Esteganografía lingüística utilizando Aprendizaje Profundo.

Trabajo Terminal No. 2019-B022

*Alumno: Hernández Ruiz Augusto Ignacio**
Directores: Cortez Duarte Nidia Asunción, Alexander Gelbukh
E-mail: escom@akil.com.mx

Resumen – La confidencialidad es uno de los servicios indispensables en las comunicaciones actuales, sobre todo cuando viajan por canales inseguros. Cuando algún atacante intercepta algún mensaje cifrado, tiene motivos para sospechar que contiene datos relevantes y trataría de descifrarlos, sin embargo si se obtiene información legible que no sea de su interés, no la atacaría.

El presente trabajo terminal desarrollará un sistema que permita realizar esteganografía lingüística utilizando aprendizaje profundo, con con el propósito de dar un enfoque diferente dicha técnica y los resultados que produce.

Palabras clave – Esteganografía, Aprendizaje Profundo.

1. Introducción

La necesidad de ocultar información confidencial, ha desempeñado un papel importante dentro de la sociedad desde las primeras civilizaciones, esta práctica se puede remontar a la época de los griegos, cuando los nobles tenían que transmitir un mensaje lo tatuaban la cabeza de algún esclavo y posteriormente esperaban a que el cabello creciera de nuevo para poder enviarlo al destinatario sin levantar sospechas sobre la existencia del mismo [1].

En la segunda guerra mundial, los Alemanes desarrollaron una técnica llamada "Microdot" que reducía fotografías hasta el tamaño de un punto tipográfico, para después agregarlas en algún texto que se transmitía por canales inseguros, y al viajar como textos en claro las hacía extremadamente difíciles de detectar [2].

La palabra esteganografía se deriva de los griegos *"stegos"* que significa ocultar, y *"grafia"* que significa escribir, por lo que "Escritura Oculta" puede considerarse como su definición original.

La esteganografía es el arte de ocultar información sensible dentro de algún contenedor que puede ser audio, video, texto, imágenes siendo estas últimas las más comunes debido a su frecuencia en Internet. Actualmente existen diversos algoritmos utilizados para realizar esteganografía, los cuales al ser combinados con algoritmos de cifrado, incrementan el nivel de confidencialidad aplicada a los datos [1].

Con el desarrollo de las nuevas tecnologías los campos de estudio de áreas que antes parecían completamente ajenas, han comenzado a reducir la distancia entre ellos, derivando en investigaciones y aplicaciones que involucran temas muy diferentes entre sí. Dos de las áreas que han coincidido son la Criptografía y el Aprendizaje Profundo, que fungen como base de este proyecto.

El aprendizaje profundo es una rama del aprendizaje máquina (Machine Learning), que está enfocada en el estudio y desarrollo de algoritmos inspirados por la función y estructura de las células del cerebro llamadas neuronas, intentando replicarlas en Redes Neuronales Artificiales (RNA por sus siglas en inglés) [3].

Las RNA son una simple abstracción de las neuronas biológicas implementadas como parte de un programa, su utilidad radica en la capacidad que tienen de ser entrenadas para realizar funciones de utilidad. Las llamadas redes neuronales multicapa reciben su nombre debido a que utilizan más de una capa de neuronas para procesar la información; se dividen en tres niveles importantes: entrada, intermedio (encargado del procesamiento) y salida [4].

2. Objetivo

- Objetivo General

  Diseñar e implementar un sistema que permita ocultar texto sensible dentro de un texto contenedor por medio de esteganografía lingüística generada con aprendizaje profundo y posteriormente, recuperar los datos ocultos con el fin de proveer confidencialidad.

- Objetivos Particulares

    1. Implementar un algoritmo de esteganografía lingüística.
    2. Generar un conjunto de entrenamiento para alimentar una Red Neuronal.
    3. Entrenar a una Red Neuronal con datos obtenidos del algoritmo de esteganografía lingüística.
    4. Desarrollar un algoritmo que permita recuperar los datos ocultos dentro del texto contenedor

3. Justificación

En la actualidad, internet es el medio por el que gran parte de las comunicaciones, transacciones y flujo de datos en general se realizan, debido a esto, la información confidencial como contraseñas, credenciales de acceso, llaves etc. viaja constantemente por canales inseguros. Para proteger estos datos, se han desarrollado técnicas y algoritmos para cifrarlos, de modo que aunque viajen por estos canales, se ofrezca el servicio de confidencialidad.

Sin embargo a la par de dichas técnicas de cifrado, existen diferentes tipos de ataques que partes externas a la comunicación original pueden implementar para sustraer los datos confidenciales y poder modificarlos o hacer un uso indebido de los mismos. Los atacantes al detectar que un flujo de datos viaja de forma cifrada u oculta, tienen motivos para sospechar que la información tiene relevancia para alguna de las partes involucradas en la comunicación y puede implementar un método para descifrarla.

Según un estudio hecho por McAfee Labs en 2016[5], el 11% de los ataques cibernéticos fueron realizados a SSL (Secure Socket Layer) interceptando datos antes de ser cifrados y dando acceso a información sensible como datos de tarjetas de crédito, números de seguridad social etc.

Se conoce como “Ataque del Hombre de En medio” (Man In The Middle Attack - MITM - por sus siglas en inglés) [6], a las acciones de interrumpir, interceptar, modificar y fabricar información que viaja por canales inseguros. A continuación se detalla el ataque de modificación.

1. Datos son enviados del punto A al punto B

2. El atacante implanta herramientas de escucha (listeners) en las transmisiones, interceptando datos que son específicamente etiquetados como valiosos.

3. Los datos interceptados pueden ser modificados en el proceso de transmisión, para intentar engañar al usuario final y hacer que proporcione información confidencial como credenciales de acceso etc.

4. Un vez que se obtuvo la información o que se almacenó para su posterior procesamiento, el flujo se reanuda con los datos originales hacía el destinatario.

El presente proyecto propone una manera diferente de proteger la información que se transmite por algún medio, haciendo uso de esteganografía lingüística, de modo que los datos al no ir cifrados pasen desapercibidos y puedan llegar sin problema a su destinatario. Además utilizando Aprendizaje Profundo para generar la técnica propuesta, se espera que el comportamiento del sistema se aproxime al de algoritmos existentes.

El proyecto incorpora conocimientos de diversas áreas con un enfoque primordial en el área de Criptografía y Redes Neuronales

4. Estado del Arte

La esteganografía, en la actualidad, debido a los avances tecnológicos y a la manera en la que se distribuye y se procesa la información, es aplicada generalmente en imágenes, en las cuales, al modificar el contenido de la misma a nivel binario, se puede incrustar información confidencial sin que la imagen resultante, denote una manipulación evidente sobre la misma.

Sin embargo, mi aproximación es por un medio de comunicación ampliamente usado, pero poco explorado para sus uso con esteganografía; los textos.

Algunas de las aproximaciones para la aplicación de esteganografía lingüística, se han dado por medio de un método llamado "Sustitución de sinónimos", la Universidad de Cambridge, realizó una publicación [7] en la que presentan su aproximación al

problema, haciendo uso de este método, y ayudados por medio del corpus n-gram de google, con el fin de garantizar que la sustitución de las palabras clave por sus sinónimos fueran imperceptibles para el usuario. Los resultados presentados mostraron un análisis tanto hecho por humanos, como un esteganálisis realizado por computadora para medir el nivel de seguridad que su algoritmo presentaba, diciendo que habían alcanzado un nivel de seguridad aceptable, dentro del rango de poder ocultar 2 bits de información por oración

En la Conferencia de Métodos Empíricos del procesamiento de lenguaje natural, se hizo la presentación de un trabajo [8] por parte de la Universidad de Edinburgh, en el cual, por medio de una red neuronal recurrente, generaban poesía china, utilizando selecciones de contenido, y analizando su comportamiento dentro del contexto en el que se encontraban, para poder después, generar texto de una longitud mayor a los trabajos previos, logrando generar poemas de una longitud determinada, con coherencia entre sus enunciados.

Por otra parte, existe una página de internet *"http://spammimic.com",* la cual te permite ocultar texto dentro de un texto que la misma página genera, teniendo una arquitectura de encoder-decoder, la cual te permite tanto ocultar, como obtener la información que introdujiste.[9]

5. Productos o Resultados esperados

El sistema contará con la arquitectura de un autoencoder (ver Figura. 1) el cual tiene un comportamiento descrito en la Figura 2.

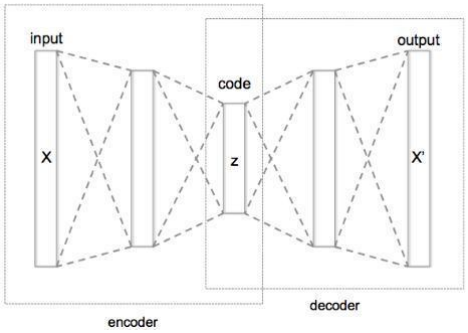

Figura 1. Arquitectura de un autoencoder

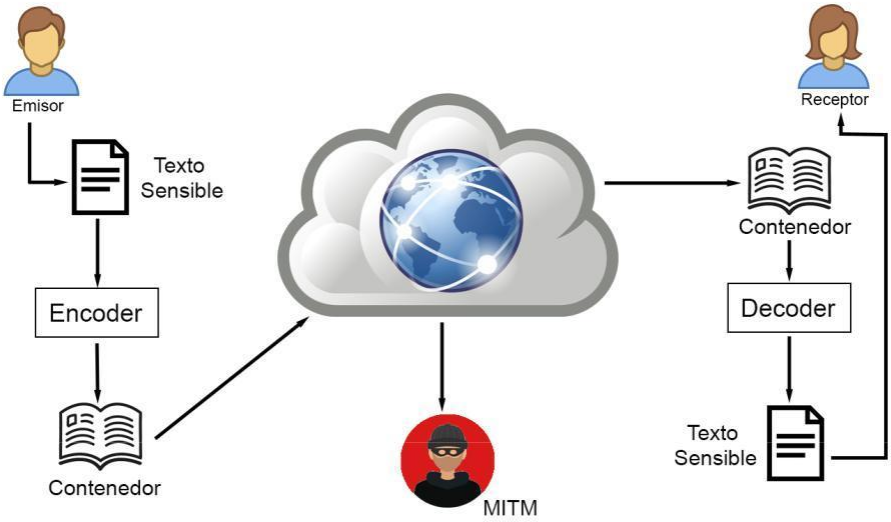

Figura 2. Diagrama del sistema.

Además de contar con un módulo independiente que será el encargado de generar el conjunto de entrenamiento con el que la RNA será entrenada para su correcto aprendizaje. El diagrama de dicho módulo de muestra a continuación. (ver Figura 3)

Figura 3. Diagrama de Módulo Generador de Conjunto de Entrenamiento.

Productos a entregar:

- Aplicación que realice esteganografía usando Aprendizaje Profundo.
- Documentación del sistema.
- Manual de usuario.

6. Metodología

La metodología que se utilizará en el desarrollo del sistema será Prototipos Evolutivos, ya que es parte del trabajo estudiar y seleccionar las herramientas y algoritmos que se usarán para la construcción del proyecto, este modelo es ideal para su construcción. Además al ir creando prototipos y haciendo iteraciones para refinarlos se genera una retroalimentación que nos permitirá enriquecer los requerimientos y acercar el desarrollo del prototipo final en un menor tiempo.

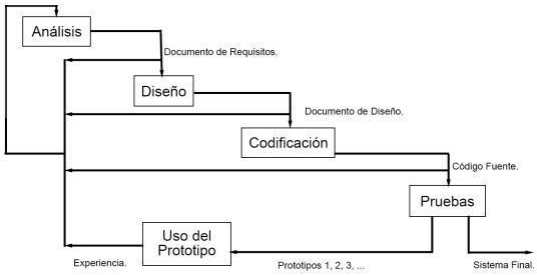

Figura 5. Diagrama de Metodología de Algoritmos Evolutivos.

7. Cronogramas

Nombre del alumno(a): Hernández Ruiz Augusto Ignacio TT No. : **2019-B022**
Título del TT: **Esteganografía Lingüística utilizando Aprendizaje Profundo.**

Actividad	FEB	MAR	ABR	MAY	JUN	AGO	SEP	OCT	NOV	DIC
Análisis y Especificación										
Construcción de Prototipo Inicial (Textos procesados para ser parte del conjunto de entrenamiento)										
Verificación de Requerimientos										
Modificación y actualización de Prototipo (Desarrollo de RNA, algoritmo de esteganografía, encoder, decoder)										
Elaboración de la Documentación										
Evaluación de TT I.										
Retroalimentación de Requerimientos										
Supervisión de pruebas										
Elaboración de presentación de resultados										
Evaluación de TT II.										

8. Referencias

[1] T Morkel, JHP Eloff and MS Olivier, "An Overview of Image Steganography," in Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005), Sandton, South Africa, June/July 2005 (Published electronically)

[2] Maninder Singh Rana, Bhupender Singh Sangwan, Jitendra Singh Jangir, “Art of Hiding:: An Introduction to Steganography”, International Journal of Engineering and Computer Science, Vol 1, Pag. 11-20, Octubre 2012.

[3] Brownlee Jason . (agosto16, 2016). What is Deep Learning?. octubre 8, 2017, de machinelearningmastery Sitio web: https://machinelearningmastery.com/what-is-deep-learning/

[4] Approximation by Superpositions of a Sigmoidal Function, Cybenko G, Springer Vertag, New York, EUA, 1989.

[5] Symantec. (2016). Internet Security Threat Report. Internet Security Threat Report, 21, Pag. 25, octubre 11, 2017.

[6] Symantec . (2016). What Is A Man In The Middle Attack?. octubre 11, 2017, de Norton Sitio web: https://us.norton.com/internetsecurity-wifi-what-is-a-man-in-the-middle-attack.html

[7] Practical Linguistic Steganography using Contextual Synonym Substitution and a Novel Vertex Coding Method, Ching-Yun Chang, Stephen Clark, University of Cambridge, UK, 2014.

[8] Chinese Poetry Generation with Recurrent Neural Networks, Xingxing Zhang, Mirella Lapata, University of Edinburgh, UK, 2014.

[9] Sitio web: http://spammimic.com/

9. Alumnos y Directores

Hernández Ruiz Augusto Ignacio.- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2012630497, Tel. 54466585, Email: escom@akil.com.mx

FUNDAMENTO LEGAL: Art. 3, fracc. II, Art. 18, fracc. II y Art. 21, lineamiento 32, fracc. XVII de la L.F.T.A.I.P.G. PARTES CONFIDENCIALES: No. de boleta y Teléfono.

Firma:_____

Cortez Duarte Nidia Asunción, Maestra en Ciencias en Computación CINVESTAV-IPN 2009, Ing. en Sistemas Computacionales ESCOM-IPN 2006, Profesora Titular en ESCOM Depto. de Ingeniería en Sistemas Computacionales. Áreas de interés: criptografía, seguridad de información, hardware reconfigurable, aritmética computacional, diseño digital y redes de computadoras. Teléfono: 57-29-6000 ext. 52032 email: nidiacortez3@gmail.com

Firma:_____

Alexander Gelbukh, M. en C. en matemáticas (con distinción) por la Universidad Estatal M.V. Lomonósov de Moscú, Rusia, 1990. Doctor en ciencias (ciencia de la computación) por el Instituto de la Información Científica y Técnica de Toda Rusia, 1995. Profesor-investigador y Jefe del Laboratorio de Lenguaje Natural y Procesamiento de Texto del Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN), México, desde 1997. Profesor invitado de la Universidad Nacional de Colombia. Investigador Invitado de la Universidad Waseda, Japón. Miembro de la Academia Mexicana de Ciencias desde 2000. Investigador Nacional de México (SNI) desde 1998, actualmente con el nivel 3. Galardonado con el Diploma a la investigación, por el IPN. Presidente de la Sociedad Mexicana de Inteligencia Artificial, expresidente de la Asociación Mexicana del Procesamiento del Lenguaje Natural.
gelbukh@cic.ipn.mx

Firma:_____