**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

UNIDAD ACADÉMICA PROFESIONAL TIANGUISTENCO

INGENIERÍA EN SOFTWARE

**“Detección de phishing”**

Q U E P R E S E N T A

**PABLO CESAR SELVA CONTRERAS**

DOCENTE: Rene Arnulfo García Hernández

TIANGUISTENCO, MÉX. DICIEMBRE 2023

Contenido

[Planteamiento del Problema 4](#_Toc152713978)

[Función de aptitud 4](#_Toc152713979)

[Codificación del individuo: 5](#_Toc152713980)

[Operadores Genéticos utilizados. 6](#_Toc152713981)

[Pruebas a mano: 6](#_Toc152713982)

[Experimentación 8](#_Toc152713983)

[Explicación del mejor individuo 14](#_Toc152713984)

[Comparación con las pruebas a mano 14](#_Toc152713985)

[Conclusiones 15](#_Toc152713986)

Introducción:

En la era digital actual, la seguridad de la información se ha convertido en un componente esencial para salvaguardar la integridad y la confidencialidad de datos sensibles. Sin embargo, la sofisticación de las amenazas en línea, en particular el phishing, ha desafiado constantemente los métodos convencionales de detección. El phishing, una técnica maliciosa que busca engañar a los usuarios para revelar información confidencial, como contraseñas o información financiera, representa una amenaza persistente y en constante evolución.

Este trabajo se centra en abordar el desafío de la detección de phishing mediante la aplicación de algoritmos genéticos. Los algoritmos genéticos, inspirados en los procesos evolutivos naturales, ofrecen una perspectiva única para abordar problemas complejos y dinámicos como la identificación de sitios web fraudulentos. Al aprovechar la capacidad de los algoritmos genéticos para realizar búsquedas eficientes en espacios de soluciones amplios y no lineales, se busca mejorar la precisión y la eficacia en la identificación de intentos de phishing.

La investigación se estructura en torno a la comprensión profunda de los mecanismos de phishing, la revisión crítica de los enfoques existentes en detección de phishing y la implementación práctica de algoritmos genéticos como una solución efectiva. Con este estudio, se busca contribuir al campo de la ciberseguridad al ofrecer una herramienta robusta y adaptable para la identificación temprana de amenazas de phishing, fortaleciendo así la resiliencia de los sistemas de seguridad informática en un entorno digital dinámico y desafiante.

Planteamiento del Problema:

Los ataques cibernéticos, representa un desafío crítico para la seguridad de la información en la actualidad. El phishing, caracterizado por su capacidad para engañar a los usuarios y obtener información confidencial, ha evolucionado de manera significativa, adaptándose continuamente a las medidas de seguridad tradicionales. La detección efectiva de estos ataques se ha vuelto crucial, ya que las consecuencias de una falla en este sentido pueden resultar en la pérdida de datos sensibles, daños financieros y compromisos de la privacidad.

Los métodos convencionales de detección de phishing a menudo se basan en patrones estáticos y firmas previamente conocidas, lo que los hace vulnerables a las tácticas cada vez más sofisticadas empleadas por los perpetradores. Además, la dinámica naturaleza del phishing, donde nuevos métodos y estrategias surgen constantemente, destaca la necesidad de enfoques de detección más adaptables y proactivos.

En este contexto, surge la oportunidad de explorar y aplicar algoritmos genéticos como una alternativa innovadora para abordar los desafíos inherentes a la detección de phishing. Estos algoritmos, inspirados en los procesos evolutivos naturales, tienen el potencial de proporcionar soluciones más dinámicas y eficaces al problema de la detección de amenazas en línea. Sin embargo, la implementación específica de algoritmos genéticos en el contexto de la detección de phishing aún no ha sido exhaustivamente explorada, lo que constituye una brecha en la investigación actual.

# Función de aptitud

En detección de phishing el individuo, se específica con tres características (longitud del URL, presencia de "HTTPS" y coincidencia con palabras clave).

​

La función de aptitud se utiliza como criterio para seleccionar qué individuos sobreviven y se reproducen en cada generación del algoritmo genético, ya que los individuos más aptos tienen más probabilidades de pasar sus características beneficiosas a la siguiente generación.

# Codificación del individuo:

|  |  |  |
| --- | --- | --- |
| 1 | 2 | 3 |

1. Longitud del URL. URL extremadamente largos

|  |  |  |
| --- | --- | --- |
| ID | URL | Longitud |
| 1 | http://signin-att-100454.weeblysite.com/ | 42 |
| 2 | https://att-login7762.weeblysite.com/ | 35 |
| 3 | https://mail.google.com/ | 23 |

1. Presencia de HTTPS

|  |  |
| --- | --- |
| ID | URL |
| 1 | http://signin-att-100454.weeblysite.com/ |
| 2 | https://att-login7762.weeblysite.com/ |
| 3 | https://mail.google.com/ |

1. Coincidencia de palabras clave, acortadores, faltas de ortografía

|  |  |
| --- | --- |
| ID | Palabras clave |
| 1 | Loginn |
| 2 | Loging |
| 3 | htttps |
| 4 | Googlle |
| 5 | Securee |
| 6 | www2 |
| 7 | Facebookk |
| 8 | Registeer |
| 9 | Auth |
| 10 | uio.fly |

# Operadores Genéticos utilizados.

**Selección:**

La selección determina qué individuos serán elegidos para reproducirse y formar la próxima generación. Los individuos se seleccionan de acuerdo con su aptitud: aquellos con una mayor aptitud tienen más probabilidades de ser seleccionados.

Selección por Torneo: Este método selecciona varios individuos al azar y elige al mejor de ellos. Es simple y eficaz.

**Cruzamiento:**

Cruzamiento de un punto: Divide los cromosomas de los padres en un solo punto de cruce y combina las partes antes y después de ese punto para formar a los descendientes.

**Mutación:**

Mutación bit a bit: Cambia aleatoriamente algunos bits en el cromosoma. Ajusta la tasa de mutación para equilibrar la exploración del espacio de búsqueda y la explotación de soluciones.

# Pruebas a mano:

Se realizo un análisis de varios URL´S extraídos de un data set:

|  |  |  |  |
| --- | --- | --- | --- |
| ID | URL | Longitud | Observaciones |
| 1 | https://signin-att-100454.weeblysite.com/ | 41 | phishing |
| 2 | https://att-100404.weeblysite.com/ | 34 | phishing |
| 3 | https://mail-100080.weeblysite.com/ | 35 | phishing |
| 4 | https://att-login7762.weeblysite.com/ | 37 | phishing |
| 5 | https://att-104875.weeblysite.com/ | 34 | phishing |
| 6 | https://concisebookkeepingservices.com.au/ | 42 | Seguro |
| 7 | https://webmail-102865.weeblysite.com/ | 38 | Seguro |
| 8 | https://att-login4152.weeblysite.com/ | 37 | phishing |
| 9 | https://sign-in-att-103732.weeblysite.com/ | 42 | phishing |
| 10 | https://home4serve-105926.square.site/ | 38 | Seguro |
| 11 | https://ssgzx.com/ | 18 | Seguro |
| 12 | https://firstpag.com.br/wp-includes/certificates/service1/pp-nov/propular/popular/07977cb6f552bd6ccab6cb563008ab20 | 114 | phishing |
| 13 | https://firstpag.com.br/wp-includes/certificates/service1/pp-nov/propular/popular/07977cb6f552bd6ccab6cb563008ab20/ | 37 | phishing |
| 14 | https://att-102985.weeblysite.com/ | 115 | phishing |
| 15 | http://sp-tanupdate.net/ | 34 | Seguro |

Prueba 2:

|  |  |  |  |
| --- | --- | --- | --- |
| ID | URL | Longitud | Observaciones |
| 1 | https://cloudflare-ipfs.com/ipfs/bafybeiaizekonc5stugisvgirw3ac6slc635bpxw7jqfjieytde7ru6s2i/qop80.html | 103 | Phishing |
| 2 | http://www.news2023xlqptzx.matzid.com/ | 38 | Seguro |
| 3 | https://news2023ewxggu6.matzid.com/ | 35 | Seguro |
| 4 | https://cloouds-bar-84ac.slrheeibtuebsid.workers.dev/&data=05%7C01 | 66 | Phishing |
| 5 | https://negociebra.com.br/ | 26 | Seguro |
| 6 | http://steam-endpoint-store-cf-dev.steamdata.ai/ | 48 | Seguro |
| 7 | http://s-facebook-3-v2.vercel.app/ | 34 | Phishing |
| 8 | https://metamaskwalletsss-1.gitbook.io/ | 39 | Seguro |
| 9 | https://verifyme-commbank.com/intl/secure.php/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/ | 244 | Phishing |
| 10 | https://verifyme-commbank.com/intl/secure.php/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/ | 266 | Phishing |
| 11 | https://verifyme-commbank.com/intl/secure.php/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/ | 288 | Phishing |
| 12 | https://verifyme-commbank.com/intl/secure.php/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/ | 310 | Phishing |
| 13 | https://verifyme-commbank.com/intl/secure.php/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/verifyme-commbank.com/ | 156 | Phishing |

Experimentación  
  
**tam\_poblacion=100**

**prob\_cruce=0.3**

**prob\_mutacion=0.9**

**num\_generaciones=20**

**Tiempo de ejecución 0.5s**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

**tam\_poblacion=300,**

**prob\_cruce=0.3,**

**prob\_mutacion=0.9,**

**num\_generaciones=220)**

**Tiempo de ejecución 0.8s**

**Interfaz de usuario gráfica

Descripción generada automáticamente**

**Tam\_poblacion=500,**

**prob\_cruce=0.6,**

**prob\_mutacion=0.5,**

**num\_generaciones=220**

**Tiempo de ejecución 10s**

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

**tam\_poblacion=900,**

**prob\_cruce=0.6,**

**prob\_mutacion=0.5,**

**num\_generaciones=220**

**Interfaz de usuario gráfica

Descripción generada automáticamente**

**tam\_poblacion=1400,**

**prob\_cruce=0.2,**

**prob\_mutacion=0.9,**

**num\_generaciones=220**

**tiempo de ejecución 2 minutos**

**Interfaz de usuario gráfica

Descripción generada automáticamente**

**tam\_poblacion=2000,**

**prob\_cruce=0.7,**

**prob\_mutacion=0.9,**

**num\_generaciones=220**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

# Explicación del mejor individuo

Después de cada generación, se evalúa la aptitud de cada individuo en la población utilizando la función de aptitud definida para el problema específico. La función de aptitud asigna un valor numérico a cada individuo, indicando qué tan "bueno" es en términos de la tarea que está resolviendo el algoritmo genético.

La selección del mejor individuo se basa en los valores de aptitud calculados. Se busca el individuo cuya aptitud sea la más alta. Este proceso se realiza comparando los valores de aptitud de todos los individuos en la población y seleccionando aquel con el valor más alto, Una vez identificado el mejor individuo, se puede examinar su representación específica en términos de la codificación utilizada.

La interpretación del mejor individuo depende de la aplicación específica. En el contexto de la detección de phishing, las características del mejor individuo (longitud del URL, presencia de "HTTPS", coincidencia con palabras clave) pueden analizarse para comprender qué aspectos específicos han contribuido a su alto rendimiento según la función de aptitud.

# Comparación con las pruebas a mano

Conforme se fueron realizando pruebas con el algoritmo los resultados de las posibles detecciones de phishing fueron mejores que, con las pruebas a mano, ya que en el url se detectaban falsos positivos algunos urls contaban con la coincidencia de la palabra https el cual complicaba la detección correcta de phishing, en cambio con el algoritmo se evaluaban la cantidad de caracteres y algunas otras palabras las cuales hacían que se detectara de manera correcta el phishing en el url.

Texto

Descripción generada automáticamente

# Conclusiones

En conclusión, este estudio sobre la detección de phishing mediante un algoritmo genético ha arrojado resultados significativos, Las características seleccionadas, como la longitud del URL, la presencia de "HTTPS" y la coincidencia con palabras clave, jugaron un papel crucial en el éxito del algoritmo genético. Sin embargo, se reconocen algunas limitaciones en este estudio, como la necesidad de abordar escenarios más complejos y la posibilidad de mejorar la interpretación de las características identificadas. Este trabajo destaca la importancia de la aplicación de algoritmos genéticos en la ciberseguridad y ofrece una perspectiva prometedora para la detección de amenazas en línea. Las lecciones aprendidas durante la investigación proporcionan valiosas ideas para mejorar enfoques futuros.