Utilización de Watermarking para Seguridad en la Nube: el Caso de las Imágenes Médicas

Laura Mónica Vargas[1,2], María Alejandra Di Gionantonio[3]

laura.monica.vargas@unc.edu.ar, ing.alejandradg@gmail.com

[1] Laboratorio de Redes y Comunicaciones de Datos, Departamento de Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

[2] Laboratorio de Procesamiento de Señales, Departamento de Matemática, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

[3] Laboratorio de Investigación de Software, Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba, Universidad Tecnológica Nacional

I. RESUMEN

En los últimos años, acompañando el rápido avance de las telecomunicaciones se ha desarrollado la telemedicina por medio de la cual los médicos pueden transferir y compartir los datos digitales de los pacientes en forma remota para determinar un diagnóstico definitivo. Por otra parte, actualmente la tendencia es llevar la información médica que se almacenaba en el propio centro de salud a la nube siendo esencial en estos casos proteger los datos médicos intercambiados. En las plataformas de Cloud Computing, la seguridad es todavía un problema importante a resolver. Se debe garantizar que las imágenes médicas se puedan compartir en forma segura preservándolas de cualquier intento de distorsión, como así también proporcionar privacidad en las cadenas de datos de las historias clínicas de salud o Electronic Health Records (EHR). Una alternativa de solución es la inserción de marcas de agua en las imágenes médicas, técnica conocida como *watermarking*, cuya aplicación en imágenes digitales empezó hace unas décadas. En este trabajo se propone que acompañe a las clásicas técnicas criptográficas para perfeccionar la

Palabras clave: seguridad informática, watermarking, imágenes digitales, cloud computing, telemedicina.

II. CONTEXTO

El presente trabajo se realiza en los siguientes ámbitos: Laboratorio de Redes y Comunicaciones de Datos perteneciente al Departamento de Computación y en el Laboratorio de Procesamiento de Señales del Departamento de Matemática, ambos en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba, así como en el Laboratorio de Investigación de Software, Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

La temática de watermarking ha estado presente en los Proyectos aprobados por SeCyT-UNC de ID, todos ellos Categoría A: **Reducción de ruido, compresión y seguridad en la transmisión de** señales **1-D y 2-D,** Código 05M/160, años 2010-2011, **Restauración, reconocimiento de patrones y transmisión segura de imágenes**, Código 05/M232, años 2012-2013 y **Restauración, reconocimiento de patrones y transmisión segura de imágenes,** Código 05M/296, años 2014-2015, de todos los cuales la primera autora ha sido **SubDirectora.**

La primera autora realizó su tesis de doctorado en la FCEFyN-UNC aplicando métodos de Inteligencia Artificial a algoritmos de marcado y la segunda autora logró el grado de Especialista en [illegible] la FRC-UNC con una tesis sobre la utilización de marcas de agua para autenticación de imágenes médicas.

La primera autora ha dirigido tesis para alcanzar el grado de Ingeniero en Computación de la FCEFyN consistentes en diseño e implementación de marcas de agua durante los años 2012-2013. También ha participado de congresos y jornadas presentando algoritmos de implementación de marcas de agua (INMAT 2008, Congreso

Internacional de Ingeniería 2010, MACI 2013 y 2015, ACCN 2015 entre otros), publicado en IEEE y en Int. Journal of Advances in Processing Images Techniques. Ambas autoras han publicado un artículo sobre el tema en la Revista de la FCEFyN-UNC en 2016.

III. INTRODUCCIÓN

En la etapa actual, la investigación se centra en el análisis del problema de seguridad de la información contenida en las imágenes médicas cuando estas son alojadas en un framework de Cloud Computing.

Cloud Computing es un mecanismo que creció en los últimos años, basado en la Web que permite escalar y virtualizar recursos de TI que son proporcionados como servicios a través de la red. Características inherentes y esenciales que deben ser provistas por las aplicaciones de cloud computing son: servicio bajo demanda, acceso ubicuo, escalabilidad, elasticidad, independencia del usuario respecto al mantenimiento y pago por uso, siendo la seguridad todavía un desafío[1] [2].

En las últimas décadas, antes del desarrollo de la nube, se logró un progreso significativo en el uso de tecnologías de comunicación para almacenar y distribuir datos médicos bajo formatos digitales [3]. Al comienzo de los años 80 apareció el PACS (Picture Archiving and Communication System/ Sistema de Comunicación y Almacenamiento de Imágenes). Inicialmente, estos se desarrollaron para cubrir necesidades específicas de los centros médicos, tales como adquisición de datos y visión de estos en estaciones de trabajo de poca capacidad. En un principio solo se asociaban a radiología, pero luego abarcaron todo tipo de imágenes médicas [4]. Rápidamente se extendieron en los centros médicos, pero presentaban problemas de interoperabilidad ya que eran desarrollados en forma independiente por distintos proveedores que seguían sus propias reglas. Surgió entonces la Norma DICOM (Digital Imaging and Communications in Medicine), desarrollada por el ACR (American College of Radiology) en conjunto con NEMA (National Electrical Manufacturers Association) [5]. Tras varios intentos se aprobó en 1993 y desde entonces sufre una actualización constante. Esta norma define el acceso a la web, la estructura de intercambio de datos, las capas de comunicación y los comandos para manejo de imágenes médicas que deben respetar todos los fabricantes para obtener interoperabilidad. Su aceptación hizo que se "dicomizaran" los PACs. Actualmente los fabricantes de equipos para imágenes médicas, siguiendo indicaciones de la norma, los acompañan de un CS (Conformance Statement) que asegura que cumplen con la misma. Esta norma permite el acceso remoto a archivos en formato dicom (extensión dcm) utilizando los ya clásicos protocolos TCP/IP, el Protocolo HTTP (Hypertext Transfer Protocol) o HTTPS (Hypertext Transfer Protocol Secure). Si bien aseguró interoperabilidad entre los distintos sistemas y demostró cierta flexibilidad en entornos que manejan imágenes médicas, no hizo un aporte significativo a la seguridad, ni al acceso de datos por fuera de instituciones médicas

El empleo de las imágenes no es seguro cuando los datos, médicos o de otro origen, circulan libremente por redes abiertas como Internet, expuestos a que los mismos sean alterados o mal utilizados. Esto sucede cuando se utilizan servicios de teleconsulta o telediagnóstico, los que se están difundiendo por todo el mundo con el aporte de las Tecnologías de Información y Comunicación.

En un principio, la norma DICOM no presentó ninguna disposición respecto a seguridad ya que no era una preocupación en el momento de su generación. La parte 15, agregada posteriormente, actualizada por última vez en 2016, pide el uso de protocolos de comunicación seguros, encapsulamiento en formatos seguros, encriptación de los datos y empleo de firmas digitales. Sin embargo, la firma digital puede ser separada del archivo por lo que son varios los autores que señalaron la necesidad de agregar otro nivel de seguridad, incluso antes de que el almacenamiento de la información médica se hiciera en la nube [6]. Almacenar en la nube es una buena alternativa ya que permite a los centros médicos desentenderse del hardware y software usado en los sistemas de archivos, pero conlleva un mayor riesgo de violación de autenticidad e integridad en los registros del paciente [7]. En la actualidad, los sistemas de información no son centralizados, sino distribuidos y, por lo tanto, el control también debe

estar distribuido. La protección primitiva, que controlaba el acceso mediante claves, no es suficiente. Por otro lado, la confidencialidad de los datos es una necesidad ética en el campo de la salud que puede lograrse por medio de la encriptación clásica [8], herramienta que, sin embargo, no resulta suficiente para solucionar todos los problemas de protección de datos digitales.

En 1996, se dictó en EEUU, la HIPAA (Health Insurance Portability and Accountability Act) que indica qué requisitos se deben cumplir para las transacciones de datos de salud con el objetivo de que los datos médicos se almacenen y se puedan recuperar a largo plazo, evitando abusos y fraude. Microsoft en 2007 y Google en 2008 ofrecieron portales Health a los usuarios que querían que sus EHRs estuvieran disponibles para sus servicios de salud y para ellos mismos. En 2010, IBM y Aetna en conjunto anunciaron un nuevo uso de la plataforma de cloud computing de IBM diseñada para ayudar a los profesionales de la salud a acceder rápidamente a la información del paciente: registros médicos, recetas, y datos de laboratorio recolectados de múltiples fuentes para crear un registro detallado del mismo. Se estima que en el año 2020 el 80% de los datos se habrá mudado a la nube. El uso de estas plataformas y otras permite a los centros médicos desentenderse de problemas técnicos (actualización y mantenimiento de software y hardware), económicos y legales relacionados al manejo de datos lo que le conviene más allá de los riesgos que corre. Entre los inconvenientes se encuentra la latencia, la dificultad para tener el servicio disponible todo el tiempo, y la seguridad [9]. Se debe tener especialmente en cuenta que los datos almacenados en la nube son vulnerables a ataques internos. La identidad y ubicación de intermediarios y de los proveedores de servicio está disimulada, oculta, por la nube.

En forma paralela a estos avances en telecomunicaciones, se empezó a desarrollar en la década del 90, el watermarking o marcado de productos multimedia, imágenes, videos, audio, gráficos, etc., como forma de protección de propiedad intelectual [10]. Consiste en embeber bits en el archivo, sea imagen, video o audio, de forma visible (audible) o invisible (no audible). Estos bits extra constituyen la marca y en las primeras implementaciones permitían identificar al propietario, utilizándose posteriormente para alcanzar otros propósitos como detección de adulteraciones, aseguramiento de integridad e incorporación de metadatos. Algunos métodos permiten en el caso de haberse producido una alteración, determinar en qué sector del archivo sucedió. Así, en estos días, la marca de agua aparece como un medio eficiente para asegurar integridad y verificar autenticidad. Es usual que la marca sea la firma digital del archivo o metadatos encriptados siguiendo técnicas clásicas. Mientras que con la encriptación se espera que no se entienda lo que se ve, con el watermarking hay datos ocultos que no se pueden ver.

El watermarking puede ser reversible o irreversible. El reversible permite recuperar el archivo original. Para lograrlo es preciso haber guardado cierta información. En la Figura 1 se presenta un esquema básico de watermarking reversible, que es el que nos interesa ya que permite recuperar la imagen original algo necesario en imágenes legales como lo son las médicas.

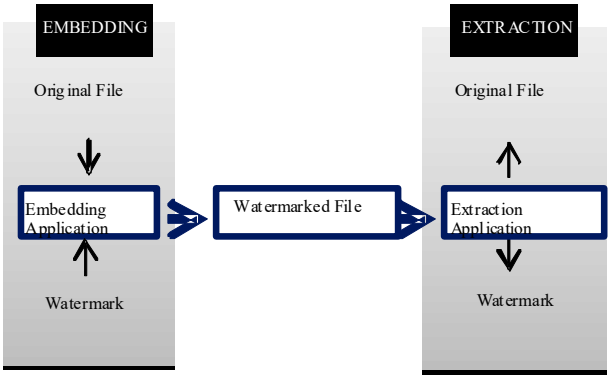

Figura 1. Esquema de Marcado Reversible

En cuanto a los algoritmos de embebido se han realizado numerosos desarrollos, en el dominio espacial (modificación del bit menos significativo o LSB) y en el de la frecuencia, en particular. En este último dominio son de especial interés los desarrollos efectuados usando la transformada wavelet y la transformada discreta de coseno (DCT). Entre los algoritmos clásicos para imágenes se cuentan el de Tian [11] de expansión de la diferencia entre pares de píxeles y el de Ni [12] de corrimiento del histograma. Siempre es necesaria la validación teniendo en cuenta que para distintos tipos de imágenes pueden ser convenientes distintos

algoritmos, no existiendo uno de utilidad universal.

La evaluación de las marcas de agua reversibles se hace considerando el tiempo de ejecución del algoritmo, la capacidad y el error medio cuadrático MSE o relación señal/ruido PSNR para cuantificar imperceptibilidad, la que también debe ser objeto de una evaluación subjetiva por parte de expertos. Estos parámetros responden a las siguientes expresiones:

$$Capacidad(bpp) = \frac{N_{bm}}{Cantidad\ total\ \ pixeles\ imagen}$$

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_W(i,j) - I(i,j)\right)^2}{MN}$$

$$PSNR(dB) = 10\log_{10}\frac{\left[I_{pico}\right]^2}{MSE}$$

donde N_{bm} indica la cantidad de bits de la marca, M la cantidad de filas y N la cantidad de columnas de la imagen, I(i,j) el valor de la intensidad en la posición del píxel fila i, columna j e $I_w(i,j)$ indica el valor de la intensidad en la posición i,j luego del marcado.

La importancia del watermarking en imágenes médicas se destaca desde hace dos décadas [13], hasta la actualidad [14]. Es de particular interés que se embeban los datos del paciente en sus imágenes médicas personales. Las amenazas, como la destrucción de sistemas de software y violación en los accesos, están surgiendo con frecuencia en la plataforma de la nube, por lo que se hace absolutamente necesario tomar medidas para contrarrestarlas. Se recomienda en telemedicina, la combinación de watermarking con técnicas de criptografía clásicas [13-17]. La encriptación puede impedir problemas en los nodos intermedios, pero no en los puntos finales que deben poder desencriptar los datos y si el proveedor del servicio confía, a su vez, en otros proveedores entonces los datos del usuario pueden ser leídos por muchas entidades en la nube. Lo que se precisa para incrementar la confianza en la nube siguiendo esta línea de razonamiento es algún mecanismo que pueda detectar y castigar cualquier problema relativo a la confidencialidad. El usuario final debe confiar en la entidad que administra los EHR (sea Google Health, Microsoft Health Vault u otro proveedor de nube). El administrador de EHR debe tener los medios para detectar y castigar las violaciones a la confidencialidad que se hubieran producido. Las nuevas técnicas deben tratar especialmente de mitigar los riesgos de que la información sufra ataques internos en la nube [18].

La propuesta actual de investigación consiste en diseñar e implementar un mecanismo que combine watermarking y encriptación para lograr asegurar integridad y autenticidad de datos médicos. Se propone utilizar un sistema de archivos distribuidos Hadoop, y MapReduce, modelo de programación para el manejo de grandes bases de datos [19].

IV. OBJETIVOS Y LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los objetivos de este proyecto de investigación los podemos dividir en:

A. Objetivo General

-Implementación y evaluación de técnicas de watermarking junto con las criptográficas tradicionales, para otorgar seguridad a las imágenes digitales médicas almacenadas en la nube.

B. Objetivos Específicos

-Analizar el estado del arte de cloud computing para servicio de almacenamiento de imágenes médicas.
-Analizar las soluciones implementadas para obtener seguridad en el almacenamiento de imágenes médicas en la nube.
- Utilizar infraestructuras como Hadoop para simular las aplicaciones.
-Difundir los resultados obtenidos para realimentar el proceso de desarrollo de los algoritmos de watermarking.
-Formar recursos humanos específicos en las áreas objeto de estudio.

V. MATERIALES Y MÉTODOS

En nuestra investigación proponemos hacer frente al problema de la seguridad de los datos contenidos en imágenes médicas alojadas en Cloud Computing, utilizando la técnica de marca de agua en el EHR. Luego se envía la imagen con marca de agua al proveedor de la Nube. En particular se trabajará con imágenes médicas en formato dicom y

con marcas reversibles indetectables.

Es común que una imagen médica sea diagnosticada antes de que la misma sea almacenada en un almacenamiento a largo plazo, de este modo la parte significativa de la imagen, conocida como ROI (Region of Interest), es determinada en ese momento. El embebido de información extra se hará fuera de esta zona.

VI. FORMACIÓN DE RECURSOS HUMANOS

La segunda autora está realizando su tesis de Maestría en Sistemas de información en la FRC-UTN con el tema "Análisis comparativo de múltiples plataformas de Health Cloud Computing para hosting de imágenes médicas con marcas de agua".

Se dirigirán trabajos finales sobre la temática abiertos a estudiantes de Ingeniería en Sistemas de Información e Ingeniería en Computación.

REFERENCIAS

[1] Youssef *et al.* "Toward a Unified Ontology of Cloud Computing". Grid Computing Environment Workshop, IEEE, 2008.

[2] Jadeja and Modi. "Cloud Computing – Concept, Architecture and Challenges". International Conference on Computing, Electronics and Electrical Technologies, IEEE, 2012.

[3] Acharya *et al*. "Compact Storage of medical Images with patient Information". IEEE Transactions on Information Technology in Biomedicine, vol.5, pp. 320-323, 2001.

[4] Bharath. "Introductory Medical Imaging". Ed. John Enderle, University of Connecticut, 2009.

[5]Pianykh. "Digital Imaging and Communications in Medicine (DICOM). A Practical Introduction and Survival Guide". 1st Ed., Ed Springer, 2008.

[6] Ahmed and Abdullah. "Telemedicine in a Cloud - A Review". IEEE Symposium on Computers and Informatics, 2011.

[7] Ahuja, Mani and Zambrano. "A Survey of the State of Cloud Computing in Healthcare", Network and Communications Technologies, Vol. 1 N°2. Canadian Center of Science and Education, 2012.

[8] Stallings. "Cryptography and Network Security", Ed. Prentice Hall, 4th Ed., 2005.

[9] Zhang and Liu. "Security Models and Requirements for Healthcare Application Clouds". 3rd International Conference on Cloud Computing", IEEE, 2010.

[10] Cox *et al.* "Digital Watermarking". Ed. Morgan Kauffmann. 2002.

[11] Tian. "Reversible Data Embedding using a Difference Expansion". IEEE Transactions on Circuits Systems and Video Technology, vol. 13, no. 8, pp. 890-896, 2003

[12] Ni *et al.* "Reversible data Hiding". Proceedings of the 2003 International Symposium on Circuits and Systems, vol. 2, pp. 912-915, 2003

[13] Coatrieux *et al*. "Relevance of Watermarking in Medical Imaging". Proceedings of the IEEE EMBS Conf. on Information Technology Applications in Biomedicine, Arlington, USA, pp 250-255, 2000.

[14] Aminzou *et al.* "Towards a Secure Access to Patient Data in Cloud Computing Environments". Security Days (JNS3), IEEE, 2013.

[15] Elgamal, Hikal & Abou-Chadi. "Secure Medical Images Sharing over Cloud Computing environment". International Journal of Advanced Computer Science and Applications (IJACSA), vol. 4, N°5, 2013.

[16] Bouslimi and Coatrieux. "A Joint/Encryption Watermarking System for Verifying the Reliability of Medical Images". Medical Data Privacy Handbook pp. 493-526, Springer, 2015.

[17] Al-Haj, Hussein and Abandah. "Combining Cryptography and Digital Watermarking for Secured Transmission of Medical Images". Second International Conference on Information Management (ICIM), IEEE; 2016.

[18] Garkotti *et al*. "Detection of Insider attacks in Cloud based e-healthcare". International Conference on Information Technology, IEEE, 2014.

[19] Lee *et a*l. "Implementation of MapReduce-based Image Conversion Module in Cloud Computing Environment". Proc. of Int. Conference on Advances in Computing, Control and Telecommunication Technologies, 2011.