Generalización del modelado de cadencias de tecleo con contextos finitos para su utilización en ataques de presentación y canal lateral

Doctorado en Ciencias Informáticas
Universidad Nacional de La Plata (UNLP)

Autor: Nahuel González[1]
Fecha de Defensa: 07/06/2022

Director: Dr. Jorge S. Ierache[1]
Codirector: Dr. Waldo Hasperué[2]
Asesor científico: Dr. Enrique P. Calot[1]

[1] Laboratorio de Sistemas de Información Avanzados (LSIA)
Facultad de Ingeniería, Universidad de Buenos Aires
Ciudad Autónoma de Buenos Aires, Argentina

[2] Instituto de Investigación en Informática (III-LIDI)
Facultad de Informática, Universidad Nacional de La Plata
La Plata, Provincia de Buenos Aires, Argentina

Resumen

Los patrones individuales de escritura en un teclado, denominados cadencias de tecleo, configuran un atributo biométrico comportamental que puede utilizarse como segundo factor de autenticación (2FA) para verificar en forma transparente la identidad de los usuarios. Los sistemas de autenticación basados en cadencias de tecleo son vulnerables a ataques de presentación, que utilizan muestras sintetizadas para imitar el comportamiento del usuario legítimo; adicionalmente, cuando otros sistemas filtran los tiempos de escritura, estos pueden ser recolectados en un ataque por canal lateral para luego identificar el texto ingresado o potenciar un ulterior ataque por fuerza bruta, reduciendo los candidatos posibles. A pesar de estos riesgos, los métodos de autenticación basados en cadencias de tecleo que se han propuesto en la literatura han sido generalmente evaluados bajo un modelo de esfuerzo cero, también llamado de impostores no entrenados, que subestima o ignora el riesgo de los ataques mencionados. En esta tesis se propone un sistema de detección de vida para mitigar los riesgos de un ataque de presentación y una familia de estrategias de síntesis de muestras artificiales que son utilizadas como adversarios durante el entrenamiento. También se proponen nuevas distancias basadas en los histogramas empíricos del perfil del usuario legítimo, que presentan mejor capacidad de discriminación que las distancias clásicas. El modelo de detección de vida logra tasas de falsos positivos menores al 2% contra muestras sintetizadas con perfiles interusuario y menores al 15% incluso en el caso extremo en que el atacante cuenta con el perfil biométrico completo del usuario legítimo; estos valores deben contrastarse con el 98%+ de falsos positivos alcanzado sin detección de vida. Adicionalmente, una modificación del modelo propuesto permite abordar el problema de identificación del texto ingresado utilizando sólo los tiempos de escritura, potenciando los ataques de canal lateral y de fuerza bruta ya existentes.

Palabras clave: seguridad informática, biometría comportamental, cadencias de tecleo, detección de vida, ataques de presentación, ataques por canal lateral.

1. Introducción

Las sutiles variaciones en la forma en que distintas personas teclean son suficientes para revelar su identidad. Hace cuarenta años, Gaines y colaboradores (Gaines 1980), pioneros del análisis de cadencias de tecleo, reconocieron la utilidad de este fenómeno para la autenticación de usuarios. Utilizando sólo los tiempos entre eventos de presión y liberación de teclas, es posible construir un segundo factor de autenticación (2FA) tanto para endurecer claves de usuario (Monrose 1990) como para verificar la identidad en forma continua (Bours 2009). Más recientemente, el análisis de cadencias de tecleo también ha encontrado usos fuera del dominio de la seguridad informática, como descubrir ciertas características fisiológicas o impedimentos clínicos del usuario (Milne 2018), e incluso determinar en forma aproximada las variaciones de su estado emocional mientras escribe (Epp 2011).

El desafío que hoy enfrenta el análisis de cadencias de tecleo consiste en prevenir modelos de ataque sofisticados (Rahman 2011) en donde el impostor malintencionado cuenta con acceso parcial o total al perfil biométrico del usuario y realiza un esfuerzo considerable para imitar al usuario legítimo (Gonzalez2021b). O, también, en explorar el uso de las cadencias de tecleo para potenciar otro tipo de ataques como los de canal lateral (Gonzalez 2021d) o fuerza bruta (Song 2001). Dentro de esta línea, Monaco y colaboradores (Monaco 2019) han demostrado que es posible reconstruir la consulta escrita en un motor de búsqueda empleando los intervalos entre teclas sucesivas, inferidos en base a los tiempos de arribo de paquetes de red.

1.1 Contexto

El análisis de cadencias de tecleo pertenece al ámbito de la *seguridad informática* ya que permite, como único factor o en combinación con otros, la autenticación y verificación de identidad. Los métodos de autenticación pueden dividirse en aquellos basados en el conocimiento, basados en la posesión, y *biométricos*; el tema que nos compete pertenece a estos últimos, que se dividen en convencionales, cuando estudian las características fisiológicas del usuario, y *comportamentales* cuando analizan sus patrones de comportamiento. El análisis de cadencias de tecleo también pertenece al ámbito de la *interacción hombre-máquina* ya que la fuente de información es un dispositivo de entrada: el teclado, o a veces también un dispositivo móvil.

Dentro de la *seguridad informática* se estudian los tipos de *ataque* a los que puede ser sometido un sistema de información. Estos pueden ser *directos* o *indirectos*. Los primeros son aquellos que demandan una interacción con el sistema; los segundos, aquellos que se realizan con información filtrada sin interactuar con el sistema.

1.2 Definiciones

Un *ataque de presentación* es un tipo de ataque directo, en el cuál un actor malintencionado muestra o presenta al sistema de autenticación las credenciales biométricas del usuario legítimo, ya sea copiándolas o imitándolas, con la intención de hacerse pasar por él y lograr obtener acceso privilegiado. La prevención de ataques de presentación requiere robustecer al sistema considerado con un modelo de *detección de vida*, capaz de distinguir entre el comportamiento del usuario legítimo y el de un atacante que intenta imitarlo. En particular, los sistemas de autenticación basados en cadencias de tecleo son vulnerables a ataques de presentación con *muestras sintéticas*, es decir aquellas construidas artificialmente por el atacante (en contraste con las muestras legítimas, escritas naturalmente por el usuario) con la ayuda de una herramientas de síntesis y, posiblemente, conocimiento parcial o total del comportamiento del usuario legítimo o de su perfil biométrico.

Un *ataque por canal lateral* es todo aquel que explota la información extrínseca filtrada inadvertidamente por una cierta implementación particular de un sistema. Es un tipo de ataque indirecto, y se caracteriza por tener como objetivo dicha implementación particular en lugar del algoritmo, método, o sistema implementado. Por ejemplo, en (Song 2001) se detalla cómo una implementación del protocolo SSH que transfiera por red y sin demora los caracteres escritos filtra

inadvertidamente los tiempos de escritura y cómo estos pueden emplearse para potenciar un ataque por fuerza bruta, reduciendo en dos órdenes de magnitud el tiempo y la cantidad de intentos requeridos para adivinar la clave del usuario.

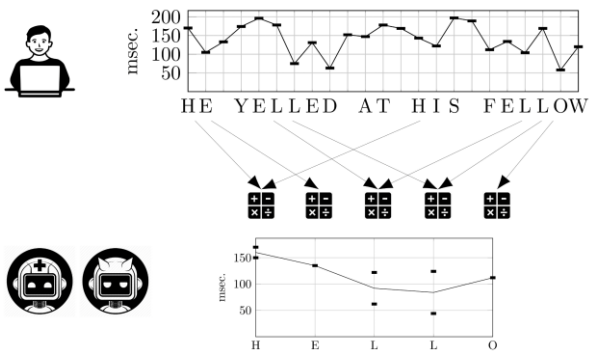

Figura 1 - *Modelado por contextos finitos*

El *modelado por contextos finitos*, que se ilustra en la Figura 1, es un método general de síntesis de cadencias de tecleo que emplea un conjunto de muestras de escritura en texto libre del usuario legítimo para reconstruir con fidelidad los atributos temporales, y quizás otros, que corresponden a cualquier secuencia de caracteres, así no existan observaciones previas de esta última entre las muestras de entrenamiento (González 2015b). Los métodos de síntesis de muestras propuestos en esta tesis (ver Sección 4) se basan en el modelado por contextos finitos.

Un *perfil intrausuario* es aquel que sólo incluye muestras del usuario a imitar. En contraste, un *perfil interusuario* es aquel que se compone de muestras de varios usuarios de una población genérica.

1.3 Motivación

Las técnicas de análisis y modelado de cadencias de tecleo han sido y continúan siendo ampliamente estudiadas por su interés intrínseco y para su aplicación como un segundo factor en la autenticación de usuarios. En particular, estos métodos tienen la ventaja de ser transparentes, en el sentido de que no demandan ninguna acción ulterior al usuario legítimo más que escribir naturalmente, y no le imponen demoras, confirmaciones, u otras tareas como los biométricos convencionales y demás esquemas de autenticación multifactor.

Sin embargo, las vulnerabilidades de los sistemas de autenticación basados en cadencias de tecleo ante ataques de presentación y falsificaciones sintéticas han recibido mucha menos atención que en el caso de los biométricos tradicionales, a pesar de que las debilidades a las que los expone no es inferior. Hoy en día y en vista de la ubicuidad de las filtraciones de datos, no es descabellado imaginar que un atacante cuenta con acceso, al menos parcial, a las muestras biométricas del usuario. Por este motivo es necesario investigar métodos sofisticados de ataque y las contramedidas de defensa necesarias para mitigarlos. Asimismo, el análisis y modelado de cadencias de tecleo es promisorio como vector de amplificación de otro tipo de ataques, de canal lateral, al permitir la filtración de información sobre el texto escrito en la forma de valores temporales correlacionados con este.

1.4 Marco

La presente tesis fue radicada en el Laboratorio de Sistemas de Información Avanzados (LSIA) de la Facultad de Ingeniería de la Universidad de Buenos Aires, y se enmarca en el proyecto de desarrollo estratégico PDE-44-2019, *Reconocimiento de Patrones de Tecleo en Ambientes Web*,

2. Definición del problema

El problema que ha guiado los experimentos de esta tesis es la creación de contramedidas de defensa eficaces ante ataques de presentación que utilicen muestras sintetizadas artificialmente, en base a un perfil intrausuario total o parcial. En el camino hacia la solución se estudiaron las distribuciones temporales subyacentes y se propusieron estrategias de síntesis de muestras capaces de engañar a los actuales sistemas de detección de vida con mayor frecuencia que los empleados usualmente para la evaluación. Por último, se descubrió que una modificación de los mismos métodos permite potenciar los ataques por canal lateral para identificación del texto ingresado.

2.1 Objetivos

Identificar las distribuciones subyacentes y los patrones de comportamiento que generan la cadencia de tecleo en texto libre.

Proponer estrategias de síntesis de muestras artificiales para su uso en ataques de presentación, que engañen a los actuales sistemas de autenticación con frecuencia suficiente como para constituir una amenaza.

Construir un sistema de detección de vida que sirva como contramedida de defensa ante las anteriores y otras estrategias de síntesis de muestras artificiales del estado del arte.

Explorar las técnicas anteriores para su uso en ataques de canal lateral, incrementando la capacidad de identificar un texto en base a los tiempos de escritura.

3 - Contribuciones

La presente tesis contribuye a incrementar la seguridad de los métodos de autenticación por medio de cadencias de tecleo, robusteciéndolos frente a ataques de presentación con muestras sintetizadas, a la vez que potencia los ataques por canal lateral a través la reconstrucción del texto ingresado en base a los tiempos de escritura.

3.1 - Métodos

Los métodos propuestos se describen en detalle en la Sección 4.

3.2 - Herramienta

Se creó una herramienta integrada que implementa los métodos propuestos en esta tesis. La herramienta y el código fuente fueron puestas a disposición del público en el repositorio del Laboratorio de Sistemas de Información Avanzados y en la publicación con revisión de pares (Gonzalez 2023a), en Software Impacts, Elsevier, que acompaña al artículo (Gonzalez 2021b).

3.3 - Producción científica

La producción científica derivada de esta tesis incluye dos publicaciones en revistas internacionales con revisión de pares: *On the Shape of Timing Distributions in Free- Text Keystroke Dynamics Profiles* (Gonzalez 2021a) en Heliyon, Elsevier, revista de primer cuartil (Q1) según Scimago Journal Ranking; y *Towards Liveness Detection in Keystroke Dynamics: Revealing Synthetic Forgeries* (Gonzalez 2021b) en Systems and Soft Computing, que acompaña a Applied Soft Computing, revista de primer cuartil (Q1) según Scimago Journal Ranking. Dentro del mismo tópico y como resultado de la investigación, dos artículos adicionales en revistas internacionales con revisión de pares fueron publicados posteriormente a la fecha de aprobación de la tesis: *KSDSLD—A tool for keystroke dynamics synthesis & liveness detection* (Gonzalez 2023a) en Software Impacts, Elsevier, y *Dataset of Human- written and Synthesized Samples of Keystroke Dynamics Features for Free-text Input* (Gonzalez 2023b) en Data in Brief, Elsevier. La publicación *Exploring Internal Correlations in Timing Features of Keystroke Dynamics at Word Boundaries and Their Usage for Authentication and Identification* (Gonzalez 2020b) en Computer Science–CACIC 2020: Revised & Selected Papers pertenece a series y capítulos de libro en idioma inglés.

Se realizaron tres presentaciones en congresos internacionales: (Gonzalez 2015b), (Gonzalez 2016), y (Gonzalez 2021d); mientras que las presentaciones en congresos nacionales y regionales fueron tres: (Gonzalez 2020a), (Gonzalez 2021c), y (Gonzalez 2022). Adicionalmente, se realizaron en colaboración seis presentaciones, dos como autor principal y cuatro como coautor, de avances de líneas de investigación en el Workshop de Investigadores de Ciencias de la Computación (WICC): (Gonzalez 2015a), (Calot 2015), (Calot 2016), (Ierache 2019), (Concilio 2020), y (Gonzalez 2021e).

3.4 - Transferencia Tecnológica

En el ámbito nacional, se ha contribuído con entidades del Ministerio de Seguridad, entre ellos la Policía de Seguridad Aeroportuaria (PSA), facilitando el uso experimental de la herramienta solamente con fines exploratorios, para su empleo en la identificación de personal para auditoría forense. Adicionalmente, se dictó un seminario de la temática de la tesis al personal de la fuerza.

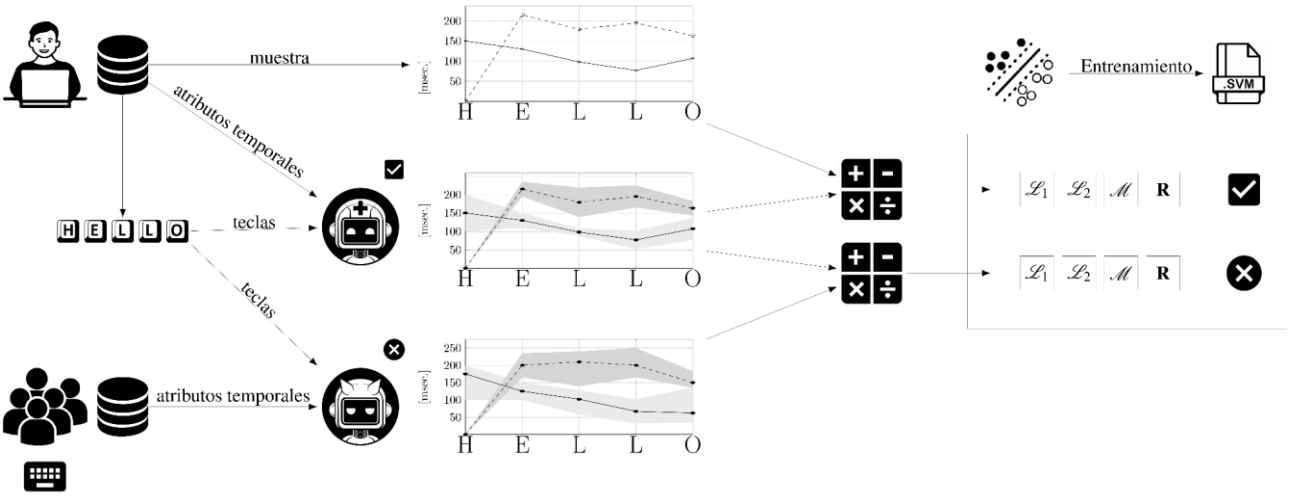

Figura 2 - *Entrenamiento del modelo de detección de vida con muestras legítimas y sintetizadas.*

En el ámbito internacional se llevó a cabo un proceso de transferencia tecnológica en contacto con empresas privadas del ámbito de la seguridad informática, con el objetivo de robustecer productos existentes para la autenticación con cadencias de tecleo y aplicarlos a la seguridad bancaria.

3.5 - Conjuntos de datos

Los conjuntos de datos de entrenamiento, evaluación, y resultados para los tres experimentos principales de esta tesis fueron puestos a disposición de la comunidad de investigadores en forma libre y gratuita, en los repositorios IEEE DataPort y Mendeley Data. Los conjuntos de datos del experimento sobre distribuciones subyacentes pueden encontrarse en (Gonzalez 2021l) y (Gonzalez 2021m), los del experimento sobre síntesis de muestras y detección de vida en (Gonzalez 2021h) y (Gonzalez 2021i), y los del experimento sobre identificación del texto ingresado en (Gonzalez 2021j) y (Gonzalez 2021k). Un conjunto de datos extendido que subsume los dos anteriores ha sido descrito en la publicación con revisión de pares (Gonzalez 2023b), que acompaña al artículo (Gonzalez 2021b).

Adicionalmente, se han hecho públicos en los repositorios IEEE DataPort y Mendeley Data los conjuntos de datos de entrenamiento, evaluación, y resultados de los siguientes artículos que forman parte de esta tesis: para (Gonzalez 2021c) en (Gonzalez 2021f) y (Gonzalez 2021g), para (Gonzalez 2020b) en (Gonzalez 2021n).

3.6 - Honores y menciones

En el marco de esta tesis, el artículo *Exploración de correlaciones internas de los parámetros temporales generados en dinámicas de tecleo* (Gonzalez 2020a), presentado en el XXVI Congreso Argentino de Ciencias de la Computación (CACIC), año 2020, ha obtenido el premio al mejor trabajo en el IX Workshop de Seguridad Informática (WSI). Asimismo, el artículo *Un método de ensamble basado en subsecuencias a nivel de palabras para la autenticación de usuarios con cadencias de tecleo en textos libres* (Gonzalez 2021c), presentado en el XXVII Congreso Argentino de Ciencias de la Computación (CACIC), año 2021, ha obtenido el premio a la mejor presentación en el X Workshop de Seguridad Informática.

4 - Métodos propuestos

4.1 - Distancias basadas en histogramas empíricos

El experimento sobre distribuciones empíricas de los tiempos de retención y latencia entre teclas en la escritura en texto libre (Gonzalez 2021a) demostró que los modelos estadísticos paramétricos no son suficientes para capturar las variaciones en la forma de estas distribuciones. En consecuencia, se propuso una familia de

distancias basadas en los histogramas empíricos del perfil biométrico del usuario, utilizando la función inversa de la distribución acumulada de probabilidad como contraparte de las distancias clásicas de Manhattan, euclídea, Canberra, y de Minkowski. Las distancias propuestas demostraron ser más sensibles que las clásicas para la verificación de identidad y la detección de muestras sintéticas (Gonzalez 2021b), a la vez que proporcionaron mayor ganancia de información a los clasificadores empleados para la detección de vida e identificación del texto ingresado.

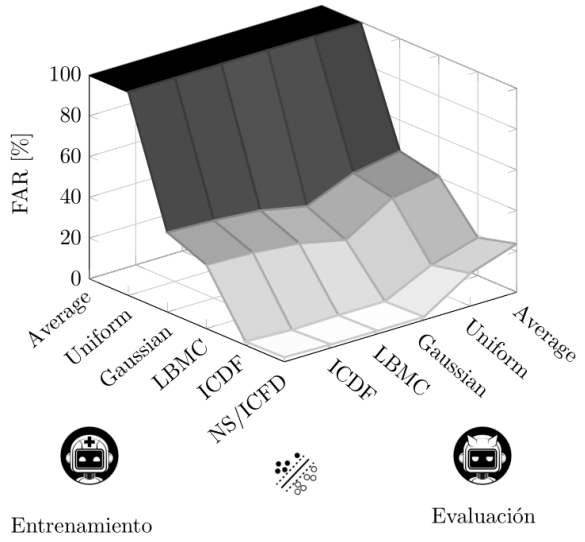

Figura 3 - *Tasas de falsos positivos de métodos de síntesis propuestos, al ser utilizados para entrenar y para atacar al modelo de detección de vida.*

4.2 - Métodos de síntesis basados en histogramas empíricos

La efectividad de las distancias basadas en histogramas empíricos para discriminar entre usuarios legítimos e impostores motivó la introducción de una familia de métodos de síntesis de muestras de cadencias de tecleo también basados en histogramas empíricos. El experimento sobre síntesis de muestras y detección de vida (Gonzalez 2021b) reveló que estos métodos, siempre que se empleen contextos de orden elevado que consideren las correlaciones internas a nivel de palabras (Gonzalez 2020a, Gonzalez 2020b), son capaces de engañar a un sistema de verificación de identidad con mayor frecuencia que otros métodos del estado del arte.

4.3 - Modelo de detección de vida con adversarios sintéticos

Se propuso un modelo de detección de vida, capaz de discriminar entre muestras auténticas y sintetizadas, que emplea los métodos de síntesis basados en histogramas empíricos y otros del estado de arte como adversarios para el entrenamiento. El experimento sobre síntesis de muestras y detección de vida (Gonzalez 2021b) demostró que un modelo tal es capaz de neutralizar un ataque de presentación con muestras sintéticas con muy bajas tasas de falsos positivos (< 2%) si el atacante no cuenta con demasiada información sobre el usuario legítimo, y con tasas aceptables (< 15%) incluso si el atacante cuenta con el perfil biométrico completo del usuario legítimo. La tasa de falsos negativos en todos los casos es menor al 5%. Al utilizar atributos derivados basados en el grado de desorden local (Gonzalez 2022) se logró mejorar el rendimiento gradualmente.

4.4 - Identificación de texto ingresado en base a tiempos de escritura

Dada una muestra de tiempos de escritura que carezca de la secuencia de teclas correspondiente, como por ejemplo la resultante de una ataque por canal lateral, es posible reconstruir el texto ingresado con alta probabilidad si el texto no es demasiado extenso (Monaco 2019). Un método para tal fin pero capaz de abordar muestras más largas (en el orden de los 100 caracteres), basado en el anterior modelo de detección de vida e ilustrado en la Figura 4, fue propuesto y evaluado como parte de esta tesis (Gonzalez2021d), alcanzando tasas de error muy bajas (< 2%) pero con la limitación de requerir una lista de candidatos parciales.

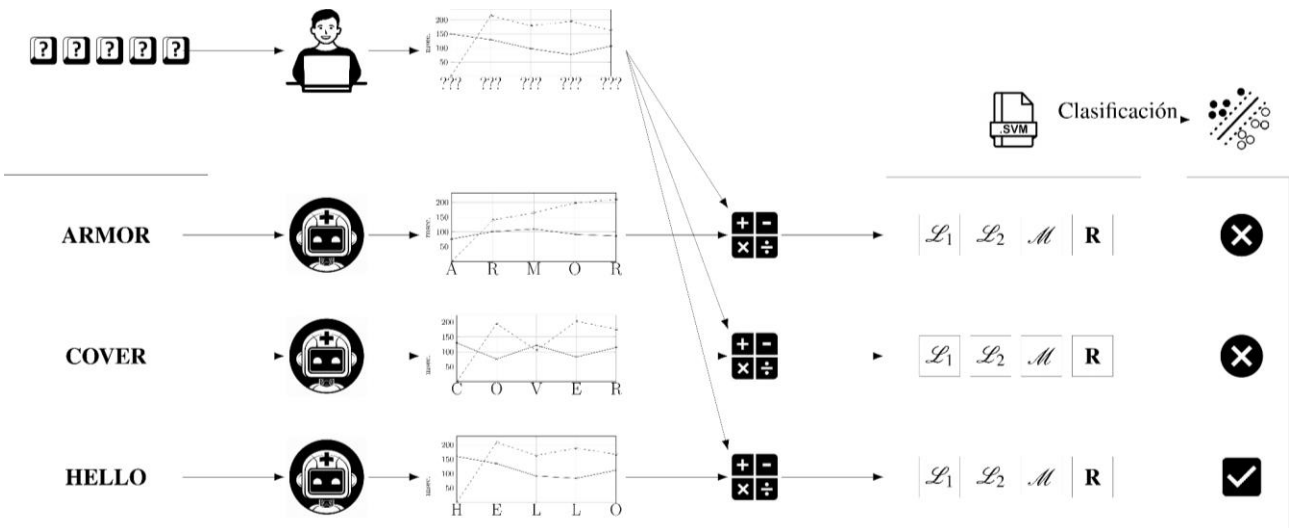

Figura 4 - *Identificación del texto ingresado en base a los tiempos de escritura*

5. Síntesis de los resultados

La conclusión principal de esta tesis consiste en haber establecido la validez de la hipótesis de trabajo; esto es, que *ninguna distribución suave es adecuada para modelar en general todos los atributos temporales que caracterizan la cadencia de tecleo; sólo las distribuciones empíricas del perfil intra-usuario pueden capturar con precisión su comportamiento característico, y este fenómeno puede ser capitalizado tanto para generar muestras sintéticas que engañan a los actuales sistemas de autenticación como para construir medidas de defensa eficaces contra ataques de presentación, al distinguir la escritura del usuario humano legítimo de una muestra construida artificialmente.*

Adicionalmente, cada experimento arrojó una conclusión particular dentro de su alcance. El experimento sobre distribuciones subyacentes nos enseñó que *la distribución log-logística es una clara ganadora entre todos los candidatos para ajustar los histogramas de tiempo, tanto de retención como de latencia, pero las tasas de rechazo de hipótesis y los méritos relativos muestran que un enfoque que considere los histogramas empíricos en su individualidad es preferible.*

De esta forma se motivaron y justificaron tanto las distancias como los métodos de síntesis basados en histogramas empíricos. Al evaluar su rendimiento en un ataque de presentación, se estableció que *las estrategias de síntesis basadas en histogramas empíricos alcanzan un mejor rendimiento que aquellas basadas en distribuciones suaves al intentar engañar a un sistema de autenticación.*

Pero también se estableció que, al utilizarlas como adversarios en un esquema de detección de vida, resulta que *las distancias basadas en histogramas empíricos alcanzan un mejor rendimiento que las tradicionales*.

Finalmente, al evaluar una extensión del método anterior se descubrió que *utilizando sólo atributos temporales es posible identificar el texto ingresado dentro de una lista de candidatos de tamaño mediano, alcanzando tasas de error muy bajas y competitivas con el estado del arte aunque tratemos con textos y listas de candidatos más largas.*

5.1 Resultados cuantitativos

Para la evaluación de los métodos propuestos, se utilizaron tres conjunto de datos (LSIA, KM, y PROSODY, dividido en GAY, GUN, y REVIEW) de gran extensión, públicamente accesibles, y que han sido empleados en estudios previos sobre cadencias de tecleo. La figuras 5 y 6 muestran las tasas de falsos positivos que cada método de síntesis alcanza contra el sistema de detección de vida, para cada conjunto de datos, con perfiles inter- e intrausuario, y sus intervalos de confianza del 95%.

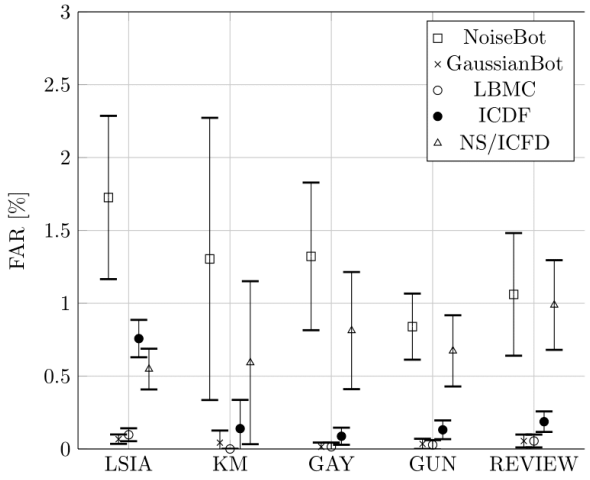

Figura 5 - *Tasas de falsos positivos para distintos métodos de síntesis de muestras, CI 95%, perfil interusuario.*

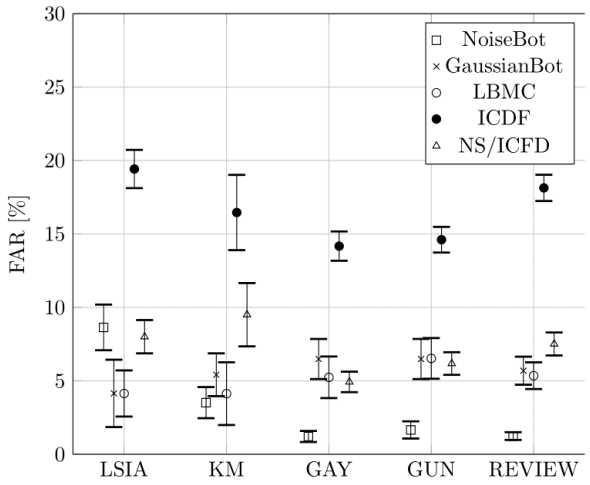

Figura 6 - *Tasas de falsos positivos para distintos métodos de síntesis de muestras, CI 95%, perfil intrausuario.*

6. Conclusiones

El resultado principal de esta tesis es un modelo de detección de vida que defiendea los sistemas de autenticación continua con cadencia de tecleo frente a ataques de presentación. Basado en un modelo de aprendizaje automático entrenado con muestras sínteticas, éste alcanza una tasa de falsos positivos menor al 2% contra muestras sintetizadas con perfiles interusuario y menores al 15% incluso en el caso extremo en que el atacante cuenta con el perfil biométrico completo del usuario legítimo; estos valores deben contrastarse con el 98%+ de falsos positivos alcanzado sin detección de vida. Para ser utilizados como adversarios, se propusieron diversos métodos de síntesis capaces de inducir tasas elevadas de falsos positivos en sistemas convencionales, y cuyas características principales son el empleo de modelos de orden elevado e histogramas empíricos recabados del perfil biométrico del usuario. Adicionalmente, se descubrió que una modificación del método de detección de vida permite reconstruir el texto ingresado en base a los tiempos de escritura, extendiendo el largo de las muestras tratables con los métodos existentes y potenciando los ataques por canal lateral que obtengan tales tiempos.

Se ha revelado la vulnerabilidad inherente en los sistemas de autenticación continua por cadencias de tecleo cuyo diseño no tiene en consideración los recursos a los que puede acceder un atacante sofisticado, justificando la necesidad de evaluar su rendimiento con modelos de ataque realistas y de incluir modelos de detección de vida como parte del proceso de verificación. Por otra parte, se ha mostrado que estos últimos hacen factible reducir la eficacia de un ataque de presentación con muestras sintéticas en un orden de magnitud o más, restaurando la confianza en la verificación de identidad con cadencias de tecleo en textos libres.

Los resultados de esta tesis, tanto para el robustecimiento de sistemas de autenticación existentes como para la auditoría forense, fueron validados en el transcurso del proceso de transferencia tecnológica con entidades del sector público nacional y privado internacional.

Como producción científica se obtuvieron dos publicaciones internacionales en revistas de primer cuartil (Q1), dos publicaciones suplementarias en revistas internacionales, una en series y capítulos de libro en inglés, tres presentaciones en congresos internacionales, tres en congresos nacionales y regionales, y seis presentaciones en workshops nacionales. Adicionalmente, se publicaron seis conjuntos de datos relacionados con los experimentos de esta tesis.

6.1 Futuras líneas de investigación

Durante el transcurso de esta tesis, los resultados han abierto diversos interrogantes. El hilo conductor de todos ellos puede

resumirse como la búsqueda de modelos, cuantitativos y explicativos, que den cuenta de las diferencias entre una muestra escrita por un usuario humano y otra sintetizada, para que el método de detección de vida explote en su tarea de discriminarlas. Lo que, en vista de los recientes desarrollos con redes neuronales adversarias generativas y *deepfakes*, plantea la crucial pregunta: *¿Es factible sintetizar muestras de cadencias de tecleo indistinguibles de sus contrapartes reales*? Intentar contestarla abre, sin duda, la más relevante y promisoria futura línea de investigación que puede derivarse de los resultados de esta tesis.

Referencias

(Bours 2009) Patrick Bours y Hafez Barghouthi. "*Continuous authentication using biometric keystroke dynamics*". En: The Norwegian Information Security Conference (NISK). Vol. 2009. 2009.

(Calot 2015) Enrique P. Calot y col. "*Líneas de investigación del Laboratorio de Sistemas de Información Avanzados*". En: Proceedings del XVII Workshop de Investigadores en Ciencias de la Computación (WICC). Salta, Argentina, jun. de 2015, pág. 4. ISBN: 978-987-633-134-0. URL: hdl.handle.net/10915/46107

(Calot 2016) Enrique P. Calot y col. "*Avances en educción de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro-computadora*". En: Proceedings del XVIII Workshop de Investigadores en Ciencias de la Computación (WICC). Entre Ríos, Argentina, jun. de 2016, págs. 872-876. ISBN: 978-950-698-377-2.

(Concilio 2020) Germán Concilio y col. "*Avances en reconocimiento de patrones de tecleo para la identificación de personas en ambientes web*". En: XXII Workshop de Investigadores en Ciencias de la Computación (WICC). El Calafate, Santa Cruz, Argentina, 2020, págs. 842-846. ISBN: 978-987-3714-82-5.

(Epp 2011) Clayton Epp, Michael Lippold y Regan L Mandryk. "*Identifying emotional states using keystroke dynamics*". En: Proceedings of the sigchi conference on human factors in computing systems. 2011, págs. 715-724

(Gaines 1980) R Stockton Gaines, William Lisowski, S James Press y Norman Shapiro. "*Authentication by keystroke timing: Some preliminary results"*. Inf. téc. Rand Corp Santa Monica CA, 1980.

(Gonzalez 2015a) Nahuel González y col. "Educción de dinámica de tecleo centrado en el contexto emocional de un individuo". En: Proceedings del XVII Workshop de Investigadores en Ciencias de la Computación (WICC). Salta, Argentina, jun. de 2015, pág. 5. ISBN: 978-987-633-134-0. URL: hdl.handle.net/10915/4626

(Gonzalez 2015b) Nahuel González y Enrique P. Calot. "*Finite Context Modeling of Keystroke Dynamics in Free Text*". En: Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the, IEEE. Septiembre 2015. ISBN: 978-3-88579-639-8. doi: 10.1109/BIOSIG.2015.7314606.

(Gonzalez 2016) Nahuel González, Enrique P. Calot y Jorge S. Ierache. "*A replication of two free-text keystroke dynamics experiments under harsher conditions*". En: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE. ISBN: 978-1-50900-780-6. doi: 10.1109/BIOSIG.2016.7736905.

(Gonzalez 2020a) Nahuel González, Germán Concilio, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué "*Exploración de correlaciones internas de los parámetros temporales generados en dinámicas de tecleo*". En: XXVI Congreso Argentino de Ciencias de la Computación (CACIC), 2020, págs. 726-735. ISBN: 978-987-4417-90-9.

(Gonzalez 2020b) Nahuel González, Germán Concilio, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*Exploring Internal Correlations in Timing Features of Keystroke Dynamics at Word Boundaries and Their Usage for Authentication and Identification*". En: Computer Science–CACIC 2020: 26th Argentine Congress, CACIC 2020, San Justo, Buenos Aires, Argentina, October 5–9, 2020, Revised Selected Papers. Vol. 1. Communications in Computer and Information Science, Springer Nature, págs. 321-332. ISBN: 978-3-030-75835-6. doi: 10.1007/978-3-030-75836-3_22.

(Gonzalez 2021a) Nahuel González, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*On the shape of timing distributions in free-text keystroke dynamics profiles*". En: Heliyon 7.11 (2021), Elsevier. ISSN: 2405-8440. doi: 10.1016/j.heliyon.2021.e08413.

(Gonzalez 2021b) Nahuel González, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*Towards liveness detection in keystroke dynamics: Revealing synthetic forgeries*". En: Systems and Soft Computing (2022), Elsevier. ISSN: 2772-9419. doi: 10.1016/j.sasc.2022.200037.

(Gonzalez 2021c) Nahuel González, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*Un método de ensamble basado en subsecuencias a nivel de palabras para la autenticación de usuarios con cadencias de tecleo en textos libres*". En: XXVII Congreso Argentino de Ciencias de la Computación (CACIC), 2021, págs. 685-694. ISBN: 978-987-633-574-4.

(Gonzalez 2021d) Nahuel González, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*The Reverse Problem of Keystroke Dynamics: Guessing Typed Text with Keystroke Timings Only*". En: 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2021, IEEE. ISBN: 978-1-6654-4231-2. doi: 10.1109/ICECET52533.2021.9698782.

(Gonzalez 2021e) Nahuel Gonzáez y col. "Avances en robustecimiento ante ataques de presentación y falsificación para sistemas basados en el análisis de patrones de tecleo". En: XXIII Workshop de Investigadores en Ciencias de la Computación (WICC). Chilecito, La Rioja, 2021, pág. 177. ISBN: 978-987-24611-4-0.

(Gonzalez 2021f) Nahuel González. Dataset for An Ensemble Method for Keystroke Dynamics Authentication in Free-Text Using Word Boundaries. Ver. 1. 2021. doi: 10.17632/np9hhy6gt7.1

(Gonzalez 2021g) Nahuel González. Dataset for An Ensemble Method for Keystroke Dynamics Authentication in Free-Text Using Word Boundaries. Ver. 1. 2021. doi: 10.21227/jdzh-4m97

(Gonzalez 2021h) Nahuel González. Dataset for Towards Liveness Detection in Keystroke Dynamics: Revealing Synthetic Forgeries. Ver. 1. doi: 10.17632/xvg5j5z29p.1.

(Gonzalez 2021i) Nahuel González. Dataset for Towards Liveness Detection in Keystroke Dynamics: Revealing Synthetic Forgeries. Ver. 1. doi: 10.21227/1ka3-er49.

(Gonzalez 2021j) Nahuel González. Dataset for The Reverse Problem of Keystroke Dynamics: Guessing Typed Text with Keystroke Timings. Ver. 1. doi: 10.17632/94dwkbxf2d.1.

(Gonzalez 2021k) Nahuel González. Dataset for The Reverse Problem of Keystroke Dynamics: Guessing Typed Text with Keystroke Timings. Ver. 1. doi: 10.21227/7616-7964.

(Gonzalez 2021l) Nahuel González. Dataset of Timing distributions in free text keystroke dynamics profiles. Ver. 1. doi: 10.17632/sjk7kz35nh.1. URL: https://data.mendeley.com/datasets/sjk7kz35nh/1 (visitado 04-03-2021)

(Gonzalez 2021m) Nahuel González. Dataset of Timing distributions in free text keystroke dynamics profiles. Ver. 1. doi: 10.21227/ngv9-fa18.

(Gonzalez 2021n) Nahuel González. Dataset for Exploring internal correlations in timing features of keystroke dynamics at word boundaries and their usage for authentication and identification. Ver. 1. doi: 10.17632/vx83444p8n.1.

(Gonzalez 2022) Nahuel González, Jorge S. Ierache, Enrique P. Calot, y Waldo Hasperué. "*Atributos derivados para la clasificación de cadencias de tecleo en texto libres basados en el grado de desorden local*". En: XXVIII Congreso Argentino de Ciencias de la Computación (CACIC), 2022.

(Gonzalez 2023a) Nahuel González. "*KSDSLD—A tool for keystroke dynamics synthesis & liveness detection*". En: Software Impacts (2022), Elsevier. doi: 10.1016/j.simpa.2022.100454.

(Gonzalez 2023b) Nahuel González. "*Dataset of Human-written and Synthesized Samples of Keystroke Dynamics Features for Free-text Inputs*". En: Data in Brief (2022), Elsevier *(en prensa)*.

(Ierache 2019) Jorge S. Ierache y col. "*Líneas de investigación del Laboratorio de Sistemas de Información Avanzados: Dinámica de Tecleo, Computación Afectiva, Extracción de Relaciones Semánticas, Blockchain, y Smart Contracts*". En: XXI Workshop de Investigadores en Ciencias de la Computación (WICC).

(Milne 2018) A Milne, K Farrahi y MA Nicolaou. "*Less is more: Univariate modelling to detect early Parkinson's disease from keystroke dynamics*". En: International Conference on Discovery Science. Springer. 2018, págs. 435-446.

(Monaco 2019) John V Monaco. "*What are you searching for? a remote keylogging attack on search engine autocomplete*". En: 28th {USENIX} Security Symposium ({USENIX} Security 19). 2019, págs. 959-976.

(Monrose 1990) Fabian Monrose y Aviel D Rubin. "*Keystroke dynamics as a biometric for authentication*". En: Future Generation computer systems 16.4 (2000), págs. 351-359.

(Rahman 2011) KA Rahman, KS Balagani y VV Phoha. "*Making impostor pass rates meaningless: A case of snoop-forge-replay attack on continuous cyber-behavioral verification with keystrokes*". En: CVPR 2011 workshops. IEEE. 2011, págs. 31-38.

(Song 2011) DX Song, DA Wagner y X Tian. "*Timing analysis of keystrokes and timing attacks on ssh*." En: USENIX Security Symposium. Vol. 2001. 2001.