partir de la aplicación de un algoritmo que revierte el proceso anterior, requiere también el conocimiento de la clave o llave

Todo este conjunto estructurado de elementos recibe el nombre de **Criptosistema**.

Se denomina **alfabeto** al conjunto de caracteres posibles de encontrar en el texto claro, que no necesariamente coincide con el juego de caracteres que se usa en el mensaje cifrado. Finalmente, en algunos algoritmos, el texto cifrado se escribe en base a bloques de longitud constante, denominados grupos, lo que se usa como verificación adicional al comprobar que el texto cifrado tiene un número entero de los mismos, para ello se puede aplicar un **padding** o relleno del texto claro inicial con ceros o algún carácter especial.

	<p style="text-align: center;">UNIVERSIDAD NACIONAL DE SAN AGUSTIN FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMA</p>	
<p style="text-align: center;">Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación</p>		
<p>Aprobación: 2022/03/01</p>	<p>Código: GUIA-PRLD-001</p>	<p>Página: 3</p>

2. Preprocesado

Se denomina preprocesado al conjunto de funciones aplicadas sobre el texto inicial para generar el texto claro antes de aplicarle el algoritmo de cifrado, con la finalidad de que este corra apropiadamente y el texto cifrado resultante sea más resistente al criptoanálisis o ataque para romper el cifrado, dentro de ello tenemos

a) *Modificación del alfabeto por sustitución*

Hay algoritmos de cifrado definidos sobre algún alfabeto en particular lo que exige sustituir los caracteres que no corresponden a este por otros, es típico los problemas del español con las letras **h,j,ñ,k,u,w,y** que no pertenecen al alfabeto latino, por lo que deben sustituirse por otros; también es común los alfabetos que usan sólo mayúsculas

b) *Inclusión de caracteres nulos*

Son palabras o grupos de caracteres sin significados insertados en medio del texto claro o para romper letras repetidas, por ejemplo, **rr, ll** u otro.

c) *Eliminación de caracteres no significativos*

Eliminar signos de puntuación o espacios en blanco

d) *Reemplazo de carretillas*

Reemplazar por otro texto palabras o frases muy comunes o repetitivas

e) *Reemplazo numérico*

Los caracteres deben ser reemplazados por valores numéricos ASCII, UNICODE-8, como, por ejemplo

"Hello World" = 0x48656C6C6F20576F726C64



o cada palabra por el valor numérico resultante de sumar la posición de cada carácter en el alfabeto por el peso según la posición del carácter en la palabra

"HIJO" = $7 \times 26^3 + 8 \times 26^2 + 9 \times 26^1 + 14 \times 26^0 = 128688$

3. Método Kasiski (tabla de frecuencias)

Cuando en Criptografía se reemplaza algún carácter por otro (sustitución), es evidente que el comportamiento estadístico del carácter que se ha insertado (frecuencia de aparición) es el mismo que el del carácter inicial en el alfabeto de origen, lo que define su aparición en el texto claro.

En el caso del español, la letra E es la más frecuente en el texto en claro, entonces en el nuevo texto que se ha creado reemplazando caracteres según una clave de longitud L constante, con mucha probabilidad la letra E seguirá siendo también aquí la más frecuente.

	<p style="text-align: center;">UNIVERSIDAD NACIONAL DE SAN AGUSTIN FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMA</p>	
<p style="text-align: center;">Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación</p>		
<p>Aprobación: 2022/03/01</p>	<p>Código: GUIA-PRLD-001</p>	<p>Página: 4</p>

Para que esta propiedad se manifieste de forma adecuada y en consecuencia nos entregue información válida que permita iniciar un ataque, se deberá contar con la cantidad suficiente de texto cifrado. En un módulo de cifra de 27 letras, las mayúsculas del alfabeto español, se ha demostrado que con 100 o más letras del texto cifrado, poco más de 3 veces el tamaño del alfabeto, ya se manifiesta la redundancia del lenguaje.

II. EJERCICIO RESUELTO

Conversión a números.

Algunos algoritmos de cifrado necesitan convertir los caracteres en números. Se pueden seguir distintas estrategias. Por ejemplo:

- Podríamos usar una tabla de conversión de los caracteres en números usando algún sistema de codificación de caracteres como ASCII o Unicode. Por ejemplo, el mensaje "Hello World" usando Unicode-8 quedaría:

```
48 65 6C 6C 6F 20 57 6F 72 6C 64
```

Esta cadena de bytes podríamos convertirla en un número concatenando obteniendo:

```
0x48656C6C6F20576F726C64=87521618088882533792115812
```



- Se podría considerar que los caracteres están ordenados según un criterio e interpretar la cadena como un número con base el número de caracteres del alfabeto. Por ejemplo, si consideramos que solo hay caracteres en mayúsculas y los ordenamos según el orden alfabético tendríamos, por ejemplo:

```
"HIJO" 7x263+8x262+9x261+14x260=128688
```

III. EJERCICIOS PROPUESTOS

Sobre el texto claro mostrado a continuación:

*Mi corazón oprimido
Siente junto a la alborada
El dolor de sus amores
Y el sueño de las distancias.
La luz de la aurora lleva
Semilleros de nostalgias
Y la tristeza sin los ojos
De la médula del alma.*

	<p style="text-align: center;">UNIVERSIDAD NACIONAL DE SAN AGUSTÍN FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMA</p>	
<p style="text-align: center;">Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación</p>		
<p>Aprobación: 2022/03/01</p>	<p>Código: GUIA-PRLD-001</p>	<p>Página: 5</p>

*La gran tumba de la noche
 Su negro velo levanta
 Para ocultar con el día
 La inmensa cumbre estrellada.
 ¡Qué haré yo sobre estos campos
 Cogiendo niños y ramas
 Rodeado de la aurora
 Y llena de noche el ama!
 ¡Qué haré si tienes tus ojos
 Muertos a las luces claras
 Y no ha de sentir mi carne
 El calor de tus miradas!
 ¿Por qué te perdí por siempre
 En aquella tarde clara?
 Hoy mi pecho está reseco
 Como una estrella apagada.*

ALBA, Federico García Lorca

Implementar las siguientes operaciones de preprocesamiento, en cada caso debe mostrar el código de la implementación de la operación y la salida parcial resultante en cada paso (podrá usar como herramienta de desarrollo C++, python o java, **pero no por una librería**)

4.1 Realizar las siguientes **sustituciones**: axo, hxi, ñxm, kxl, uxv, wxv, zxy, xxr (tanto mayúsculas como minúsculas).

4.2 **Elimine** las tildes

4.3 **Convierta** todas las letras a mayúsculas

4.4 **Elimine** los espacios en blanco y los signos de puntuación Indique cuál sería el alfabeto resultante y cuál su longitud



GUARDE EL RESULTADO EN EL ARCHIVO "POEMA_PRE.TXT" (el que deberá ser adjuntado)

4.5 Abra el archivo generado e implementar una función que calcule una tabla de frecuencias para cada letra de la 'A' a 'Z'. La función deberá definirse como **frecuencias(archivo)** y deberá devolver un diccionario cuyos índices son las letras analizadas y cuyos valores son las frecuencias de las mismas en el texto (número de veces que aparecen). Reconozca en el resultado obtenido los cinco caracteres de mayor frecuencia

4.6 Obtener la información que el método **Kasiski** requiere para implementar un ataque, para ello deberá recorrer el texto preprocesado y hallar los trigramas en el mismo (sucesión de tres letras seguidas que se repiten) y las distancias (número de caracteres entre dos trigramas iguales consecutivos)

4.7 Volver a preprocesar el archivo cambiando cada carácter según **UNICODE-8**

4.8 Volver a preprocesar el archivo cambiando cada carácter según alfabeto de su elección

	<p align="center">UNIVERSIDAD NACIONAL DE SAN AGUSTIN FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMA</p>	
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		
Aprobación: 2022/03/01	Código: GUIA-PRLD-001	Página: 6

Volver a preprocesar el archivo insertando la cadena **AQP** cada 20 caracteres, el texto resultante deberá contener un número de caracteres que sea múltiplo de 4, si es necesario rellenar (*padding*) al final con caracteres **X** según se necesite

IV. CUESTIONARIO

1. Describa alguna otra operación o función de preprocesamiento que se implemente sobre el texto claro en los criptosistemas, justifique ¿por qué esta etapa es necesaria?
2. ¿Qué riesgo implicaría el implementar el preproceso de la información?

V. REFERENCIAS Y BIBLIOGRAFÍA RECOMENDADAS:

Pousa, A. (2011). Algoritmo de cifrado simétrico AES (Doctoral dissertation, Universidad Nacional de La Plata).

TÉCNICAS E INSTRUMENTOS DE EVALUACIÓN	
TÉCNICAS: <i>Ejercicios propuestos</i>	INSTRUMENTOS: <i>Lista de cotejo</i>
CRITERIOS DE EVALUACIÓN Y LOGROS ALCANZADOS <i>Niveles de logro: inicio, proceso, logro esperado, logro destacado.</i>	