Prevención de exposición y filtración de información sensible en dispositivos móviles

Trabajo Terminal No. _ _ - _ _

*Alumna: *Jeon Jeong Paola*

Directores: Aguirre Anaya Eleazar, Juárez Gambino Joel Omar

**habeet90@gmail.com*

Resumen - En este trabajo terminal se propone identificar los documentos de texto contenidos en los dispositivos móviles Android y analizar la información para clasificarla de acuerdo con la sensibilidad de su contenido. Mediante técnicas de procesamiento de lenguaje natural, se analizarán los documentos y se identificarán las características que permitan la clasificación de la información contenida en los documentos. Los grupos en los que se clasificará la información será en: datos no sensibles, datos personales y datos personales sensibles. Por otro lado, a través de técnicas de informática forense, se identificarán los documentos que contengan información en lenguaje natural y se utilizarán herramientas antiforenses que evitarán su extracción del dispositivo móvil y criptográficas que protejan su confidencialidad. El usuario tendrá la opción de proteger la información almacenada o eliminar información que consideré es crítica para permanecer en el móvil, para que finalmente, en el dispositivo únicamente se encuentren documentos seguros y no vulnerables ante terceros.

Palabras clave - clasificación de texto, forense digital, procesamiento de lenguaje natural, técnicas antiforenses.

1. Introducción

Los dispositivos móviles se han hecho populares a lo largo de los años en la sociedad. En el 2020, 88.2 millones de usuarios tenían acceso a los teléfonos celulares, de los cuáles más de la mitad eran teléfonos inteligentes (smartphones). Los teléfonos celulares son herramientas indispensables en nuestra vida diaria, ya que realizamos llamadas, los utilizamos como agenda, como reloj, etc., dicho de otra manera, sin un teléfono celular, nos desconectamos del mundo. Una de las operaciones más comunes que realizamos al utilizar un dispositivo móvil es recibir y/o enviar archivos como multimedia, mensajes y otros documentos con información personal [1]. Dentro de esta información personal pueden existir datos sensibles que vulneren nuestra seguridad y es por eso que, en este trabajo terminal se propone identificar documentos de texto de acuerdo a su contenido y determinar la sensibilidad de estos.

Tanto es la importancia de mantener seguros los documentos que existen algunas herramientas que proveen este tipo de servicios, por ejemplo:

- ➔ Azure Information protection, es una herramienta dedicada a la nube, que clasifica, ordena y protege datos dependiendo del tipo de sensibilidad que tienen [2].
- ➔ CipherTrust de Thales, es una plataforma que permite a los usuarios finales mantener una fuerte propiedad de sus datos en las instalaciones y en la nube, al igual que en el traslado de cargas de trabajo y datos confidenciales a la nube.[3]
- ➔ ManageEngine DataSecurity Plus es un software el cual evita que archivos que contienen información sensible sean copiados a dispositivos USB.[4]

Como podemos observar, las herramientas mencionadas anteriormente, son muy eficaces y proveen una variedad de servicios, sin embargo, se enfocan principalmente en la nube, algo que difiere sobre el trabajo terminal propuesto. Los dispositivos móviles transmiten, procesan y almacenan información sensible que puede ser extraída por malware o físicamente por técnicas forenses. En este sentido, resulta indispensable identificar la información

sensible y protegerla, es por eso que la protección de la información, se podrá realizar por medio de la identificación de los documentos que contengan información en lenguaje natural a través de técnicas forenses, se analizarán su contenido por algoritmos de procesamiento en lenguaje natural y finalmente se protegerá por medio de algoritmos de cifrado y técnicas antiforenses.

Cuando damos a conocer nuestra información, estamos violentando la privacidad, es por eso que tener protegidos nuestros datos personales, nos protege a nosotros [5]. Monitorear los que estamos enviando y recibiendo y almacenando evita que la situación no llegue a consecuencias mayores.

2. Objetivo

Diseñar un control de seguridad que proteja la información sensible contenida en dispositivos móviles por medio de técnicas de procesamiento de lenguaje natural, herramientas de informática forense y algoritmos criptográficos que evite su exposición y exfiltración por terceras entidades no deseadas.

Objetivos específicos

- Extraer el contenido de los documentos de texto.
- Separar la información sensible por medio de taxonomías establecidas.
- Cifrar información sensible

3. Justificación

Los trabajos relacionados con la identificación de los datos contenidos en un documento imponen un impacto fuerte en la seguridad personal [6]. Existen varias herramientas que se dedican a la clasificación de documentos, sin embargo, muchos de estos están enfocados en la nube [2]. Por otro lado, la implementación de la seguridad preventiva a través de herramientas y técnicas de informática antiforense, ayuda a que los documentos no sean vulnerables en manos de agentes externos que utilizan la información sensible como una amenaza. Un aspecto que hace que este proyecto sea importante, es el análisis del contenido de documentos que se encuentran en un dispositivo móvil, el cual es una herramienta que se utiliza por muchas personas en el mundo.

4. Productos o resultados esperados

Resultados:

- Identificación de archivos con contenido en lenguaje natural en dispositivos móviles
- Clasificación de datos sensibles contenidos en documentos de texto.
- Clasificación de datos sensibles en datos personales y datos personales sensibles
- Control anti exfiltración de información sensible por accesos no autorizados

Productos

- Prototipo de una herramienta de protección que previene la exposición de información sensible implementada en dispositivos móviles. Se seleccionarán los componentes que presenten un grado de innovación para el proceso de solicitud de registro. Cabe destacar que esta propuesta es un trabajo terminal de investigación, no corresponde a un desarrollo de software tradicional, por lo tanto, el nivel de madurez de los componentes de software serán de un prototipo en el nivel 3 o 4 del TRL

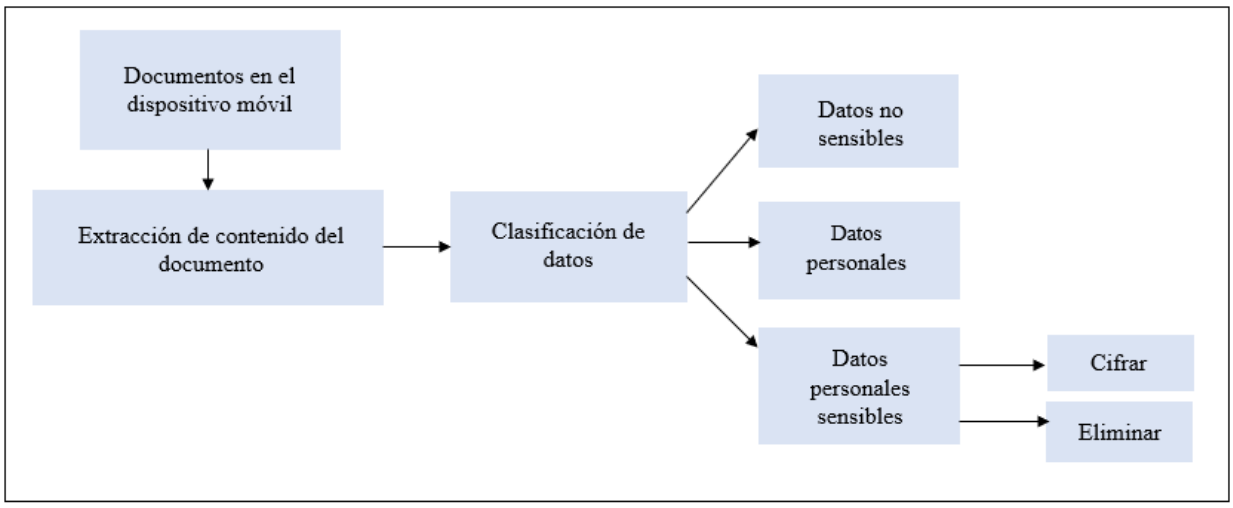

Figura 1. Diagrama del modelo general de la propuesta.

5. Metodología

La metodología que se utilizará en este trabajo será a través de la combinación de los métodos científicos y experimentales. Este tipo de metodología empieza cuando se establece una hipótesis, se trabaja con esta hipótesis a través de experimentos y los resultados de los experimentos realizados son analizados para identificar variables y comportamientos que permitirán el diseño de control preventivo [7].

El uso del método científico en el trabajo propuesto, permite realizar experimentos las veces necesarias para resolver todos los problemas que involucran la protección de la información frente a la exposición y exfiltración del dispositivo móvil..

Las etapas de esta metodología son:

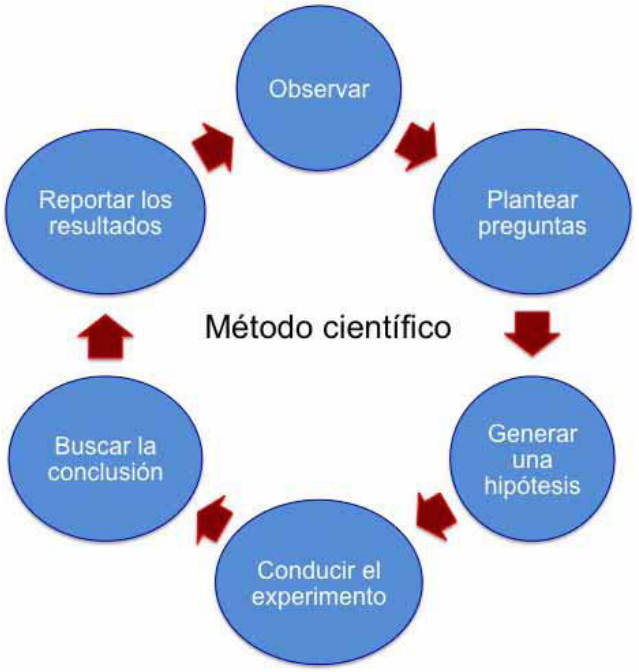

Figura 2. Diagrama de la metodología.

6. Cronograma

CRONOGRAMA Nombre del alumno (a): Jeon Jeong Paola TT No.:
Título del TT: Prevención de exposición y filtración de información sensible en dispositivos móviles

Actividad	Agosto	Septiembre	Octubre	Noviembre-Diciembre	Enero	Febrero	Marzo	Abril	Mayo-junio
Análisis de herramientas de procesamiento lenguaje natural									
Recuperación de los documentos de texto									
Extracción de la infromación contenida en los documentos de texto									
Identificación de datos de interés para el proceso de clasificación									
Evaluación de Trabajo Terminal I									
Análisis de clasificadores									
Clasificación de los documentos de texto									
Cifrado de los documentos de txexto									
Evaluación de Trabajo Terminal II									

7. Referencias

[1] INEGI. En México hay 84.1 millones de usuarios de Internet y 88.2 millones de usuarios de teléfonos celulares: ENDUTIH 2020, presentado en el comunicado de prensa núm. 352/21. Ciudad de México, CDMX, 2020, pp.20. Disponible en https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/OtrTemEcon/ENDUTIH_2020.pdf

[2] Microsoft (2017). Azure Information protection: End user adoption guide, 22, 4-5.

[3] Thales. (2022). Thales aumenta la confianza de los clientes en la nube (1er ed.). [En línea]. Disponible en: https://prensariotila.com/thales-aumenta-la-confianza-de-los-clientes-en-la-nube/

[4]S. Cooper. (2021). 6 best sensitive data discovery tools (1er ed.) [En línea]. Disponible en: https://www.comparitech.com/data-privacy-management/best-sensitive-data-discovery-tools/

[5]Claypoole, T. F. (2014). Privacy and Social Media. *Business Law Today*, 1–4. http://www.jstor.org/stable/businesslawtoday.2014.01.05

[6]R. Mitson. (2022). Data security: Importance, measures and best practices (1er ed.). [En línea]. Disponible en: https://www.sherpany.com/en/resources/digital-transformation/cloud-computing/data-security-importance/

[7] Stiles, K. A. (1942). What Is the Scientific Method? *Bios*, *13*(1), 13–20. http://www.jstor.org/stable/4604621

8. Alumna y Directores

Jeon Jeong Paola. - Alumna de la carrera de Ing. en Sistemas Computacionales en ESCOM Boleta: 2020630193 Tel. (55) 4524 1679, email: habeet90@gmail.com

Firma: _____

Joel Omar Juárez Gambino. – Doctor en Ciencias de la Computación por el CIC, IPN. Sus áreas de estudio son: Procesamiento de Lenguaje Natural y Aprendizaje Automático. Departamento de Ciencias e Ingeniería de la Computación, ESCOM, Tel. 57296000 Ext. 52022, email: jjuarezg@ipn.mx

Firma: Omar Juárez Gambino

Eleazar Aguirre Anaya. - Doctor en comunicaciones y electrónica por ESIME, IPN. Sus áreas de estudio son: Seguridad en sistemas operativos y forense digital. Laboratorio de Ciberseguridad del CIC, Tel. 5526905660 Ext. 56607, email: eaguirre@cic.ipn.mx

Firma: _____