Captura de datos para análisis de la dinámica del tecleo de números para sistema operativo Android

Selene Nieto-Ruiz, Yonic A. Gómez-Sánchez, Asdrúbal López-Chau, Carlos A. Rojas

Ingeniería en Computación, Centro Universitario UAEM Zumpango, Zumpango, Estado de México, México

Resumen. El análisis de la dinámica del tecleo es una técnica usada para verificar la identidad de usuarios de sistemas informáticos. Actualmente, existen pocos conjuntos de datos públicamente disponibles de usuarios reales, y los que se encuentran corresponden a teclados tipo QUERTY. El teclado tipo numérico, que es usado en diversas aplicaciones importantes, ha recibido poca atención por parte de la comunidad científica. En este trabajo se presenta el desarrollo de un sistema para sistema operativo Android, cuya finalidad es la captura de los tiempos de tecleo usados comúnmente en biometría informática. Los datos generados por usuarios reales usando este sistema, se han publicado en la Internet para su libre descarga. En este artículo se presenta un resumen de los datos y se presentan trabajos futuros.

Palabras clave: dinámica de tecleo, biometría, Android.

1. Introducción

La biometría informática se refiere al conjunto de técnicas, principalmente estadísticas, que son aplicadas para verificar la autenticidad de individuos, tomando como base sus rasgos físicos o de conducta. La biometría informática se divide en estática y dinámica. La primera usa las características físicas de los sujetos, mientras que la segunda emplea sus rasgos conductuales. Entre los rasgos físicos más utilizados se encuentran las huellas dactilares y algunas partes del ojo como la retina o iris. Algunos ejemplos de características conductuales incluyen la voz, la forma de firmar, y el modo de caminar.

Recientemente, el grupo de hackers "*Chaos Computer Club*"[1] demostró que utilizar el tipo de biometría informática estática para proteger sistemas ya no es confiable. Ya que mediante simples fotografías tomadas con teléfonos inteligentes es posible copiar las huellas dactilares. Debido a lo anterior, la biometría informática dinámica cobra mayor importancia en diversas aplicaciones, y es actualmente usada cada vez más en sitios Web [2] y sistemas informáticos[3].

[1] http://www.ccc.de/en/

[2] Por ejemplo, https://mega.co.nz

[3] Ejemplo, Sistemas de Administración Tributaria en México

La dinámica del tecleo (*keystroke dynamics*)[13], que consiste en usar la forma o el ritmo en que un usuario escribe en un teclado para autenticar su identidad, es una de las principales técnicas de la biometría informática dinámica. Numerosos trabajos y proyectos han sido publicados recientemente [4], [13], ya sea proponiendo nuevos algoritmos [7] , implementando o aplicando la dinámica del tecleo en diferentes dispositivos. Generalmente, los estudios con dinámica de tecleo se han centrado en teclados tipo QUERTY [14]. Sin embargo, son pocos los trabajos que se enfocan en teclados tipo numérico. En la actualidad, son varias las aplicaciones en las que se utiliza este tipo de teclado, por ejemplo, cajeros automáticos, cajas fuertes, teléfonos [3] y para captura en hojas de cálculo.

En este artículo, se presenta la implementación completa de un sistema para la captura y almacenamiento de la dinámica del tecleo en dispositivos con sistema operativo (SO) Android, para teclados tipo numérico. El sistema propuesto, fue implementado y probado con 14 usuarios reales.

2. Biometría informática

El término biometría, proveniente del griego "*bios*"vida y "*metron*"medida, se refiere a sistemas de reconocimiento humano a través de la medición de sus rasgos físicos y/o de conducta [5] . El uso de técnicas para la identificación de individuos se remonta a tiempos antiguos. La medición de algunas partes del cuerpo fue usada primero, y luego la impresión de huellas dactilares [10]. A principios de la década 2000, otras características físicas que empezaron a utilizarse fueron el iris del ojo y los patrones vasculares de la retina, ya que se descubrió que son únicos para cada ser humano [12] .

Existe dos modos de emplear la biometría, para identificación o para verificación. En el primero, conocido como igualación 1 a N, un sistema identifica a un individuo de entre una población que se encuentra registrada. Para ello se realiza una búsqueda de coincidencia, basándose en algún rasgo particular. En el modo de verificación, se comprueba la identidad de un individuo, usando un patrón que ha sido previamente registrado del mismo; esto también es llamado igualación 1 a 1.

La biometría se clasifica en dos grandes tipos [9] :

Biometría Estática

Se caracteriza por considerar parámetros derivados de la medición directa de alguna característica fisiológica compleja de cuerpo humano. El adjetivo estática es debido a que, bajo condiciones normales, los valores de la característica medida no varían significativamente durante el ciclo de vida de un ser humano. Las principales aplicaciones y estudios se basan en los sistemas biométricos de huellas dactilares, geometría de la mano, iris, entre otras.

Biometría Dinámica

Los psicólogos han mostrado que los seres humanos somos predecibles en nuestros comportamientos al desempeñar tareas repetitivas o rutinarias. La biometría dinámica es la encargada de estudiar las características o rasgos de la conducta para verificar la identidad de las personas, a diferencia de la biometría

estática en la dinámica los rasgos o características varían en una misma persona a lo largo del tiempo.

En la siguiente sub-sección, se presenta una de las técnicas de biometría dinámica que se ha empleado cada vez más en aplicaciones recientes, y que se estudia en este artículo.

2.1. Dinámica del tecleo

Actualmente, es común para la mayoría de las personas contar con una computadora, teléfono inteligente o tableta electrónica. Una de las principales actividades que se realizan en estos dispositivos, es la escritura por medio del teclado tipo QUERTY, que es el predominante en este tipo de sistemas.

Otro tipo de teclado que es empleado en algunas aplicaciones, es el numérico, cuya apariencia es presentada en la Figura 1. Ejemplo de estas aplicaciones son calculadoras simples, pantallas para ingreso al sistema y pantallas para captura de números, como en la banca electrónica.

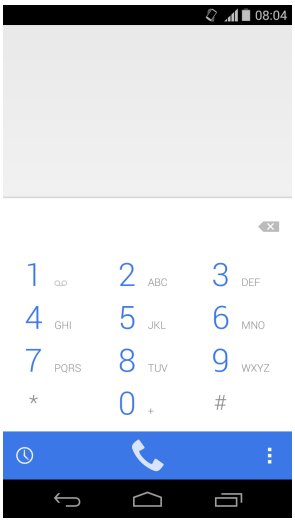

Fig. 1: Teclado tipo numérico de Android

- Tiempo de presión (*dwell-time*): Es el intervalo que transcurre cuando el usuario mantiene presionada y libera la misma tecla.
- Tiempo de cambio (*flight-time*)): Se considera el tiempo que transcurre cuando el usuario suelta una tecla y presiona la tecla siguiente.

A diferencia de las características físicas de una persona, que son inmutables a lo largo del tiempo (bajo condiciones normales, es decir, sin accidentes,

enfermedades deformadoras, etc.), la forma de teclear de una persona puede sufrir variaciones notables a lo largo de su vida. Existen distintas razones por lo que sucede esto, por ejemplo, el grado de atención al teclear, o estado de salud de la persona. Las características físicas del teclado utilizado (tamaño, separación de teclas, etc.) tienen también una influencia notable sobre el valor de *dwell-time* y *flight-time*.

3. Sistema Operativo Android

Android es un sistema operativo (S.O.) orientado a dispositivos móviles, y está basado en Linux [6]. Fue desarrollado en 2003, en Palo Alto, California, por Andy Rubin, Rich Miner, Nick Sears y Chris White. Desde 2005, Android es propiedad de Google.

El código fuente de Android sobrepasa los 10 millones de líneas, escritas en varios lenguajes de programación como C, C++ y Java. Este sistema operativo está compuesto por los siguientes componentes: aplicaciones, marco de aplicaciones, bibliotecas, máquina virtual de Java reducida (llamada Dalvik) y núcleo del sistema.

Al crear un proyecto nuevo en el entorno Android, automáticamente se generan los siguientes directorios [8]:

- src: Contiene en su totalidad el código fuente (Java) de la aplicación desarrollada.
- res: En este directorio se encuentran los recursos necesarios para generar la aplicación.
  - res/drawable: Almacena imágenes en diversos subdirectorios dependiendo de su resolución.
  - res/raw: Contiene los archivos que son de propósito general, excepto los formatos XML.
  - res/layout: Se encuentran los archivos que definen la interfaz gráfica, generalmente siempre en XML.
  - res/values: Guarda datos, tales como colores, cadenas de texto, estilos.
- gen: Se almacena un conjunto de archivos al momento de que el proyecto se compila, con el fin de dirigir los recursos de la aplicación.
- assets: Se encuentran los archivos indispensables para el correcto funcionamiento de la aplicación, por ejemplo, archivos de datos o de configuración.
- Archivo AndroidManifest.xml: Se considera el archivo más importante para la creación de una aplicación. Este se genera automáticamente y se encuentra definida la configuración del proyecto XML.

La creación de interfaces gráficas puede realizarse utilizando únicamente código, o combinando código fuente y XML.

4. Trabajos relacionados

Son variados los trabajos que tratan sobre verificación de la autenticidad de usuarios basada en la dinámica del tecleo o uso del dispositivo apuntador ratón.

De acuerdo a el artículo [1] , se presenta una nueva forma de sistema biométrico, basado en la dinámica del mouse o ratón. Se desarrolla una técnica para modelar características del comportamiento de los datos capturados con ayuda de redes neuronales artificiales. Esta investigación se centra en la extracción de características de comportamiento que son relacionadas con el usuario. Para ello se utiliza técnicas de estadísticas como lo es redes neuronales. Se muestra que se logra un FAR (por sus siglas en inglés False Acceptance Rate) de 2.4946 % y un False Rejection Rate (FRR) de 2.4624 %.
El FAR y FRR son métricas para determinar el desempeño de un sistema de detección de intrusiones.

- FAR: mide la probabilidad de que un impostor puede ser erróneamente aceptado por el sistema.
- FRR: mide la probabilidad de que un usuario genuino pueda ser rechazado por el sistema.

Para su realización se les pidió a participantes jugar un juego donde tenía que hacer clic lo más rápido posible en un cuadro en movimiento durante un periodo de tiempo fijo. Lo cual permitía recoger las coordenadas del ratón y calcular varias características tales como velocidad, aceleración, velocidad angular, curvatura, entre otros más. Para la autentificación del individuo se hace mediante la comparación de dicha información contra algunos umbrales.

En el artículo [11] se presenta un modelo para la autentificación de usuarios en función de pulsaciones de teclado en dispositivos móviles.

Este modelo captura el tiempo el tiempo necesario para que el usuario presione y suelte las teclas cuando este escribiendo, para que de esta manera se compare la información con el registro almacenado anteriormente. Entonces, si el resultado es el mismo al registrado, el usuario se considera autentico e idóneo para acceder al sistema de lo contrario es considerado como un impostor por consecuencia el sistema queda bloqueado. Con ayuda del método de distancia Euclidiana alcanzo un FAR de 12.97 % y un FRR de 2.25 %.

Cuando se introduce la contraseña el usuario ejecuta su propia característica a través del acto de empuje y soltar. Lo que permite la identificación y autentificación mediante su ritmo de escritura (velocidad, tiempo de pulsación de cada tecla y liberación de cada uno).

La realización de este tipo de sistemas permite el aumento de seguridad móvil con el mejoramiento de control de acceso.

5. Principal aportación

La principal aportación de este proyecto es generar un conjunto de datos con tiempos de presión y tiempos de cambio de usuarios reales, lo anterior con la finalidad de que estos datos sean usados en estudios de biometría dinámica del tecleo. A diferencia de otros conjuntos de datos similares, en los que se emplean teclados tipo QUERTY para los experimentos, en nuestra propuesta los tiempos registrados corresponden a teclado tipo numérico. De acuerdo a la investigación documental realizada previamente a la realización de este trabajo, no existe actualmente un conjunto de datos públicamente disponible como el que se presenta en este artículo.

6. Sistema desarrollado para la captura de datos de dinámica del tecleo

El teclado tipo QUERTY es uno de los más utilizados en dispositivos móviles actualmente. Sin embargo, todavía son varias las aplicaciones importantes que utilizan teclados de tipo numérico. Algunas de ellas son las cajas fuertes, teclados en cajeros automáticos y otros sistemas de control de acceso. Pocos trabajos han prestado atención a la biometría dinámica con este tipo de teclado. En este trabajo, se desarrolla una aplicación para SO Android [4], que permite capturar los tiempos de presión y de cambio. Estos son almacenados en una base de datos local, para su posterior análisis.

La interfaz gráfica de usuario (GUI) que se diseñó, es similar a las de acceso a sistemas informáticos, y se muestra en la Figura 2.

Fig. 2: Interfaz para captura de usuario y contraseña

[4] http://www.alchau.com/research/2015/keystrokeDinamycsAndroid.rar

Esta GUI tiene controles para que los usuarios ingresen un nombre de usuario y una contraseña. Para preservar el anonimato de los usuarios y al mismo tiempo para distinguir entre los datos generados por cada uno, se usaron números como identificadores para los usuarios. De esta forma, el primer usuario tiene como identificador el valor 1, el segundo el 2, y así sucesivamente.

Con la finalidad de que los datos generados por cada usuario puedan ser comparables con los otros, se decidió que el nombre del usuario y la contraseña fueran iguales en todos los casos. El nombre de usuario es **94255701**, mientras que la contraseña es **416850293**.

6.1. Arquitectura del sistema

El diagrama de clases del sistema implementado es mostrado en la Figura 3. La clase *Gestor de usuarios* controla el acceso y consultas a la base de datos local. Cada usuario es identificado con un número entero. La clase *Usuario* mantiene los datos de cada usuario, como el id y contraseña. Una de las clases más importantes es la denominada *EscucharTeclado*. Esta clase contiene los métodos para calcular los tiempos de presión y de cambio. La API (*Application Programming Interface*) de Android ofrece una serie de clases para facilitar la captura de eventos del teclado. Una combinación de los eventos *onKeyDown*() y *onKeyUp*() [2] fue usada para capturar los tiempos requeridos.

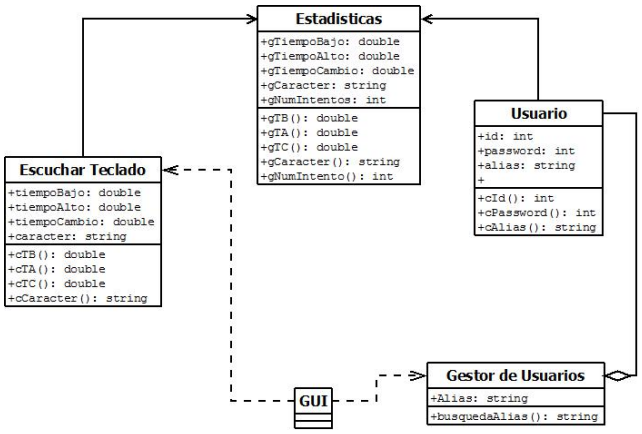

Fig. 3: Diagrama de clases del sistema implementado

7. Datos generados

El dispositivo en el que se probó el sistema, fue un teléfono inteligente modelo XT1058, con procesador ARMv7 (v71), patalla de 4.7 pulgadas y resolución de

720 × 1184 p. El sistema operativo es Android 4.4.4 .

Se solicitó a 14 personas, que introdujeran el nombre de usuario (94255701) y la contraseña (416850293) un total de 10 veces. Para los casos en los que un usuario se equivoca, el sistema elimina ese registro y solicita que se intente nuevamente. Cada intento exitoso fue almacenado en una base de datos local en el teléfono inteligente.

Un resumen de los tiempos de presión promedio al introducir es el nombre de usuario y la contraseña es mostrado en las Figuras 4 y 5, respectivamente. Los tiempos de cambio promedio, son presentados en las Figuras 6 y 7.

Los datos correspondientes a todos los intentos de los 14 usuarios, han sido publicados en el sitio http://www.alchau.com/research/2015/keystrokeData-1/data.zip, y se encuentran disponibles para su descarga.

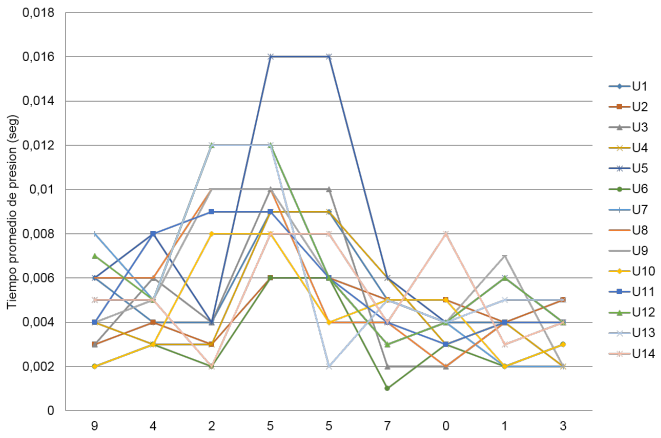

Fig. 4: Tiempos de presión promedio al introducir el nombre de usuario

8. Conclusiones y trabajo futuro

El análisis de la dinámica del tecleo ha sido estudiado ampliamente en la última década. En la literatura se reportan avances significativos sobre identificación de usuarios usando teclados tipo QUERTY. Sin embargo, de acuerdo a la investigación realizada en la presente investigación, no se ha prestado gran atención a estudios similares con teclados de tipo numérico.

En este artículo, se desarrolló una aplicación para sistema operativo Android, que permite capturar los tiempos de presión y de cambio en este tipo de teclado. Se realizaron pruebas con 14 usuarios reales, a quienes se solicitó capturar un mismo nombre de usuario y contraseña un total de 10 veces. Los datos generados para cada intento exitoso se han puesto disponibles públicamente en la Internet, para ser utilizados en futuros experimentos.

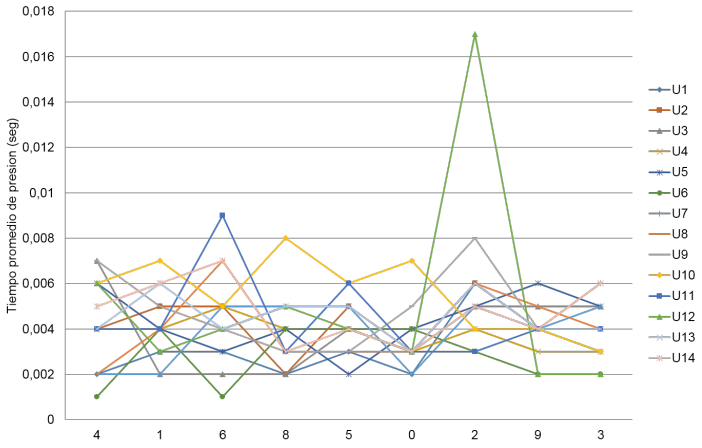

Fig. 5: Tiempos de presión promedio al introducir la contraseña

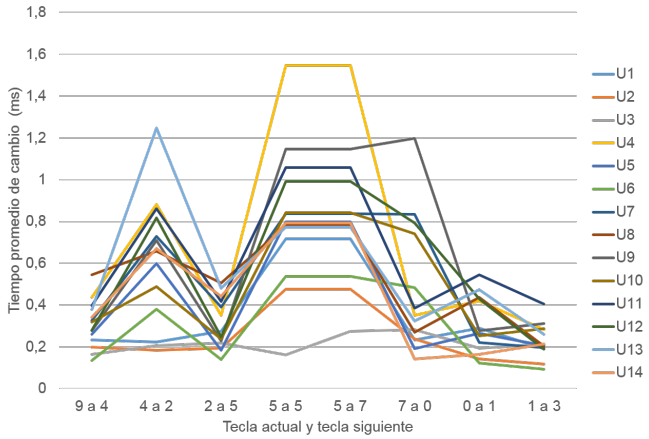

Fig. 6: Tiempos de cambio promedio al introducir el nombre de usuario

Como trabajo futuro, se tiene planeado probar algoritmos para verificar la identidad de usuarios basados en la dinámica de su tecleo. También se tiene considerado implementar este tipo de sistemas en dispositivos electrónicos, para ser utilizado con cerraduras electromecánicas. Por último, pero no menos importante, se tiene la intensión de desarrollar un algoritmo propio para identificación de impostores.

Agradecimientos. Los autores de este artículo agradecen a los revisores anónimos por sus valiosas observaciones para mejorar este trabajo. También agradecen a la Universidad Autónoma del Estado de México por su apoyo econónico, a través del proyecto de investigación 3790/2014/CID.

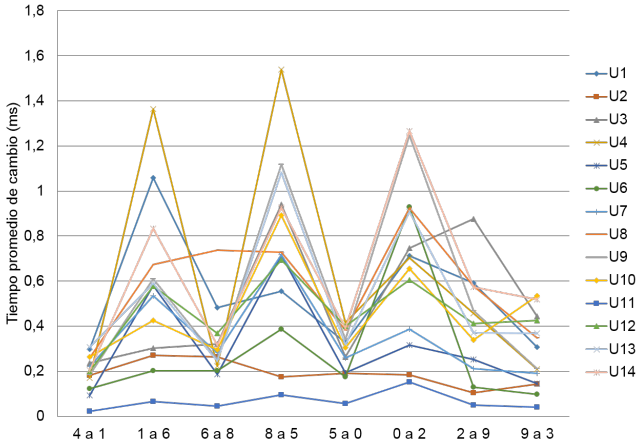

Fig. 7: Tiempos de cambio promedio al introducir la contraseña

Referencias

1. Ahmed, A.A.E., Traore, I.: A new biometric technology based on mouse dynamics. IEEE Transactions on Dependable and Secure Computing (2007)
2. Android: View.onkeylistener (Diciembre 2014), `http://developer.android.com/reference/android/view/View.OnKeyListener.html`
3. Antal, M., Szabó, L.Z., László, I.: Keystroke dynamics on android platform. ELSEVIER (2014)
4. Gaines, R., Lisowski, W., Press, S., Shapiro, N.: Authentication by keystroke timing: some prelimary results. Rand Corporation (1980)
5. Galván, G.I.: Sistema de autenticación de dispositivos móviles basado en biometría de comportamiento de tecleo (Junio 2007), `http://delta.cs.cinvestav.mx/~francisco/tesisIglesias.pdf`
6. Gironés, J.T.: El gran libro de Android. Barcelona (2012)
7. Obaidat, M.S., Macchiarolo, D.T.: An on-line neural network system for computer access security. IEEE Transactions on Industrial Electronics (1993)
8. Soriano, J.E.A.: Android: Programación de dispositivos móviles a través de ejemplos. Barcelona (2012)
9. UNAM: Clasificación de los sistemas biométricos (Octubre 2014), `http://redyseguridad.fi-p.unam.mx/proyectos/biometria/clasificacionsistemas/clasificaciontipo.html`
10. UNAM: Fundamentos de biometría (Noviembre 2014), `http://redyseguridad.fi-p.unam.mx/proyectos/biometria/fundamentos/antecedentes.html`
11. Urtiga, E., Moreno, E.: Keystroke - based biometric authentication in mobile devices. IEEE Latin America Transactions (2011)
12. Yampolskiy, R.V., Govindaraju, V.: Behavioural biometric: a survey and classification. Int. J. Biometrics (2008)
13. Yu, E., Cho, S.: Keystroke dynamics identity verification - its problems and practical solutions. ELSEVIER (2004)
14. Zhong, Y., Deng, Y., Jain, A.K.: Keystroke dynamics for user authentication