Metodología para usar la esteganografía como medio de acreditar la validez de la documentación publicada electrónicamente

Paz Menvielle, María Alejandra [1]; García Neder, Héctor Jorge[2]; Groppo, Mario Alberto[3]; Gibellini, Fabian[4]; Sánchez, Cecilia Beatriz[5]; Pozzi, José Wuiler[6]; Broda, Noel[7]; Gudiño, María Belén[8]

Universidad Tecnológica Nacional, Facultad Regional Córdoba, Departamento Sistemas de Información: ([1]) pazmenvielle@yahoo.com.ar; ([3]) WICC@groppo.com.ar; ([4]) speaker@bbs.frc.utn.edu.ar; ([5]) csanchezjuriol@hotmail.com; ([6]) wuiler@gmail.com; ([7]) noel_fb@hotmail.com; ([8]) mb_gud89@hotmail.com

Universidad Tecnológica Nacional, Facultad Regional Río Grande: ([2]) hgneder@frrg.utn.edu.ar

Resumen

Actualmente, diversos organismos proveen gran cantidad de documentación mediante medios computarizados. Se presenta información académica y/o administrativa, se presentan formularios, noticias, resoluciones del ámbito público, etc., y en muchos casos esos formularios son utilizados para generar posteriores trámites administrativos.

Sin embargo, es poco común que esa información, o documentos, sean validados para comprobar su autenticidad, integridad, etc. Las revisiones son no automatizadas y el elemento se presupone correcto solamente por las características que el receptor puede reconocer, ello implica riesgo. La información puede resultar corrupta, los documentos alterados y podría conducirse a disputas de índole administrativo/legal.

Es necesario que cualquier documento/información difundido por medios masivos, tenga la posibilidad de ser verificado en su originalidad e integridad. Con este objetivo, buscamos establecer una metodología, usando herramientas esteganográficas, que brinden una respuesta a esta necesidad.

Los desafíos son importantes y originales. Las herramientas existentes son ineficientes cuando el documento emitido se transfiere en diferentes soportes (impresión, fotocopia, etc.).

Buscamos una metodología integral, ya que la sola herramienta no es suficiente para abarcar las múltiples situaciones posibles.

Desarrollaremos el sistema de soporte informático necesario, y aunque se aplicarán técnicas esteganográficas, se pretende que sean solamente el soporte técnico de la metodología: nuestro principal objetivo.

Contexto

El proyecto se encuentra acreditado en la Facultad Regional Córdoba y en la Facultad Regional Río Grande de la Universidad Tecnológica Nacional (Argentina), siendo éstas las instituciones financiadoras del proyecto. Código: IFN1267 (Aprobado Disposición SCYT Nro: 187/10)

Introducción

La esteganografía es el arte y la ciencia de la ocultación de información mediante la inserción de la misma en mensajes aparentemente inofensivos. Es el arte de encubrir o de la escritura oculta.

La esteganografía no es la criptografía (el arte de la escritura secreta), cuyo objetivo es hacer un mensaje ilegible por un tercero, pero no oculta la existencia de la comunicación secreta. El objetivo de la esteganografía es encubrir la comunicación para ocultar un mensaje a terceros. La esteganografía esconde el mensaje encubierto, pero no el hecho de que las partes se comunican entre sí. La esteganografía difiere de la criptografía en que el objetivo de la esteganografía es ocultar la existencia de un mensaje mientras que la criptografía es el objetivo de confundir el sentido del mensaje.

Aunque la esteganografía es distinta de la criptografía, hay analogías entre ellas y algunos autores la clasifican como una forma de criptografía porque la comunicación oculta es una forma de escritura secreta [i].

La criptografía implica la transformación de un mensaje original para que toda persona que llegue a encontrar el mensaje transformado, no sea capaz de comprenderlo sin conocer el método correcto para la transformación inversa, generalmente a través de algún contacto o acuerdo con el codificador original.[ii]. En cambio, la

esteganografía es utilizada para ocultar la existencia misma del mensaje. [i]

A diferencia de cifrado, la esteganografía no puede ser detectada. Por lo tanto, se utiliza cuando el cifrado no está permitido. O, más comúnmente, la esteganografía se utiliza para complementar el cifrado. Un archivo cifrado todavía puede ocultar información a través de la esteganografía, así, incluso si el archivo cifrado se descifrara, el mensaje oculto no se ve.

Técnicas de esteganografía

El proceso de la esteganografía en general consiste en colocar un mensaje oculto en algún medio de transporte, llamado el transportista. El mensaje secreto está incrustado en el transportista para formar el medio de la esteganografía. [iii]

Aunque el término esteganografía (literalmente escrito encubierto), fue acuñado al final del siglo XV, el uso de la esteganografía se remonta a varios milenios. La práctica de la esteganografía antigua consistía en el grabado de mensajes en madera y cubriéndolos con cera. Otro método consistía en el tatuaje en la cabeza rapada de mensajeros. Cuando su pelo crecía, viajaban a los campamentos del ejército llevando los mensajes secretos bien escondidos bajo la cabeza llena de cabello. A continuación, se los afeitaba de nuevo cuando llegaban a su punto de contacto. [iv]

Muchos otros tipos de técnicas de esteganografía se han inventado y utilizado desde aquellos tiempos. Algunos ejemplos incluyen tinta invisible, micro puntos fotográficos, escritos en seda, etc.

La tinta invisible ha estado en uso durante siglos para la diversión de los niños y estudiantes y para el espionaje por parte de espías y terroristas. Los micro puntos y el microfilm, elementos básicos de la guerra y películas de espionaje, se introdujeron tras la invención de la fotografía. [v] [vi] [vii] [viii]

La esteganografía mediante computadoras se realiza sustituyendo por bits de información útil, pero invisible, los bits de datos de áreas no utilizadas, en archivos comunes, tales como gráficos, sonido, texto, HTML, o incluso disquetes. Ejemplos de esta información oculta puede ser texto plano, texto cifrado, o imágenes. [i]

Muchos métodos esteganográficos ocultan el contenido en las áreas menos importantes o en el ruido de los mensajes de transporte. Se han desarrollado algoritmos de software para detectar y derrotar a estos métodos esteganográficos. [ix]

El Texto plano

El texto plano es un término general que incluye cualquier material escrito, no el texto codificado en cualquier idioma. Los e-mails, carteles, documentos de procesadores de texto, código de lengua, equipo de origen y las páginas web, son todos ejemplos de documentos de texto plano. Los datos digitales de inserción en los documentos de texto plano, se pueden lograr a través de la manipulación de las propiedades de la lengua, como las partes de la sustitución de expresión.

Marcas de Agua

La esteganografía proporciona algunas funciones muy útiles y comercialmente importantes en el mundo digital, en especial la marca de agua o filigrana digital. En esta aplicación, el autor puede incorporar un mensaje oculto en un archivo de manera que la titularidad de la propiedad intelectual puede ser afirmada más tarde y/o para garantizar la integridad del contenido. Un artista, por ejemplo, podría publicar obras de arte originales en una página Web. Si alguien roba el expediente y realiza alegaciones del trabajo como suyo, el artista puede demostrar la propiedad, ya que sólo él puede recuperar la marca de agua ([x] [xi] [xii]).

Aunque conceptualmente es similar a la esteganografía, la filigrana digital normalmente tiene objetivos técnicos diferentes. En general, sólo una pequeña cantidad de información repetitiva se inserta dentro de la portadora, no es necesario ocultar la información de marca de agua, y es conveniente que pueda ser eliminada sin afectar la integridad de la portadora.

Clasificación de las técnicas

Esteganografía técnica: Utiliza métodos científicos para ocultar un mensaje, tales como el uso de la tinta invisible o micro puntos y otros métodos de reducción de tamaño.

Esteganografía lingüística: Oculta el mensaje en el portador de alguna manera no evidente.

Sema gramas: Ocultan información por el uso de símbolos o signos. Un sema grama visual utiliza objetos físicos de aspecto inocente o cotidianos para transmitir un mensaje, como garabatos; o el posicionamiento de objetos en un escritorio o sitio Web. Un texto sema grama esconde un mensaje modificando la apariencia del texto transportado, tales como cambios sutiles en el tamaño de la fuente o tipo, añadiendo espacios adicionales o diferentes adornos en letras o el texto escrito a mano.

Códigos abiertos: Ocultan el mensaje en un mensaje portador legítimo, de formas que no son evidentes para un observador desprevenido. El mensaje portador se llama a veces la comunicación abierta, mientras que el mensaje oculto es la comunicación secreta. Esta categoría se subdivide en códigos de la jerga y cifrados encubiertos.

Código de la jerga: Utiliza un lenguaje que se entiende por un grupo de personas, pero no tiene sentido para los demás. Los códigos de Jerga incluyen el warchalking (símbolos utilizados para indicar la presencia y el tipo de señal de red inalámbrica [xiii]), la terminología de lenguajes bajos, o una conversación inocente que transmite un significado especial porque los hechos son conocidos sólo por los interlocutores. Un subconjunto de los códigos de la jerga son los códigos de referencia, en donde ciertas frases preestablecidas transmiten el significado.

Los encubridores o sistemas de cifrado de ocultamiento: Ocultan un mensaje abiertamente en el medio portador de manera que pueda ser recuperado por cualquiera que conozca el secreto de cómo se ha ocultado. Un sistema de cifrado parrilla emplea una plantilla que se utiliza para cubrir el mensaje portador. Las palabras que aparecen en los inicios de la plantilla son el mensaje oculto. Un cifrador nulo esconde el mensaje según un conjunto de reglas preestablecidas, como "leer cada quinta palabra" o "mirar el tercer carácter en cada palabra."

Inserción / Detección: Mucha investigación se está realizando en las áreas de inserción y en la detección de secreto de los datos digitales, en imágenes gráficas y multimedia de los objetos de datos, como archivos de música. Hay un gran número de programas que permiten ocultar datos en imágenes o archivos. Hay numerosas y bien publicadas maneras de utilizar la esteganografía en la ocultación de la información en archivos de imagen y sonido [xiv] [xv] [xvi]. Sin embargo, un área menos considerada es el ocultamiento de la información sencilla dentro de archivos de oficina común. Estos espacios no son bien conocidos o bien documentados. Se pueden utilizar con relativa facilidad para ocultar los datos y su uso disminuye la sospecha como se ha indicado anteriormente.

Motivación

Hay dos cuestiones principales que impulsan la tecnología para ocultar información.

En el primer grupo se encuentran los que están tratando de proteger sus derechos de propiedad intelectual. Con la alta disponibilidad de la información a través de Internet es cada vez más difícil proteger la propiedad intelectual y hacer cumplir las leyes de derechos del autor.

Las marcas de agua digitales proporcionan una forma de insertar un aviso de derechos de autor en un documento o imagen. La marca de agua es a menudo una pequeña imagen o un bloque de texto que se repite con frecuencia a través de la imagen o documento. Una técnica similar es la de insertar una huella digital o el número de serie. La ventaja de una huella digital es que puede ser utilizada para rastrear la copia al original y es una herramienta poderosa para perseguir violadores de derechos de autor. [xvii]

El segundo grupo de personas que están interesados en la ocultación de información son los que desean transmitir información de forma encubierta y evitar la observación por personas ajenas. En este caso, el mensaje oculto es más importante que el transportista que se utiliza para transportarlo.

Creemos necesario indicar la necesidad de definir un tercer grupo, en el que se encuadra el presente proyecto, es aquel que utiliza ampliamente documentación oficial que se difunde o procesa mediante sistemas informáticos.

Existe la necesidad de validar o certificar que el documento utilizado representa el original, que se encuentra libre de adulteraciones. Es importante que se pueda verificar tanto por parte del emisor como por parte del receptor de esa documentación, aún cuando cambie el medio de soporte.

Analizando los múltiples estudios que durante años se han realizado en pos de la seguridad de la documentación electrónica, se verifica que las aplicaciones utilizadas son múltiples, para resolver situaciones similares. Aunque encontramos buenos resultados, la ocurrencia de falencias o el descubrimiento de debilidades/vulnerabilidades ponen en tela de juicio cada una de esas herramientas. En la documentación que se presenta electrónicamente, es muy común el cambio de formato, del medio de soporte, etc. todo ello atenta contra el uso de esas herramientas.

Deberíamos disponer de herramientas que soporten la existencia de varios formatos y que sean independientes del medio en que se presenten.

No se conoce, por lo analizado a la fecha, que se haya investigado la posibilidad de crear una metodología de validación en documentación electrónica cuyo medio de soporte haya cambiado o se haya alterado.

Con esta idea, este trabajo dará continuidad a lo ya realizado, estableciendo metodologías que permitan la adaptabilidad continua de los sistemas que se utilicen.

La construcción de un paquete de seguridad en la documentación, consiguiendo la integración de software ya desarrollado con implementaciones propias, originales de este proyecto, impactará principalmente en el ámbito de la seguridad de las entidades públicas municipales, provinciales y nacionales y en todos aquellos que actualmente presentan documentación en formato electrónico

que luego es utilizada en actos administrativos formales.

Según los resultados a los que podamos arribar, las repercusiones tecnológicas podrían dar lugar a futuros proyectos de investigación aplicada y desarrollos tecnológicos, ya que este trabajo busca tener el carácter de libre y gratuito con el fin de progresar indefinidamente.

Líneas de investigación y desarrollo

La actividad que realizamos es del tipo desarrollo experimental, cuyo campo de aplicación es la Seguridad de la Información y de la documentación.

Resultados y Objetivos

El desarrollo del trabajo se enfoca en la búsqueda de:

- Establecer métodos de verificación de la validez de la documentación que sean resistentes a adulteraciones. La metodología debe considerar múltiples procedimientos e instrumentos diferentes.

  Esta metodología debería otorgar resultados de valor independientemente de quien la utilice, basándose en el análisis del documento, mediante el uso de herramientas disponibles actualmente en la mayoría de los sistemas actuales.

- Analizar la posibilidad de la implementación como medio de mejorar la seguridad de todo el sistema de documentación.

Para lograrlo se creará una metodología y herramientas que tendrán aplicaciones múltiples: para el origen de la información o documentación, y le permitirá proveer material confiable y verificable. Asimismo, le podrá certificar que lo retornado, luego de su circuito administrativo, corresponde certeramente con lo esperado.

Para el receptor o usuario, logrará incrementar la seguridad al facilitar la validación o comprobación del material provisto.

El diseño se basará en metodologías de programación de amplia difusión y con acceso Web. El diseño flexible debe ser adaptable rápidamente a las necesarias adecuaciones que permanentemente se requieren.

Los Objetivos propuestos son:

* Lograr un paquete de soluciones para la emisión de documentos informáticos, seguro, fiable y de muy bajo costo.

* Conseguir los protocolos de operación necesarios que permitan el desarrollo de una solución factible.

* Lograr un producto con licencias libres y/o gratuitas de los desarrollos de software que se lleven a cabo.

* Realizar análisis comparativo de software, hardware necesario, y los costos de los mismos.

* Realizar la documentación de los procesos y del software.

* Instalación del software producido y análisis de su desempeño.

* Capacitar y dar soporte a usuarios en fase de prueba.

Se pretende lograr un sistema confiable y viable desde el punto de vista económico y técnico, bajo licencias gratis y/o libres, para que el mismo pueda ser modificado, adaptado y difundido sin limitaciones, y con el fin de conseguir día a día un paquete de seguridad robusto.

Se busca establecer una metodología, en base a herramientas esteganográficas, que brinde una respuesta adecuada a la necesidad de validación de la documentación.

Lo realizado

El equipo de trabajo ha desarrollado la investigación del tema que da origen al presente proyecto, desde la obtención del aval del consejo departamental al grupo de seguridad informática. Se realizaron los estudios del estado del arte en seguridad informática, sistemas de cifrado, protocolos de autenticación, sistemas de marca de agua o filigrana digital y se analizaron múltiples procedimientos que se realizan a diario en los circuitos administrativos de la Facultad Regional Córdoba. Se puso especial énfasis en aquellos trámites que tienen su origen en documentos digitalizados o generados por medios computacionales.

Se seleccionó la esteganografía como herramienta adecuada para los fines del proyecto, restando la investigación profunda sobre las particularidades que tendrá el desarrollo planteado. Se dispone de un plan de trabajo para realizar la investigación y se constituye formalmente el proyecto encontrándose en la etapa de codificación de la muestra de concepto.

Formación de Recursos Humanos

Este proyecto proveerá a la interacción entre grupos de investigación de dos Facultades: la Facultad Regional Río Grande y la Facultad Regional Córdoba. Asimismo contribuirá a la formación y crecimiento, como investigadores, de los involucrados.

Dentro del proyecto se contempla la formación de alumnos de la carrera de Ingeniería en Sistemas de Información.

Por otro lado se prevé la capacitación de alumnos becarios que formarán parte del equipo mientras el proyecto dure. De esta manera, se generará un grupo de trabajo importante que promocione este tema.

Cabe desatacar también la importancia de formar estos grupos de investigación y de trabajo para la Universidad, para la Ciudad y para el país.

El crecimiento de este grupo podrá ser el eje de los conocimientos en estos Temas y la base para el desarrollo de futuras investigaciones relativas a la seguridad de la documentación de la Universidad.

Vemos también la posibilidad de colaborar con el crecimiento profesional de los integrantes del grupo, los cuales se ven predispuestos a perfeccionarse continuamente.

El grupo está compuesto por: Director, Co-Director, tres profesores investigadores de apoyo, y tres estudiantes investigadores. Durante el presente año se incorporará un cuarto estudiante.

Referencias

I.) Bauer, F. L. Decrypted Secrets: Methods and Maxims of Cryptology, 3rd ed. Springer-Verlag, New York, 2002.

II.) Jonathan Watkins, “Steganography – Messages Hidden in Bits”, Multimedia Systems. Coursework, Dept of Electronics and CS, University of Southampton, SO17 1BJ, UK.

III.) Gary C. Kessler. An Overview of Steganography for the Computer Forensics Examiner. July 2004 – Forensic Science Communications Volume 6 – Number 3. Disponible en http://www.fbi.gov/hq/lab/fsc/backissu/july2004/research/2004_03_research01.htm

IV.) Paul Davern, “Steganography: its history and its application to computer based data files”, Dublin City University School of Computer Applications, Working Paper: CA-0795

V.) Arnold, M., Schmucker, M., and Wolthusen, S. D. Techniques and Applications of Digital Watermarking and Content Protection. Artech House, Norwood, Massachusetts, 2003.

VI.) Johnson, N. F., Duric, Z. and Jajodia, S. Information Hiding: Steganography and Watermarking: Attacks and Countermeasures. Kluwer Academic, Norwell, Massachusetts, 2001.

VII.) Kahn, D. Codebreakers: The Story of Secret Writing. Revised ed., Scribner, New York, 1996.

VIII.) Wayner, P. Disappearing Cryptography: Information Hiding: Steganography & Watermarking. 2nd. ed., Morgan Kaufmann, San Francisco, California, 2002.

IX.) Jessica Fridrich, and Miroslav Goljan, “Practical Steganalysis of Digital Images – State of the Art “, Dept ECE at SUNY Binghamton.

X.) Arnold, M., Schmucker, M., and Wolthusen, S. D. Techniques and Applications of Digital Watermarking and Content Protection. Artech House, Norwood, Massachusetts, 2003.

XI.) Barni, M., Podilchuk, C. I., Bartolini, F., and Delp, E. J. Watermark embedding: Hiding a signal within a cover image, IEEE Communications (2001) 39(8):102-108.

XII.) Kwok, S. H. Watermark-based copyright protection system security, Communications of the ACM (2003) 46(10):98-101.

XIII.) Warchalking. Warchalking: Collaboratively creating a hobo-language for free wireless networking [Online]. (December 21, 2003). Disponible: http://www.warchalking.org/

XIV.) Dumitrescu S, W. Xiaolin and W. Zhe. 2003. “Detection of LSB steganography via sample pair analysis Signal Processing.” IEEE Transactions on. 51(7): 1995 -2007.

XV.) Petitcolas, F. 2002. The Information Hiding Homepage digital watermarking & steganography. [website] Available at http://www.cl.cam.ac.uk/~fapp2/steganography

XVI.) Bender G. 1996. “Techniques for Data Hiding.” IBM Systems Journal 35 (3/4): 313-337

XVII.) Lisa M. Marvel, “Image Steganography for Hidden Communication”, Dissertation – University of Delaware, Spring 1999.