Generalización del Modelado de Cadencias de Tecleo para su Utilización en Ataques de Presentación y Canal Lateral

Nahuel González

Tesis de Doctorado en Ciencias Informáticas | Fecha de defensa: 07/06/2022
Facultad de Informática, Universidad Nacional de La Plata

Audio en YouTube

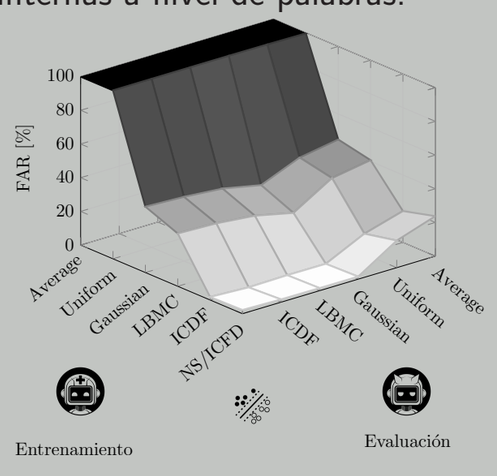
Introducción

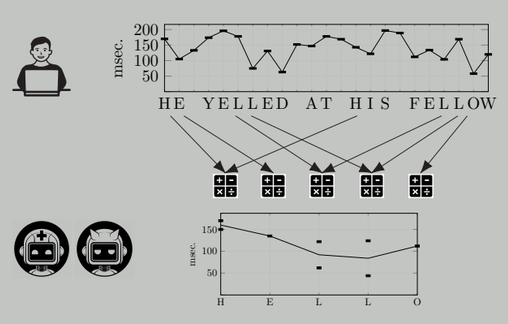
- Las **cadencias de tecleo** son un atributo biométrico comportamental que puede utilizarse como segundo factor de autenticación (2FA).
- Un **ataque de presentación** involucra a un actor malintencionado que imita las credenciales biométricas de un usuario legítimo.
- Un **ataque por canal lateral** explota la información filtrada inadvertidamente por una implementación particular de un sistema.
- Todo sistema de autenticación por cadencias de tecleo u otra modalidad biométrica es **vulnerable** a ataques de presentación y canal lateral.

El principal resultado de esta tesis es un **modelo de detección de vida** que emplea **distancias basadas en los histogramas empíricos de los tiempos de escritura** y **estrategias de síntesis de muestras** como adversarios para mitigar ataques de presentación.

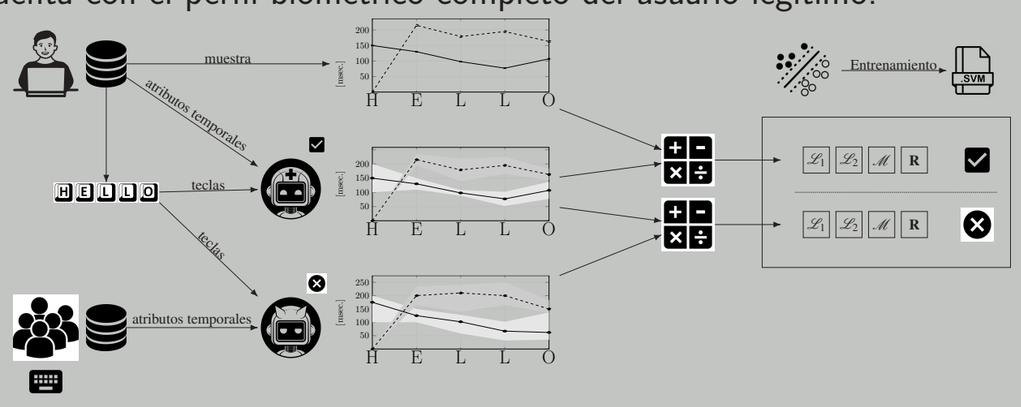
Métodos propuestos

- **Distancias basadas en histogramas empíricos**.
  - Los modelos estadísticos paramétricos no son suficientes para capturar las variaciones en la forma de las distribuciones de los tiempos de escritura.
  - Las distancias basadas en histogramas empíricos son más sensibles que las clásicas para verificar identidad y detectar muestras sintéticas.
- **Síntesis de muestras basada en histogramas empíricos**
  - Son capaces de engañar a un sistema de verificación de identidad con mayor frecuencia que los métodos del estado del arte.
  - Requieren contextos de orden elevado que consideren las correlaciones internas a nivel de palabras.





- **Detección de vida con adversarios sintéticos**
  - El modelo es capaz de neutralizar un ataque de presentación con tasas de error <2% si el atacante cuenta con poca información del objetivo.
  - La eficacia sigue siendo aceptable (<15% de error) incluso si el atacante cuenta con el perfil biométrico completo del usuario legítimo.



- **Identificación del texto ingresado con tiempos de escritura**
  - Se propuso un método capaz de abordar muestras en el orden de los 100 caracteres, basado en el anterior modelo de detección de vida.
  - Alcanzó tasas de error <2%, pero con la limitación de requerir una lista de candidatos parciales.

Motivación

- La verificación de identidad con cadencias de tecleo es transparente — no requiere acciones ulteriores del usuario ni impone demoras.
- La vulnerabilidad de los sistemas de autenticación con cadencias de tecleo ante ataques de presentación no ha recibido atención suficiente.
- La cadencia de tecleo puede emplearse para amplificar ataques de canal lateral – p.ej. reconstruyendo el texto dados los tiempos de escritura.

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 permite **potenciar los ataques por canal lateral** a través de la reconstrucción del texto ingresado en base a los tiempos de escritura.

- Herramienta de síntesis de muestras y detección de vida
  - Disponible en https://github.com/SoftwareImpacts/SIMPAC-2022-276
  - Publicada en Software Impacts, Elsevier, con revisión de pares.
- Producción científica
  - 4 publicaciones en revistas internacionales, **2 de ellas en revistas de primer cuartil (Q1) según Scimago Journal Ranking (SJR).**
  - 1 publicación en series y capítulos de libro en idioma inglés.
  - 3 presentaciones en congresos internacionales.
  - 3 presentaciones en congresos nacionales y regionales.
  - 6 presentaciones en WICC.
- Conjuntos de datos
  - Publicados en IEEE DataPort, Mendeley Data, y Data in Brief.
  - 6 conjuntos de datos de entrenamiento, de evaluación, y de resultados.
- Transferencia tecnológica
  - Ámbito nacional: Policía de Seguridad Aeroportuaria (PSA).
  - Ámbito internacional: Empresas de seguridad, sector bancario.
- Honores y menciones
  - Mejor trabajo del WSI en CACIC 2020.
  - Mejor presentación del WSI en CACIC 2021.

Conclusiones

- Sólo las distribuciones empíricas del perfil biométrico del usuario son capaces de capturar con precisión su comportamiento característico.
- Este fenómeno puede ser capitalizado para sintetizar muestras que engañan a los actuales sistemas de autenticación basados en cadencias de tecleo.
- Es posible robustecer a estos últimos con un modelo de detección de vida que emplea muestras sintéticas como adversarios.
- **El modelo de detección de vida es eficaz aún cuando el atacante cuenta con el perfil biométrico completo del usuario legítimo.**

Futura línea de investigación

En vista de los recientes desarrollos con redes neuronales adversarias generativas y *deepfakes*... **¿Es factible sintetizar muestras de cadencias de tecleo indistinguibles de sus contrapartes reales?**

Dirección de la tesis

Director	Dr. Jorge Salvador Ierache[1]
Codirector	Dr. Waldo Hasperué[2]
Asesor Científico	Dr. Enrique P. Calot[1]

El trabajo de investigación de la tesis fue radicado en [1].

[1] Laboratorio de Sistemas de Información Avanzados (LSIA)
Facultad de Ingeniería, Universidad de Buenos Aires
[2] Instituto de Investigación en Informática (III-LIDI)
Facultad de Informática, Universidad Nacional de La Plata