



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
DOS MIL DIECISEIS



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México



Instituto Politécnico Nacional

ESCUELA SUPERIOR DE INGENIERÍA MECÁNICA Y ELÉCTRICA
UNIDAD CULHUACAN

SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

“*Esteganografía en Documentos Electrónicos con Formato PDF*”

A U T O R:
Mtro. GUILLERMO DANIEL RIVERA DIAZ
e-mail: gdrd_23@yahoo.com

ASESORES:
Dr. Mario González Lee
Dr. Gabriel Sánchez Pérez





“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario
21 y 22 de abril de 2016, Cortazar, Guanajuato, México

Resumen

La seguridad de la Información a partir de la evolución, desarrollo tecnológico y masificación de los medios electrónicos y de comunicaciones, se ha convertido en parte trascendente e indispensable escolta en la prosperidad que se vive en el uso de las tecnologías de la información, ya que la información ha pasado de ser un conocimiento explícito extraído y resultado de la interacción con el entorno en además, la columna vertebral de la que dependen las organizaciones y por lo tanto uno de los activos más importantes de cualquier organización; la cual existe en muchas formas y medios y que además requiere de una protección adecuada. La seguridad de la información protege a la información de amenazas y riesgos, busca garantizar su confidencialidad, disponibilidad e integridad.

La protección y el resguardo de la información es algo de lo que el hombre se ha preocupado desde épocas inmemorables, el proteger el conocimiento debido a la ventaja que este representa para cualquier persona, ha motivado al hombre a la búsqueda de metodologías para la protección de la información; En la antigüedad se crearon las bibliotecas para resguardar y transmitir la información impidiendo que otros la obtuvieran. En la actualidad con el sobrevenir de los años, al extender el alcance de la tecnología, el cuidado de la información se ha vuelto crucial para los hombres, las organizaciones y las sociedades, por lo que el uso de disciplinas por demás antiguas como la Esteganografía o la Criptografía ha tomado gran importancia ya que el desarrollo tecnológico requiere la adaptación de dichas disciplinas a sus necesidades y tendencias.



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

La Esteganografía es una disciplina dedicada a proporcionar protección a la información ocultándola dentro de un objeto o entidad con el fin de que su existencia pase inadvertida.

En este trabajo de tesis se busca contribuir en el desarrollo de la tecnología a través de la implementación de un esquema Esteganográfico que permita ocultar información utilizado como portadora los documentos con formato PDF, con el objetivo de incrementar el catálogo de posibles repositorios para dicho fin.

Abstract

Safety Information from evolution, technological development and popularization of electronic media and communications, has become important and indispensable part escort prosperity being experienced in the use of information technology as information has evolved from an explicit knowledge extracted and result of interaction with the environment in addition, the backbone of the organizations dependent and therefore one of the most important assets of any organization, which exists in many forms and media and also requires adequate protection. The information security protects information from threats and risks, seeks to ensure its confidentiality, availability and integrity.

The protection and preservation of information is something that has concerned man since time immemorial, protecting knowledge because of the advantage that this represents for anyone, man has led to the search for methods to protect information, In ancient libraries were created to safeguard and transmit information by preventing others obtain. Today occur with the years, to extend the reach of technology, care information has become crucial for men, organizations



“CONGRESO INTERNACIONAL DE INVESTIGACION E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

and societies, so that the use of other ancient disciplines like Steganography or Cryptography has become very important as the technological development requires the adaptation of these disciplines to their needs and trends.

Steganography is a discipline dedicated to providing protection for hiding information within an object or entity so that its existence does not mark.

In this thesis seeks to contribute to the development of technology through the implementation of a steganographic scheme that allows hiding information used as carrier with PDF documents, with the aim of increasing the existing catalog of possible repositories for this purpose.

Introducción

El formato de documento portátil (PDF) es el formato de archivo nativo de la familia de productos Adobe® Acrobat®. El objetivo de estos productos es permitir a los usuarios intercambiar y ver documentos electrónicos de forma fácil e íntegra, independientemente del entorno en el que fueron creados. PDF se basa en un modelo de descripción de páginas, muy similar al del lenguaje PostScript®, para describir texto y gráficos en una forma independiente del dispositivo y de la resolución. Para mejorar el rendimiento y para una visualización interactiva, PDF define un formato más estructurado que el utilizado por la mayoría de los programas en lenguaje PostScript. PDF también incluye objetos, tales como anotaciones y enlaces de hipertexto, que no forman parte de la propia página, pero son útiles para una visualización interactiva y el intercambio de documentos.

PDF es un formato de archivo utilizado para representar documentos de forma independiente de la aplicación, del software, hardware y sistema operativo



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

utilizado para crearlos y del dispositivo de salida en el que se van a mostrar o imprimir. Un documento PDF consta de una colección de objetos que en conjunto describen la aparición de una o más páginas, posiblemente acompañado de otros elementos interactivos y de alto nivel de datos de aplicación. Un archivo PDF contiene los objetos que componen un documento PDF con la información estructural asociada, todos ellos representados como un solo auto-contenido en secuencia de bytes.

Las páginas de un documento y otros elementos visuales pueden contener cualquier combinación de texto, gráficos e imágenes. La apariencia de una página se describe mediante una secuencia de contenido PDF, que contiene una secuencia de objetos gráficos que se imprimen en la página. Este aspecto está totalmente especificado, todas las decisiones de diseño y formato ya se han definido por la aplicación que genera la secuencia de contenido.

Además de describir el aspecto estático de páginas, un documento PDF puede contener elementos interactivos que sólo son posibles en una representación electrónica. PDF es compatible con las anotaciones y muchos tipos de objetos, tales como: notas de texto, enlaces de hipertexto, el marcado, los archivos adjuntos, sonidos y películas. Un documento puede definir su propia interfaz de usuario, teclado y ratón pueden realizar acciones que son previamente especificadas por los objetos PDF. El documento puede contener campos de formulario interactivos a ser requisitados por el usuario, y puede exportar los valores de estos campos o importarlas desde otras aplicaciones.

Finalmente, un documento PDF puede contener información de más alto nivel, que sea útil para el intercambio de contenidos entre aplicaciones. Además



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
DOS MIL DIECISEIS



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

de especificar el aspecto, el contenido de un documento puede incluir la identificación y la lógica general de estructura de información que le permite realizar búsquedas, edición, o se extrae para su reutilización en otros lugares. PDF es especialmente adecuado para la representación de un documento mientras se mueve a través de etapas sucesivas de un flujo de trabajo de pre impresión.

Metodología.

Las capacidades para realizar ocultamiento de información con las que se cuenta dentro de la estructura de un documento PDF son muy amplias, en esta tesis se busca implementar un esquema de ocultamiento de información que permita insertarla de forma oculta, eficiente y masiva; se propone insertar información aprovechando los objetos texto y la capacidad de manipular sus métricas que caracterizan un glifo.

Hemos elegido implementar una combinación de técnicas; esto es, insertar información manipulando los parámetros de espaciado entre caracteres y palabras, aunque el espaciado entre palabras se puede considerar como un espaciado entre caracteres tomando en cuenta que entre cada palabra existe un carácter de espacio, por lo que se insertaría información entre el ultimo carácter de la palabra anterior y el espacio y, entre el espacio y el primer carácter de la siguiente palabra, que reflexionando se podría considerar como una sola técnica de ocultamiento de información. Así como se combinan estas dos técnicas es posible hacer una composición de diferentes técnicas e incrementar la capacidad de ocultamiento.



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario
21 y 22 de abril de 2016, Cortazar, Guanajuato, México

De acuerdo a lo expuesto, dentro de la estructura de un objeto texto se cuenta con el operador Tc (charSpace), el cual es un numero expresado en unidades del espacio del texto sin escala y que nos ayuda a establecer el espacio entre un carácter y el siguiente, su valor por defecto es cero.

Insertar espacios entre caracteres

Un inconveniente para la implementación que requerimos con el uso de este operador, es que Tc se define como un único valor por cada cadena de caracteres en cada objeto stream de texto, por lo que no sería útil debido a que la intención es manipular cada valor que define el espacio entre un carácter y el siguiente adicionando el valor de la información a ocultar, lo cual nos proporcionaría una posición para ocultar una unidad de información por cada carácter de la cadena de texto en el objeto stream de texto.

En la búsqueda de un operador de texto que nos permitiera manipular de forma conveniente para nuestra implementación encontramos al operador TJ, el cual es capaz de mostrar una o más cadenas de texto permitiendo el posicionamiento individual de cada glifo.

En este caso el operador TJ imprime los caracteres entre paréntesis, pero ajustando su posición a la cantidad fuera de los paréntesis. Esta cantidad se expresa en milésimas de unidad en el espacio de texto. La cantidad se subtrae desde la coordenada horizontal o vertical actual, dependiendo del modo de



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
 Multidisciplinario
 21 y 22 de abril de 2016, Cortazar, Guanajuato, México

escritura. En el sistema de coordenadas por defecto, si el ajuste es positivo indica que el siguiente glifo se va a desplazar hacia la izquierda.

El operador TJ muestra el caracter entre paréntesis en la posición en que se encuentre el cursor. A continuación se muestra un ejemplo del uso del esquema:

Ejemplo:

Texto a mostrar en la página del PDF = 'Hello World'

Texto a insertar de forma oculta dentro de la estructura del documento PDF =
 Estegano

El texto a ocultar se toma glifo por glifo (carácter por carácter) y se convierte a su valor decimal de acuerdo al código ASCII; esto es:

Carácter ASCII Decimal

E = 69, S = 115, t = 116, e = 101, g = 103, a = 97, n = 110, o = 111

El objeto texto de forma básica quedaría como sigue:

2 0 obj	Objeto 2 número de generación 0
<<	Diccionario de atributo
/Length 39	Llave de longitud 39 octetos
>>	
Stream	Principia el stream
BT	Op. de margen de pag. Begin Text
/F9 12 Tf	Fuente 9 tamaño 12
	Tf – define nombre de la fuente y tamaño
30 650 Td	Td—define ubicación del primer glifo



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

[(H)-69(e)-115(l)-116(l)-101(o)(W)-103(o)-97(r)-110(l) 111(d)]TJ

TJ—Operador para imprimir la cadena de caracteres entre paréntesis

ET Op. de margen de pag. End Text

Endstream Final del stream

Endobj Final del Objeto

Se observa del ejemplo anterior la forma como se manifiesta el offset que se está adicionado a la posición original de cada caracter por medio del operador TJ, el caso del carácter “d”, al que se le agregó un offset positivo provocando un desplazamiento notorio a la izquierda, a diferencia del desplazamiento de los demás caracteres que también fue notorio pero a la derecha, debido a que el offset que se le agregó fue negativo. En la ilustración se muestra dicho comportamiento.

HelloWorld

De acuerdo a lo descrito los glifos son definidos en el espacio del glifo, cuya definición varía según el tipo de fuente, las coordenadas del glifo primero son transformadas del espacio de glifo al espacio de texto; el texto se muestra en el espacio de texto, que es definido por la combinación de la matriz de texto, T_m , y los parámetros de estado de texto T_{fs} , T_h , y T_{rise} . Esto determina como las coordenadas de texto son transformadas en el espacio de usuario. Primero, un desplazamiento combinado es calculado, denotado por t_x en el modo de escritura horizontal por medio de la siguiente ecuación:

$$t_x = \left(\left(w_0 - \frac{T_j}{1000} \right) \times T_{fs} + T_c + T_w \right) \times T_h$$



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario
21 y 22 de abril de 2016, Cortazar, Guanajuato, México

Donde:

W_0 es el desplazamiento horizontal del glifo, tomando el caso más elemental $W_0=0$; esto es, supondremos que iniciamos con un desplazamiento igual a cero.

T_j es un ajuste de posición especificado por un número en una serie de T_j , en nuestro primer acercamiento podemos tomar el primer valor $T_j=-69$.

T_f y T_h son el tamaño de fuente de texto actual que para nuestra transformación del espacio de texto al espacio de usuario se traslada igual a uno.

T_c y T_w son del carácter actual, parámetros que espacian la palabra en el estado de los gráficos, su valor inicial es igual a cero.

De lo anterior tenemos que:

$$t_x = (-T_j/1000) = -(-69/1000) = 0.069 \text{ (sin unidades)}$$

En el sistema de coordenada de usuario por defecto, esto quiere decir que el glifo estándar es del tamaño de 1 unidad en el espacio de usuario o 1 / 72 de pulgada.

Por lo anterior tenemos que en el espacio de usuario tendremos un desplazamiento = $0.069 * (1/72)$ pulgadas = 0.000958 pulgadas = 0.0024 centímetros.

El desplazamiento provocado es visible para el ojo humano por lo que, será necesario llevar nuestros valores a magnitudes imperceptibles.

Ya que en la Estenografía su principal característica es que la información insertada no provoque en el documento portadora cambios visibles al ojo humano será necesario llevar los valores de offset (desplazamiento) a cantidades que así lo permitan.

Esto es, llevaremos los valores actuales a una milésima parte de su valor actual para asegurar que el desplazamiento que provoquen sea imperceptible.



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

Se tomará cada uno de los valores decimales de cada carácter del texto a ocultar y se dividirá entre 10000 para reducir el valor a centésimas de unidad quedando como sigue:

$E=0.0690$, $s=0.0115$, $t=0.0116$, $e=0.0101$, $g=0.0103$, $a=0.0970$, $n=0.0110$
 $o=0.0111$

Resultados

De lo que se obtuvo:

HelloWorld

Mensaje de salida con offset
(Con información oculta)

HelloWorld

Mensaje de salida sin offset
(Sin información oculta)

En el ejemplo anterior se muestra como al agregar un offset del orden de centésimas de la unidad no provoca cambios significativos o visibles en la apariencia de la portadora, lo cual es la base para que un esquema Esteganográfico sea eficiente.

Conclusiones

En el desarrollo de este trabajo nos encontramos con un universo de posibilidades para desarrollar un esquema de ocultamiento de información dentro de la estructura del documento PDF, las grandes capacidades con las que se cuenta internamente en la estructura del formato PDF y la posibilidad de combinarlas para incrementar la capacidad de ocultamiento. El esquema propuesto nos permitió cumplir con los objetivos establecidos pero nos deja con el deseo de compartir el conocimiento adquirido para que los esfuerzos realizados



“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”
Multidisciplinario
21 y 22 de abril de 2016, Cortazar, Guanajuato, México

hasta el momento se aprovechen y sean un motivador para que se siga trabajando sobre este tema y esta línea en la Esteganografía.

Bibliografía

- [1] R. A. Española, «Real Academia Española,» Marzo 2012. [En línea]. Available: <http://www.rae.es/rae.html>. [Último acceso: 23 Marzo 2012].
- [2] L. C. Baron, MARCAJE DE DOCUMENTOS PDF PARA LA PROTECCION CONTRA LA REDISTRIBUCION ILEGAL, Barcelona España: Universidad Polotecnica de Catalunya, 2009.
- [3] C. J. Abel, *Ocultamiento de Información en Documentos Electrónicos de Tipo Postscript*, México, D.F.: Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Culhuacan, 2010.
- [4] I. (. O. f. Standardization), «ISO (International Organization for Standardization),» News, 2008. [En línea]. Available: <http://www.iso.org/iso/news.htm?refid=Ref1141> . [Último acceso: 23 Abril 2012].
- [5] A. S. Incorporated., «PDF Reference,» de *Adobe, Portable Document Format, version 1.7*, 2006.



“CONGRESO INTERNACIONAL DE INVESTIGACION E INNOVACIÓN 2016”
Multidisciplinario
21 y 22 de abril de 2016, Cortazar, Guanajuato, México

- [6] B. Lowagie, Manning Publications, «iText in Action: Creating and Manipulating PDF.,» 2006.
- [7] F. S. F. I. GNU, «Death Master,» 2004. [En línea]. Available: <http://www.death-master.tk>. [Último acceso: 23 Noviembre 2011].
- [8] L. C. ©2012, «LinkedIn,» LinkedIn Corporation ©, 2012. [En línea]. Available: <http://be.linkedin.com/in/blowagie>. [Último acceso: 25 Octubre 2012].
- [9] I. O. f. Standardization, «ISO 32000-1:2008,» ISO, 17 12 2011. [En línea]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=51502. [Último acceso: 22 4 2012].