Análisis Temporal de Vulnerabilidades de los Sistemas Basados en Huella Dactilar

Javier Galbally, Julian Fierrez, and Javier Ortega-Garcia

Biometric Recognition Group–ATVS, EPS, Universidad Autónoma de Madrid,
C/ Francisco Tomás y Valiente 11, 28049 Madrid, Spain
{javier.galbally, julian.fierrez, javier.ortega}@uam.es

Resumen Se presenta el análisis temporal de un sistema de referencia en reconocimiento de huella dactilar basado en minucias. Estudiamos la relación entre la puntuación generada por el sistema (el NFIS2 del NIST) y el tiempo requerido para producir esa puntuación de similitud. Los resultados experimentales se obtienen sobre un subconjunto de la base de datos MCYT y muestran una correlación clara entre las dos variables (tiempo y puntuación). Así pues, de este estudio surge una nueva amenaza contra los sistemas biométricos puesto que los ataques basados en la puntuación podrían simplificarse de forma notable si se utilizase en su lugar la información temporal.

1. Introducción

En los últimos años se han realizado importantes esfuerzos de investigación en el estudio de las vulnerabilidades de los sistemas biométricos frente a ataques directos dirigidos al sensor (llevados a cabo utilizando rasgos biométricos sintéticos como dedos de goma) [1], y contra ataques indirectos (dirigidos contra alguno de los módulos internos del sistema) [2,3]. Estos estudios han desembocado en un aumento de los niveles de seguridad ofrecidos por los sistemas biométricos a través de la propuesta de nuevas contramedidas frente a los ataques analizados. Además, este interés por el estudio de las vulnerabilidades en los sistemas de reconocimiento automático de personas ha superado el campo meramente científico y han emergido a nivel internacional diferentes iniciativas que pretenden estandarizar el problema de la evaluación de seguridad en las aplicaciones biométricas, tales como la *Common Criteria* a través de diferentes Documentos de Apoyo (*Supporting Documents*) [4], o la *Biometric Evaluation Methodology* [5].

Entre las vulnerabilidades estudiadas se ha prestado especial atención a los ataques de tipo *hill-climbing* [6,7,8]. Estos algoritmos de ataque generan un determinado número de plantillas sintéticas que se modifican según un proceso iterativo de acuerdo a la puntuación que producen al ser comparadas con la plantilla atacada: si en cada iteración la puntuación aumenta los cambios se mantienen y si no se descartan. De esta forma el *score* va aumentando hasta que se alcanza el umbral de aceptación y se rompe el sistema.

A pesar de que los ataques de tipo *hill-climbing* han probado su eficiencia a la hora de romper los sistemas biométricos, presentan la fuerte restricción de necesitar la puntuación devuelta por el comparador para poder acceder a la aplicación. De hecho, incluso en el caso de que el atacante obtenga la medida de similitud, el ataque aún se puede prevenir cuantizando la puntuación de forma que el algoritmo hill-climbing no obtenga el *feedback* necesario por parte del sistema que permita ejecutar el proceso iterativo de aumento de la medida de similitud.

Los sistemas biométricos se verían amenazados por un riesgo aún mayor si pudieran ser atacados utilizando algún dato fácilmente cuantificable tal como el tiempo de comparación, o la energía consumida por el sistema en el proceso de comparación. Este tipo de información, que ya ha sido utilizada con éxito para atacar sistemas de seguridad criptográficos [9,10], está siempre accesible a un eventual atacante y es difícil de manipular o distorsionar por el diseñador del sistema (al contrario de lo que sucede con la puntuación de similitud utilizada en los algoritmos *hill-climbing* tradicionales).

En el presente trabajo llevamos a cabo un análisis temporal de un sistema de referencia de reconocimiento de huella dactilar (el NFIS2 del NIST [11]), para determinar si existe una relación entre la puntuación de similitud devuelta por el módulo de comparación, y el tiempo requerido para generar esa puntuación. El estudio permitirá determinar la viabilidad de desarrollar ataques basados en la información temporal contra este tipo de sistemas, y la necesidad o no de considerar esta amenaza no sólo a la hora de diseñar los sistemas biométricos sino también a la hora de evaluar su nivel de seguridad.

El resto del artículo está estructurado de la siguiente forma. En la Sec. 2 se describe el sistema analizado en los experimentos. La base de datos utilizada y la evaluación de rendimiento del sistema estudiado se presentan en la Sec. 3. Los resultados experimentales del análisis temporal se dan en la Sec. 4, y las conclusiones se extraen finalmente en la Sec. 5.

2. Sistema de Referencia Analizado (NFIS2)

El sistema basado en minucias analizado en los experimentos es el *Fingerprint Image Software 2* (NFIS2) del NIST [11]. Este paquete software público se utiliza en muchos trabajos como un sistema de referencia con el que se comparan nuevas soluciones de verificación de huella dactilar. Es un sistema de procesado y reconocimiento de huella dactilar que funciona sobre PC y formado por distintos módulos *software*. El extractor de características genera un fichero de texto que contiene la posición (coordenadas x e y) y orientación (ángulo con respecto al eje positivo x) de cada minucia de la huella. El comparador utiliza luego este fichero para generar la puntuación de similitud. El algoritmo de comparación calcula distancias relativas entre grupos de minucias y por tanto es invariante a la rotación y a la traslación.

Figura 1. Ejemplos típicos de huellas dactilares que se pueden encontrar en MCYT adquiridos con un nivel de control bajo (izquierda), medio (centro), y alto (derecha)

3. Base de Datos y Evaluación del Rendimiento

El análisis temporal se ha llevado a cabo utilizando un subconjunto de la base de datos MCYT [12]. El subconjunto contiene 10 muestras de los dedos índices derecho e izquierdo de 75 usuarios ($75 \times 2 \times 10 = 1,500$ imágenes), capturados con el sensor óptico UareU de Digital Persona (500 dpi). Seis de las muestras de cada dedo fueron adquiridas con un nivel de control alto (se permitía únicamente un pequeño desplazamiento del dedo respecto del centro del sensor), otras dos con un nivel de control medio, y las dos restantes con un nivel de control bajo (en la Fig. 1 se muestran ejemplos de algunas imágenes de huellas de la base de datos).

Para estimar el rendimiento en verificación del sistema sobre esta base de datos se calcularon una serie de puntuaciones de impostor y de usuario. Para ello se utilizó una de las muestras de bajo control como plantilla y las otras 9 imágenes del mismo dedo como pruebas para generar las puntuaciones genuinas, lo que lleva a $150 \times 9 = 1,350$ puntuaciones de usuario. Las puntuaciones de impostor se obtuvieron comparando cada plantilla con una muestra del resto de dedos en la base de datos, de forma que tenemos $150 \times 149 = 22,350$ puntuaciones de impostor. En la Fig. 2 mostramos las dos distribuciones de las puntuaciones (izquierda), y las curvas de FA (Falsa Aceptación) y FR (Falso Rechazo) del sistema evaluado (derecha). El EER (*Equal Error Rate*, o Tasa de Igual Error) del sistema es 1.47 %.

4. Análisis Temporal

El objetivo del estudio experimental es determinar si existe alguna correlación entre la puntuación devuelta por el sistema analizado y el tiempo de comparación requerido para producir esa puntuación. Para ello se han llevado a cabo dos experimentos, el primero para determinar si hay algún tipo de correspondencia

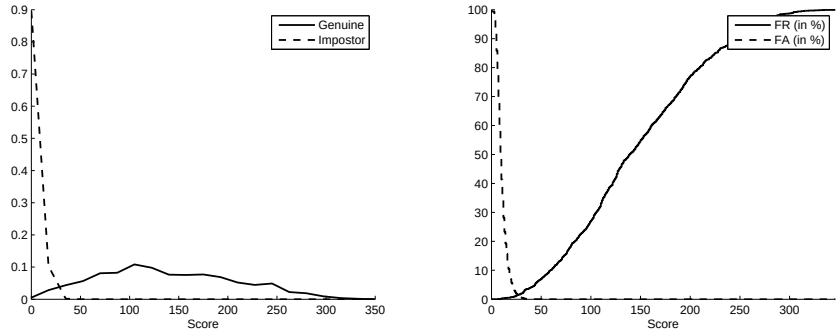

Figura 2. Distribuciones de las puntuaciones (izquierda), y curvas de FA y FR (derecha) del sistema NFIS2.

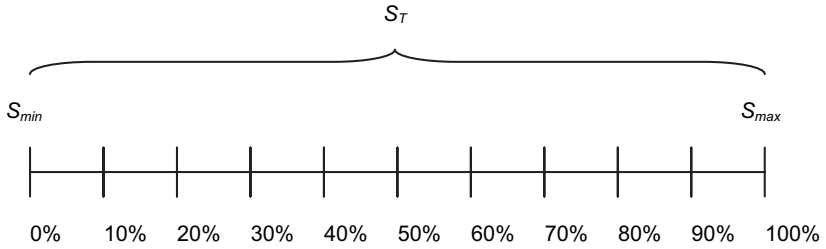

Figura 3. División del rango total de puntuaciones S_T utilizado en el experimento 1.

entre puntuación y tiempo, y el segundo para identificar el comportamiento del tiempo de comparación cuando la puntuación aumenta o disminuye.

4.1. Experimento 1: relación entre puntuación y tiempo

Tal y como se explicó en la Sec. 3, a partir de la base de datos utilizada se generó un conjunto de puntuaciones genuinas y de impostor, que contienen respectivamente 1,350 y 22,350 medidas de similitud. Cada una de estas puntuaciones tiene asociado un valor temporal que se corresponde al tiempo de comparación empleado por el sistema para producir esa puntuación. Así pues, hay un total de 23,700 puntuaciones (independientemente de si son genuinas o de impostor) y la misma cantidad de valores temporales.

En este experimento dividiremos el rango total de puntuaciones (S_T) en segmentos equiespaciados (que se corresponderán con distintos porcentajes del rango total) tal y como se muestra en la Fig. 3. Las distribuciones de los valores temporales asociados a las puntuaciones pertenecientes a cada uno de los segmentos se muestran en la Fig. 4, desde las regiones más distantes (aquellas distribuciones de tiempos que se corresponden al primer y último 10 % de S_T),

hasta las bandas más cercanas (aquellas distribuciones de tiempos que corresponden a las puntuaciones en las bandas [40-50] % y [50-60 %]). De esta forma podremos determinar si distribuciones de puntuaciones más distantes se corresponden con distribuciones de tiempo más separadas.

No se realiza distinción entre puntuaciones de usuario o de impostor ya que queremos establecer la relación entre una cierta puntuación s y su valor temporal asociado t_s, independientemente de si esa puntuación fue producida por una huella genuina o de impostor.

En la Fig. 4 podemos ver que el grado de superposición de las distribuciones temporales se incrementa entre los paneles *(a)* y *(e)*. Esto es, las distribuciones temporales que se corresponden con las puntuaciones que pertenecen a segmentos más distantes, están más separadas que aquellas que se corresponden con puntuaciones de bandas más cercanas. Además, en todos los paneles, la distribución que corresponde a una banda más alta (pintada con una línea discontinua) presenta una valor medio más alto que la correspondiente a la banda más baja de puntuaciones (dibujada con línea continua). Estas observaciones sugieren que puntuaciones más altas producen valores temporales mayores.

4.2. Experimento 2: relaciónentre varciaciones de la puntuación y del tiempo

Este segundo experimento tiene dos objetivos, en primer lugar confirmar las conclusiones extraídas del experimento previo (esto es, que valores temporales mayores se corresponden con puntuaciones más altas), y en segundo lugar determinar si existe algún tipo de correlación entre una variación en la puntuación y la variación en el tiempo necesario para generar esa puntuación. Con este propósito, se degradaron siguiendo un proceso iterativo 50 plantillas correspondientes a 50 usuarios diferentes de la base de datos. En cada una de las iteraciones del algoritmo de degradación se aplican sobre la huella dactilar una de estas dos modificaciones A) perturbar una minucia o, B) sustituir una minucia, y la plantilla modificada se compara con la original al tiempo que se mide el tiempo necesario para producir la puntuación generada en la comparción.

Las dos modificaciones A y B se tomaron de [2]. El hecho de que sólo se permitan las acciones de modificar o sustituir una minucia (y no otras modificaciones como añadir o eliminar una minucia), garantiza que el número de minucias en la plantilla permanezca constante a lo largo de las iteraciones y que los cambios en el tiempo de comparación no sean debidos a variaciones en el número de puntos singulares de la huella.

De esta forma, en la primera iteración del proceso la huella dactilar se compara consigo misma alcanzando la puntuación más alta posible para esa plantilla. A lo largo de las siguientes iteraciones, hasta alcanzar el máximo permitido (300), la puntuación decrece paulatinamente como consecuencia de los cambios introducidos en la plantilla. Al final del experimento tenemos 50 conjuntos de 300 puntuaciones (correspondientes a la evolución de la puntuación para cada una de las 50 huellas), y sus 50 conjuntos asociados de 300 valores temporales. En la Fig. 5 mostramos la evolución de la puntuación (en negro) y del tiempo

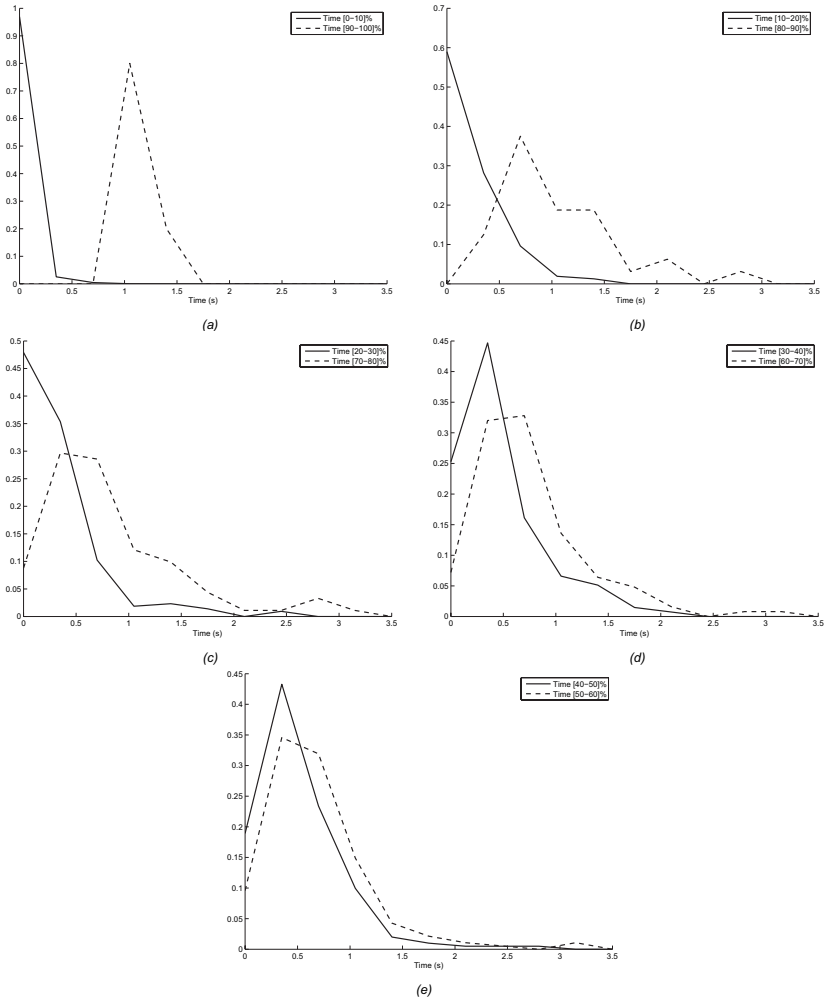

Figura 4. Distribuciones de los valores temporales asociados a las puntuaciones pertenecientes a: *(a)* bandas [0-10] % y [90-100] %, *(b)* bandas [10-20] % y [80-90] %, *(c)* bandas [20-30] % y [70-80] %, *(d)* bandas [30-40] % y[60-70] %, y *(e)* bandas [40-50] % y [50-60] %.

(en gris) para cuatro de los casos estudiados. Como los rangos de la puntuación y el tiempo son diferentes, la escala de la puntuación aparece a la izquierda de cada panel (también en negro), y la escala temporal a la derecha (en gris).

En la fila superior de la Fig. 5 hemos dibujado dos casos en los que se puede ver una clara relación entre la puntuación y el tiempo: cuanto más alta es la puntuación, mayor es el tiempo requerido por el sistema para generarla. Sin embargo, en los dos ejemplos de la fila inferior no se observar tal relación, de tal

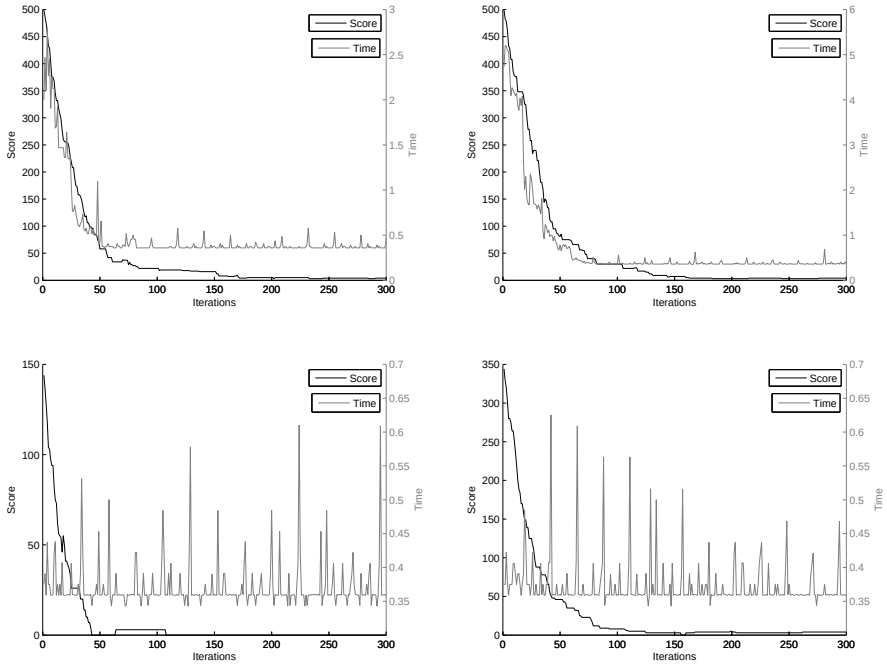

Figura 5. Evolución de la puntuación (negro) y el tiempo (gris) para cuatro de las huellas utilizadas en el experimento 2.

forma que puntuación y tiempo parecen totalmente incorrelados. Con el objetivo de estudiar el comportamiento del sistema desde un punto de vista estadístico, se calculó la media de los 50 conjuntos de puntuaciones (con 300 medidas de similitud cada uno), así como de sus correspondientes 50 conjuntos de valores temporales. Ambos valores medios se muestran en la Fig. 6 de forma análoga a la utilizada en la Fig. 5.

En la evolución de la puntuación y del tiempo mostradas en la Fig. 6 se pueden distinguir dos zonas de funcionamiento (separadas por una recta vertical discontinua) donde el sistema se comporta de forma diferente. A la izquierda de la línea discontinua podemos ver que existe un clara correlación entre variaciones de la puntuación y del tiempo: un aumento de la puntuación genera un incremento en el tiempo (en media). Sin embargo, a la derecha de la línea de separación (una vez que la puntuación ha caído aproximadamente por debajo de 30), el tiempo y la puntuación parecen ser independientes, y mientras que la puntuación continúa decreciendo, el tiempo fluctúa en torno a un valor constante.

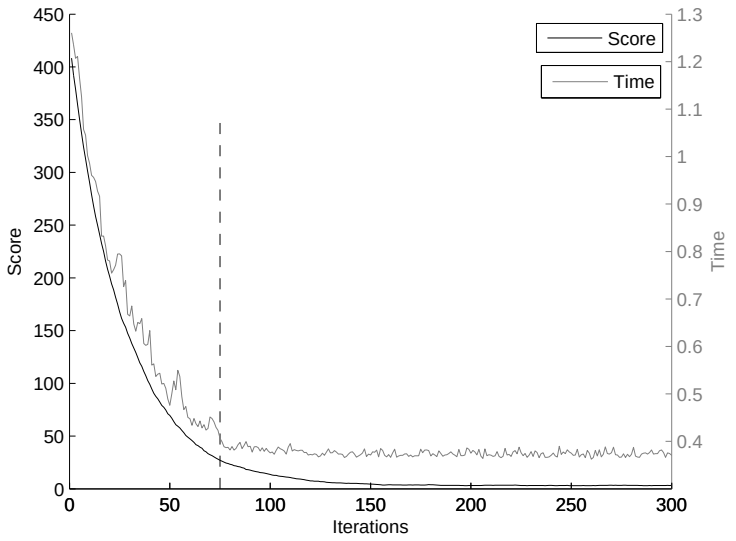

Figura 6. Media de la evolución de la puntuación (en negro) y del tiempo (en gris) de las 50 huellas consideradas en el experimento 2.

5. Conclusiones

Se ha llevado a cabo un análisis temporal de un sistema de referencia de reconocimiento de huella dactilar (el NFIS2 del NIST). Los experimentos se realizaron sobre un subconjunto de la base de datos pública MCYT y muestran que existe una clara correlación entre la puntuación devuelta por el sistema y el tiempo necesario para generar esa puntuación. Estos resultados revelan un nuevo tipo de vulnerabilidad de los sistemas biométricos ya que el tiempo de comparación (fácil de medir) puede ser utilizado para simplificar ataques inicialmente concebidos para explotar las puntuaciones de similitud (no siempre accesibles).

El presente trabajo puede resultar de especial interés no sólo a diseñadores (para incluir las contramedidas necesarias en los sistemas biométricos), sino también a evaluadores de sistemas de seguridad (para tener en cuenta esta nueva vulnerabilidad en sus análisis), y a aquellas instituciones encargadas de desarrollar estándares tales como la *Common Criteria* [4], o su asociada *Biometric Evaluation Methodology* [5].

6. Agradecimientos

J. G. es el beneficiario de una beca FPU del Ministerio Español de Educación y Ciencia. Este trabajo fue parcialmente financiado a través de los proyectos

Contexts (S2009/TIC-1485) de la CAM, Bio-Challenge (TEC2009-11186) del MICINN, y de la Cátedra UAM-Telefónica.

Referencias

1. Galbally, J., Fierrez, J., et al.: On the vulnerability of fingerprint verification systems to fake fingerprint attacks. In: Proc. IEEE of International Carnahan Conference on Security Technology (ICCST). (2006) 130–136
2. Uludag, U., Jain, A.K.: Attacks on biometric systems: a case study in fingerprints. In: Proc. SPIE-IE. Volume 5306. (2004) 622–633
3. Hill, C.J.: Risk of masquerade arising from the storage of Biometrics, B.S. Thesis. Autralian National University (2001)
4. CC: Common Criteria for Information Technology Security Evaluation. v3.1 (2006)
5. BEM: Biometric Evaluation Methodology. v1.0 (2002)
6. Adler, A.: Sample images can be independently restored from face recognition templates. In: Proc. Canadian Conference Electrical and Computing Engineering (CCECE). Volume 2. (2003) 1163–1166
7. Martinez-Diaz, M., Fierrez, J., et al.: Hill-climbing and brute force attacks on biometric systems: a case study in match-on-card fingerprint verification. In: Proc. IEEE of International Carnahan Conference on Security Technology (ICCST). (2006) 151–159
8. Galbally, J., Fierrez, J., Ortega-Garcia, J.: Bayesian hill-climbing attack and its application to signature verification. In: Proc. IAPR International Conference on Biometrics (ICB), Springer LNCS-4642 (2007) 386–395
9. Kocher, P.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Proc. 16th ICCAC, LNCS 1109 (1995) 104–113
10. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Proc. Crypto 99, LNCS 1666 (1999)
11. Watson, G.I., Garris, M.D., et al.: User's guide to NIST Fingerprint Image Software 2 (NFIS2). National Institute of Standards and Technology (2004)
12. Ortega-Garcia, J., Fierrez-Aguilar, et al.: MCYT baseline corpus: a bimodal biometric database. IEE Proc. Vision, Image and Signal Processing **150** (2003) 391–401