Una mejora de la clave para el cifrado con llave simétrica, utilizando algoritmos genéticos

Diego Gómez, Leandro Ulloa
Ataque y protección de sistemas informáticos, ITCR
Cartago, Costa Rica
luillidei@gmail.com
joslup9@gmail.com

***Abstract*—La seguridad informática, es una de los campos de la ciencia de la computación que ha tenido un fuerte auge en el día-día de las empresas y de los individuos del mundo. Esto a raíz de los elementos de la seguridad, el cual se deben de proteger. La disponibilidad, la integridad y la confidencialidad, tres aristas de la arquitectura informática, que se equilibran para alcanzar un sentimiento de seguridad según las necesidades inmediatas de los individuos. Esta investigación se basó sobre el tema de algoritmos genéticos para la mejora de claves para cifrar en algoritmos de cifrado simétrico. Se desarrolló un algoritmo genético canónico, para realizar los procesos básicos de codificación, selección, cruce y mutación, así con ello generar una nueva clave introducida por un usuario base y como resultado, una clave con baja probabilidad de ser descifrada con técnicas de fuerza bruta, como por ejemplo el uso de diccionarios, también bajando casi por completo la eficacia de estrategias de ingeniería social, por el simple hecho, que la contraseña al ser cifrada con el AG, la clave resultante que entra a la función simétrica, no será la misma, por lo tanto, el AG tendría un rol de llave privada, portada únicamente por el propietario y siendo también la única llave que descifra la clave cifrada mediante el algoritmo simétrico.**

I. INTRODUCCIÓN

La seguridad de la información juega un papel muy importante para quienes se interesan por proteger sus datos. Existe una variedad de algoritmos de cifrado, los cuales permiten ocultar la información, modificando el contenido, sin alterar la integridad de los datos. Es decir, es posible recuperar el documento encriptado. Según [2] existen dos técnicas básicas de cifrado, simétrico o asimétrico.

Según [1], el cifrado asimétrico o cifrado de clave pública es una técnica de securización de información que en lugar de proteger la información tras una clave de acceso a la misma, requiere de dos claves relacionadas entre sí. Una de estas claves se usa para codificar (proteger) la información y la otra se usa para decodificar (acceder) la información previamente cifrada.

También, como menciona [3], un sistema de cifrado simétrico es un tipo de cifrado que usa una misma clave para cifrar y para descifrar. Las dos partes que se comunican mediante el cifrado simétrico deben estar de acuerdo en la clave a usar de antemano. Este documento basa la solución al problema, en algoritmos que se basen esta técnica de cifrado.

A. Algoritmo Genético

En esta sección veremos algunos términos que se utilizarán a lo largo de la investigación, el cual se tratará sobre los algoritmos genéticos y sus características, como operan y cuáles serán los objetivos para la investigación.

Según J.L. Galindo, y A. Carnicero [7], los algoritmos genéticos fueron propuestos por Rechenberg con el fin de resolver problemas de búsqueda y optimización, basado en la leyes de evolución de las especies por Darwin en en su libro "*El origen de las especies"* publicado en 1859. Dichas leyes, mencionan que las especies evolucionan cruzando sus cromosomas con el fin de reproducir y mutar para crear individuos mejor adaptados y que entre mayor adaptado se encuentre el nuevo individuo, mayor probabilidad tendrá de procrear la especie y que en caso contrario, simplemente muere. Ahora bien, en el capítulo del temario de"*temagenetico"* de Intelligent Systems Group [4], los algoritmos genéticos son capaces de

crear soluciones para problemas del mundo real, que dicho sea de paso, las soluciones llegan a ser valores óptimos del problema original, dicho en otras palabras [7], los algoritmos genéticos, se basan en dos principios, en la *exploración* y en la *explotación*. La exploración consiste en investigar nuevos y mejores individuos en el espacio de soluciones de la función objetivo del problema a solucionar y la explotación se usa para aprovechar el conocimiento que se tenía de los puntos anteriormente encontrados para encontrar puntos mejores.
En el principio de exploración [4] [5], a cada elemento del espectro, se le asigna un número real o una simple puntuación que representa la probabilidad de adaptación del individuo en el problema, esto radica en la probabilidad de ser un candidato para que sea seleccionado para reproducirse y cruzado con otro elemento seleccionado de la misma forma. El cruce de individuos producen nuevos individuos el cual comparte algunas características de los padres. Así de esta manera, se generan nuevas poblaciones con material genético mejorado produciendo mejores soluciones.

B. Algoritmo Genético Canónico.

La solución de la investigación se basa en el algoritmo genético simple o canónico, el cual se dará un breve descripción de su composición y como opera.
Este algoritmo necesita una codificación del problema que resulte adecuada, además requiere una función de ajuste o adopción al problema, el cual asigna un valor a cada posible solución codificada.
Intelligent Systems Group [4], menciona que una vez que el algoritmo es ejecutado, los padres son seleccionados para producir generaciones, a estos se le aplica una función de cruce para generar dos hijos en el que se le aplicación la función de mutación que finalmente el nuevo conjunto solución resultante de la combinación de los individuos originalmente seleccionados, son posibles soluciones al problema y consecuentemente, la evolución del algoritmo genético formará una nueva población.

Aarti Soni en su publicación [5], menciona cuales son los tres principales componentes de un algoritmo genético canónico, el cual son: *Selección, Cruce y Mutación*, y que serán descritos en esta sección.

1) Codificación

Intelligent Systems Group [4], plantea que existe un primer método del algoritmo genético simple, este consiste en la codificación de los elementos de la población, el cual son representados como genes, y los valores de estos representa cromosomas, y para efectos del tratamiento del algoritmo, se debe codificar cada elemento del gen en binario, conteniendo un arreglo de cromosomas con valor de {0,1}.

2) Selección

Es criterio cuantitativo basado en el principio de exploración [7], demuestra que existe un conjunto de elementos llamados individuos, que comparten dentro de una misma población. Cada individuo está compuesto por un conjunto de cromosomas formado por genes que contiene el ADN del individuo y representa toda la información genética de la especie. Al ser un número finito de individuos, se utiliza un valor de número real o simplemente un valor puntual que se le asigna a un genoma para ser elegido como un genoma padre que irá a reproducir con otro individuo de la población, seleccionado de la misma forma.[5], como se mencionó anteriormente, entre más alto sea el valor del genoma, mayor probabilidad tendré en ser elegido para la reproducción[7].

3) Cruce

Según Galindo y Carnicero [7], durante la reproducción, los genes de los padres se cruzan, formando nuevos cromosoma, por lo tanto, los hijos contienen la información genética de los padres.
En esta operación se tienen dos cromosomas ya seleccionados anteriormente, estos son dos individuos de la población original. El cruce se genera mediante la adopción de algunos atributos del primer cromosoma y el resto del segundo cromosoma [4] [5].
Veamos el ejemplo: tenemos los valores de 11110101 y de 01110011, estos podrían ser cruzados después de tercer bit, el cual es elegido aleatoriamente, en consecuencia se se observan los siguientes resultados:

Padre_1.

1	1	1	1	0	1	0	1

Padre_2.

0	1	1	1	0	0	1	1

Punto de cruce del Padre_1

1	1	1		1	0	1	0	1

Punto de cruce del Padre_2

0	1	1		1	0	0	1	1

Nueva Población
Hijo_1

1	1	1	1	0	0	1	1

Hijo_2

0	1	1	1	0	1	0	1

Figura1. Ejemplo de Operador de Cruce en un punto

4) Mutación

Una vez cruzado, existe la probabilidad de ser mutado, depende de las necesidades del problema. La mutación puede ser beneficiosa porque contribuye en la diversidad genética de la especie, además provee de las funciones objetivo o problemas iniciales de no limitar a solo un conjunto de soluciones, si no más allá de esto, por la combinatoria y generación de nuevas especies. Por lo tanto, la mutación consiste en modificar ciertos genes de forma aleatoria (J. Arranz de la Peña y A. Parra Truyl) [6].

Por ejemplo, para el caso de una mutación binaria, este consiste en la inversión de un único gen que corresponderá con un bit, así como se observa en la figura 2. [5] [6].

Mutación en la segunda posición

0	0	0	0	0	0	1	0

Cadena mutada

0	1	0	0	0	1	0	0

Figura 2: Ejemplo de mutación

C. Algoritmos de Cifrado.

Es esta sección, se brinda una síntesis sobre los algoritmos de cifrado, en específico detalle sobre los algoritmos de cifrado simétrico, el cual se basa este experimento.

En la tesis de posgrado de Redes y Seguridad, el Lic. Adrián Pousa [8], describe detalladamente los conceptos importante de la criptografía, y como define criptografía como *"... es el arte o ciencia de cifrar (encriptar) y descifrar (desencriptar) información utilizando técnicas que hagan posible el intercambio de mensajes de manera segura que solo puedan ser leídos por las personas a quienes van dirigidos",* dentro de dicha definición también agrega que existen dos tipos de cifrados, estos son los, sistema de cifrado *simétricos* y sistemas de cifrado *asimétricos*.

1) Sistemas de cifrado simétrico

Los sistemas de cifrado simétricos, con una técnica de criptografía para cifrar datos o información, esta se caracteriza porque utilizan la misma clave para cifrar y descifrar, y a esta clave se le conoce como llave pública. Al mismo tiempo, existen dos modos de operación básicos, en el que se observan a continuación:

i) Cifrado en bloques: es cuando la información que se van a procesar se dividen en tamaños de palabra de 8, 16, 32, 64, 128 y 256 bytes. Cuando ya se poseen los bloques listos, el algoritmos de cifrado cifra cada bloque utilizando la clave secreta, como ya mencionados anteriormente, ejemplos de esta arquitectura son los algoritmos de AES, DES.

ii) Cifrado de flujo: este modo de operación consiste cuando la información fluye en pequeños fragmentos a tiempo real, los tamaños puede ir desde 1 byte hasta 8 bytes de tamaño de palabra por fragmento enviado dentro del flujo de información. Para cifrar el contenido, se genera una secuencia pseudo-aleatoria de bits, esta nueva cadena de valor, se combina con el valor de los fragmentos, cifrando el contenido, un ejemplo de esto es el algoritmo de RC4 aplicado en el fluido de datos telefónicos.

Para efectos del experimento, nos basaremos en la solución de los algoritmos simétricos.

2) Sistemas de cifrado asimétricos

Según esta misma tesis [8], los sistemas asimétricos, son otro tipo de algoritmo de la familia criptográfica, al igual que los algoritmos de cifrado simétricos utiliza una clave pública que se encuentra disponible al público, con la diferencia que utiliza además una clave privada y es conocida por un solo individuo, algunos ejemplos de estos algoritmos son los RSA y DSA por mencionar alguno. Existen dos métodos para cifrar la información, el método de *encriptación* y el de *autenticación*, y ambos serán expuestos a continuación.

i) *Encriptación:* consiste en enviar un mensaje cifrado con la clave pública del receptor a un destinatario y el mensaje una vez receptado, se descifra con la clave privada del mismo receptor.

ii) *Autenticación*: consiste en enviar un mensaje cifrado con clave privada del emisor a uno o varios destinatarios, una vez que el destinatario obtiene el mensaje, éste lo descifra con la clave

pública del emisor, garantizado autenticación e integridad de la información.

II. PLANTEAMIENTO DEL PROBLEMA

Cuando se cifra un documento utilizando algoritmos de clave simétrica, toda la seguridad de la información recae en la clave utilizada, es decir, de nada le sirve a un atacante conocer el algoritmo que se está utilizando, sólo si conoce la clave, necesita saber cuál es el algoritmo.

Según [3], dado que toda la seguridad está en la clave, es importante que sea muy difícil adivinar el tipo de clave. Hoy por hoy, los ordenadores pueden adivinar claves con extrema rapidez, y ésta es la razón por la cual el tamaño de la clave es importante en los «criptosistemas» modernos. El algoritmo de cifrado DES usa una clave de 56 bits, lo que significa que hay 2^{56} claves posibles. 2^{56} son 72.057.594.037.927.936 claves. Esto representa un número muy alto de claves, pero una computadora de uso general puede comprobar todo el espacio posible de claves en cuestión de días. Una máquina especializada lo puede hacer en horas. Por otra parte, algoritmos de cifrado de diseño más reciente como 3DES, Blowfish e IDEA usan todos claves de 128 bits, lo que significa que existen 2^{128} claves posibles. Esto representa muchas, muchísimas claves más, y aun en el caso de que todas las máquinas del planeta estuvieran cooperando, todavía tardarían más tiempo que la misma edad del universo en encontrar la clave.

Otro punto que hay que considerar [9], es la vulnerabilidad siempre latente de las técnicas de ingeniería social, esta una técnica usada para obtener información personal o confidencial mediante engaños; para el simple hecho, penetrar, robar, vulnerar, atacar, alguno de los tres elementos del triángulo de la seguridad, Disponibilidad, Integridad y Confidencialidad, y en este caso, para lograr adquirir claves o llaves , por medio de estas técnicas, que también puede ser evitadas llevando medidas preventivas y de seguridad dentro de la organización o como parte de una buena cultura como individuo.

III. SOLUCIÓN PLANTEADA

Para resolver el problema mencionado, una solución intuitiva, es hacer que la clave utilizada sea más difícil de encontrar. Ya que como se mencionó, por fuerza bruta son difíciles o imposibles de hallar. En algunos casos, se utilizan diccionarios, o si el atacante conoce a la víctima podría intuir la contraseña, inclusive, utilizar ingeniería social para averiguar.

La idea de este documento es mostrar una forma de modificar la contraseña que el usuario ingresa, mediante la utilización de algoritmos genéticos. Una vez mejorada la contraseña, esta es la que se utiliza en el algoritmo de cifrado. Una contraseña modificada, hará que la posibilidad de encontrarla por fuerza bruta, diccionarios o ingeniería social, la tendencia caiga a cero.

Un punto clave, es que el uso de este método de protección de contraseña, resulta útil únicamente para quien tiene el algoritmo que modifica la clave ingresada, dicho de otra manera, si una persona cifra todos sus archivos con la palabra “Montana123”, los documentos que sean pasados a través del algoritmo genético, serán cifrados con otra palabra distinta (“èÊDaËîÎmíï”), es decir, si alguien quisiera leerlos e intenta descifrarlos, utilizando como clave “Montana123”, sin pasarlo a través de la evolución del algoritmo genético, no va a funcionar. Esto quiere decir, que el algoritmo genético que mejora la contraseña escogida por el usuario, toma un rol importante, como el de una llave privada, el cual siempre va ser portada por el usuario.

A. Algoritmo para modificar la contraseña

A continuación se muestra el diagrama de flujo del algoritmo implementado.

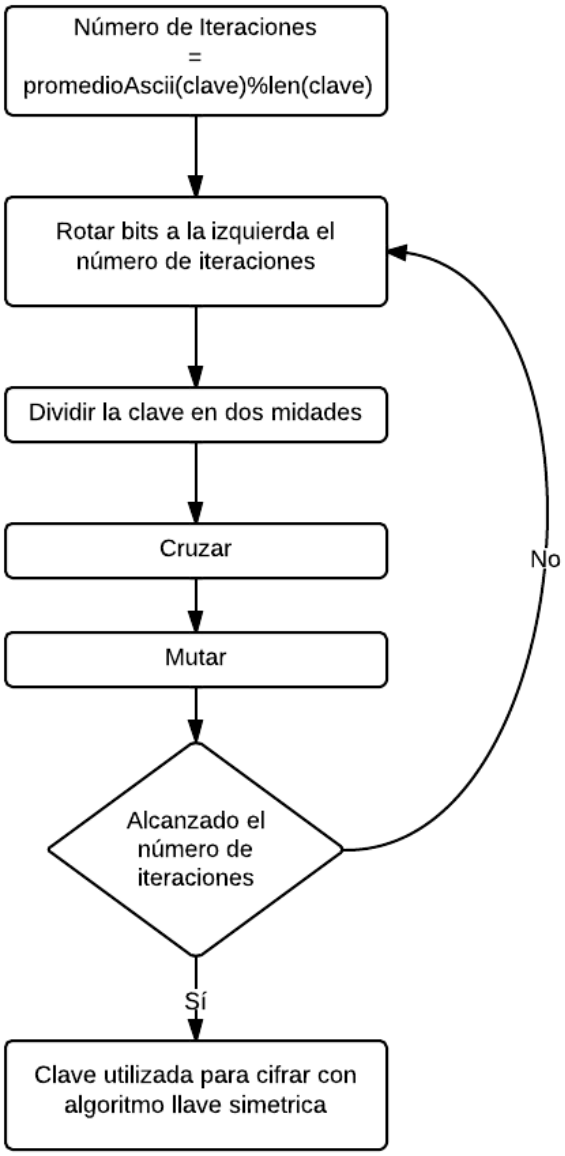

Figura 3: Flujo del código de software desarrollado

El cruce se realiza entre las mitades, haciendo un **or** y un **xor** de las mismas. La mutación depende de la paridad del número de iteración actual, se rearma la contraseña como **xorResult** + **orResult,** o como **orResult** + **xorResult**.

IV.CONCLUSIONES

Existe una ventana de posibilidades para aplicar algoritmos genéticos basados en parámetros de seguridad informática. Algunas aplicaciones que se investigaron sobre la genética y seguridad, se encontraron los siguientes casos: el uso de algoritmos genéticos para mejorar patrones de reconocimiento de rostros asociados al reconocimiento facial, también, algunas pruebas de seguridad para análisis de rastros con algoritmos genéticos y diseño e implementación de redes de seguridad mediante el uso de AG, por mencionar algunas.

Los algoritmos de cifrado simétricos, tiene la flexibilidad de usar diferentes tamaños de palabra, desde 8 bytes hasta los 256. Para efectos de asegurar el cifrado, se recomienda usar el uso de 256 bytes de tamaño, ya que la cantidad de opciones o candidatas de llaves aumenta exponencialmente, esto provoca, que ataques de fuerza bruta, no sean efectivos o al menos tengan muy bajas probabilidades de descifrar los mensajes.

Los algoritmos genéticos, son algoritmos especializados para la toma de decisiones, ya que cumplen características de optimización de operaciones, por ejemplo el uso de funciones objetivos. Se recomienda analizar cuidadosamente el problema, para identificar adecuadamente todas las variables y poner generar un algoritmo que optimice los valores de la función objetivo.

Para el caso de ingeniería social, es una estrategia muy utilizada para los diferentes niveles de ataques además de un sinnúmero de propósitos, para el caso de esta mejora, se disminuye la eficacia de la estrategia social, ya que al portar un llave privada, en otras palabras, al ser solo descifrado con el algoritmo genético y que solo va ser portada por el propietario, no será tan sencillo en adquirir esta llave, cabe de rescatar el la clave que el usuario usa, ya no será la misma con la que se cifra. Además se recomienda, mejorar los principios de seguridad a niveles individuales, en aspectos de cultura organizacional, con el fin de disminuir y mitigar errores cometidos con los atacantes.

V.BIBLIOGRAFÍA

[1] **Gnu Privacy Guard, gnupg,** *Anónimo,* última consulta: 06/06/2014. En línea: *https:*//www.gnupg.org/*gph/*es/manual/c190.html

[2] Support, Microsoft, ***Descripción de cifrado simétrico y asimétrico,*** *artículo: 246071 - Última revisión: domingo, 10 de marzo de 2013 - Versión: 3.3.* En línea: http://support.microsoft.com/kb/246071/es

I. **[3]*Alcántara, J. Cifrado asimétrico, wiki.versvs,*** **En línea: http://wiki.versvs.net/Cifrado_asim%C3%A9trico**

[4] *Intelligent Systems Group*, ***ALGORITMOS GENÉTICOS,*** *Recurso compartido desde (09 Nov 2004)*, En línea: *h*ttp://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/temageneticos.pdf

[5] A. Soni., S. Agrawal, **Using Genetic Algorithm for Symmetric key Generation in Image Encryption,** *International Journal of Advanced Research in Computer Engineering & Technology* Diciembre 2012

[6] *Arranz de la Peña, J.; Parra Truyol, A.* **ALGORITMOS GENÉTICOS**, *Universidad Carlos III de Madrid, Dep. Ingeniería Telemática.* En línea: http://www.it.uc3m.es/jvillena/irc/practicas/06-07/05.pdf

[7] *Galindo, J.L., Carnicero, A.* **Optimización mediante algoritmos genéticos: Aplicación al diseño de celosías**, *Anales de Mecánica y Electricidad*. LXXX (V), 40-50, (Sep/2003), En línea: https://www.iit.upcomillas.es/publicaciones/mostrar_publicacion_revista.php.en?id=94

[8] *Pousa, A.*, **ALGORITMO DE CIFRADO SIMÉTRICO AES. ACELERACIÓN DE TIEMPO DE CÓMPUTO SOBRE ARQUITECTURAS MULTICORE**, *Facultad de Informática - Universidad Nacional de La Plata, Diciembre 2011*, En línea: http://postgrado.info.unlp.edu.ar/Carreras/Especializaciones/Redes_y_Seguridad/Trabajos_Finales/Pousa_Adrian.pdf

[9] *Piñato Durán, L.* **INGENIERÍA SOCIAL: Un ataque a la confianza y al servicio en el sector financiero**, *Apuntes de Investigación Vol 3, Septiembre 2012*, En línea: http://apuntesdeinvestigacion.upbbga.edu.co/wp-content/uploads/ESI-Luis-Eduardo-Pati%C3%B1o-Dur%C3%A1n.pdf