Biometría de reconocimiento de voz mediante comparación ciega

Vicente Jara Vera
Departamento de Matemática Aplicada a las TIC
Universidad Politécnica de Madrid (España)
vicente.jara@upm.es

Javier Espinosa García
Gerencia de Seguridad de la Información
Acciona S.A.
javier.espinosa@acciona.es

Carmen Sánchez Ávila
Departamento de Matemática Aplicada a las TIC
Universidad Politécnica de Madrid (España)
carmen.sanchez.avila@upm.es

***Resumen*—La utilización de los datos biométricos permite lograr algunos de los mejores resultados en los procedimientos de autenticación. En este entorno, y siendo necesario en los procesos de reconocimiento, ya sea de verificación como de identificación, el cotejo o comparación entre las plantillas biométricas, es importante plantearse la posibilidad de realizar dichas operaciones sin revelar los datos en comparación. De esta manera se protegen los datos biométricos, inherentes y específicos al sujeto. Ofrecemos un sistema de comparación ciega basado en el problema de los millonarios, dentro de la Computación Segura Multiparte, que cumple con estos requisitos, aplicado al caso de la biometría de voz y que hemos realizado con una gran base de datos, donde el procesamiento de la señal utiliza los datos del periodograma de Welch. Se han logrado resultados factibles en cuanto a su implementación real, y con complejidad lineal en su cómputo criptográfico.**

***Index Terms*—Biometría, Computación Segura Multiparte, Criptobiometría, Criptografía, Periodograma, Voz, Welch**

I. Introducción

La criptografía es una ciencia aplicada cuyo propósito es lograr la seguridad de las comunicaciones. Por medio de ella se consiguen resolver los aspectos de confidencialidad, de integridad, y hasta cierto punto los de autenticación. Para explicar esta parcialidad de competencia de la criptografía hemos de hacer referencia al llamado modelo de seguridad, el cual consiste en "algo-que-tengo" (un token, una tarjeta, un USB,...), "algo-que-sé" (un PIN, una password,...), y "algo-que-soy" (biometría: aspectos fisiológicos y comportamentales). El último de ellos es el que permite integrar la criptografía con la biometría en lo que se denomina criptobiometría. Por medio de ella se puede lograr la autenticación de manera completa, convirtiendo a la misma persona, sus aspectos comportamentales o fisiológicos propios, en el origen de las propias claves criptográficas [1][2].

Debido a que la biometría se refiere a peculiaridades inherentes identificativas propias anatómicas, estructurales, fisiológicas, o comportamentales y psico-sociales de las personas, podemos obtener a partir de ella los elementos identificativos propios de los sujetos y que pueden diferenciarlo de cualquier otro, en definitiva, autenticarlo, y al mismo tiempo impedir su repudio. Por otro lado, la biometría resuelve los problemas de pérdida, robo, copia, olvido, etc., que tenemos con lo que serían los elementos "algo-que-tengo" y "algo-que-sé".

Debido a la naturaleza personal de los datos biométricos, es deseable que estos deban manejarse con la máxima confidencialidad y anonimato. Durante el proceso de comparación entre los datos de la muestra que se haya de tomar y los valores previamente almacenados como patrón o plantilla, el proceso deberá ser lo menos intrusivo posible, evitando al máximo conocer los datos biométricos implicados. En general, no se utilizan los datos biométricos puros para hacer las comparaciones de verificación o de identificación, sino que se usan valores obtenidos a partir de ellos. Esto se logra mediante funciones de un solo sentido que impiden acceder a los datos originales. No obstante, siempre será un elemento añadido estimable el desconocer los valores utilizados en la comparación final. Es decir, estamos proponiendo un sistema de tal modo que el dato de la toma del sujeto durante el proceso de comparación con los valores almacenados previamente no permita conocer los datos biométricos de su plantilla. Así, el proceso deberá hacerse de manera ciega, sin revelarse a la otra parte el dato real con el cual comparamos, será una comparación biométrica ciega.

Nuestro sistema se dirige a proteger los datos del usuario en el proceso de verificación. El caso de un proceso de identificación sería más complicado de realizar de lo que discutimos en este estudio. La razón está en que nos obligaría a hacer una comparación con cada una de las plantillas guardadas en la base de datos, enviando y recibiendo información al mismo tiempo que se aplica el protocolo, que posteriormente expondremos. Y esto es así porque deberá de realizarse la comparación en un único y mismo momento, pues no se dispondrá de los datos del usuario para compararlos en un tiempo posterior. Esto nos quedará aclarado más tarde al ver nuestra propuesta en la sección III.

Vamos a hacer uso de la biometría de voz (aunque podrían utilizarse otras modalidades diferentes), mediante la cual aplicaremos verificación con modalidad de discurso libre e independiente del texto. Usaremos criptografía asimétrica RSA bajo un protocolo de Computación Segura Multiparte (CSM) [3], manteniendo así la confidencialidad y uso ciego de los datos.

En el proceso de comparación, aquellas personas que no se encuentren dentro de los rangos aceptables del sistema mantendrán su total anonimato y privacidad y no podrán ser

verificadas. Solo en el caso del sujeto admitido se puede concluir su autenticación, aunque se desconozcan los valores exactos de su muestra biométrica. De esta forma, logramos una mínima interacción informativa de los sujetos sometidos al sistema, realizando las comparaciones entre los datos del sistema y los datos de los usuarios de forma cifrada, con total desconocimiento de los datos claros.

II. Interacción privada entre partes

La interacción privada entre partes es un aspecto de notable interés que pronto surgió como una necesidad en los entornos digitales para un conocimiento mutuo parcial de los datos sin una divulgación completa. La teoría de la Computación Segura Multiparte ha sido ampliamente desarrollada desde finales de la década de 1980 [4][5][6]. Pronto surgirían algunos protocolos de preservación de la privacidad entre las partes, donde encontramos diferentes enfoques, como el uso de técnicas de intercambio secreto [7], o la evaluación de circuitos confusos, y en última instancia, funciones booleanas [4], y de ahí al uso de cifrado homomórfico [8][9][10].

El gran problema de los procedimientos anteriores es que requieren un costo computacional muy alto, con una inmensa cantidad de operaciones criptográficas, dependiendo multiplicativamente de un orden $O(n^2)$ en su complejidad matemática [11]. Otro enfoque utilizado para resolver el problema de la interacción privada ha sido el llamado saneamiento, pero tiene el problema de que perturba la información privada, generalizándola, borrándola e incluso transformándola más allá de lo deseable [12][13]. Se ha propuesto un método intermedio entre los anteriores, la denominada privacidad diferencial, donde el usuario interacciona con la base de datos por medio de consultas estadísticas y se adiciona ruido aleatorio para preservar la seguridad, con resultados cada vez más sobresalientes [14][15][16]. Dentro del campo biométrico, el desarrollo de estos sistemas no es muy amplio, aunque se ha aplicado al iris y a la huella dactilar [17][18][19], o al rostro [20].

III. Nuestra propuesta

III-A. Protocolo: el problema de los millonarios desconfiados

La Computación Segura Multiparte permite hacer uso de protocolos a través de los cuales un grupo de actores llega a un consenso o a una certidumbre mediante el cual pueden realizar un cálculo de una función multivariable. Cada uno de los actores proporciona una o más de las variables necesarias y el resultado es conocido por todos los participantes, sin saber nada sobre las variables de entrada de los otros actores.

A diferencia de otras propuestas anteriores, antes mencionadas, para evitar el enorme procesamiento computacional y la dificultad de utilizar los sistemas que hoy se ofrecen, realizaremos una ampliación de uno de los primeros protocolos propuestos en el campo de la interacción privada entre partes, que nos permitirá lograr los objetivos con poco esfuerzo y tiempo de computación.

En el año 1982 Andrew C. Yao (1946-) propuso el problema de los millonarios [21], que podemos redenominar mejor como el problema de los millonarios desconfiados, un problema de confidencialidad parcial compartida, es decir, en el que dos personas llegan a un acuerdo de certeza de ordinalidad (mayor que, menor que, igual) sin revelar la información de las cantidades en comparación. El enfoque de Yao se basa en un supuesto en el que dos millonarios, A y B, tienen como cantidades monetarias a y b, respectivamente. Debido a su desconfianza mutua, no quieren revelar sus respectivos montos monetarios al otro, pero quieren saber si son más o menos ricos que el otro.

Antes de exponer el protocolo, se debe considerar que se basa en el uso de criptografía de clave asimétrica, por lo que cada usuario tiene dos claves, una pública, conocida por cualquiera, también para el otro actor, y una privada que solo él sabe: ($KPub_A$, $KPriv_A$), ($KPub_B$, $KPriv_B$).

A continuación ofrecemos el protocolo original, aunque le hemos agregado doble cifrado asimétrico en todas las transacciones cifradas entre A y B para mejorar la seguridad y también agregar autenticación. Hemos considerado el cifrado asimétrico definido en un Grupo $\mathbb{Z}_q$, siendo q primo y menor que los valores que definen los grupos de ($KPub_A$, $KPriv_A$) y ($KPub_B$, $KPriv_B$). Los actores involucrados son A y B, con sus respectivas cantidades a comparar, a y b, y definimos el rango de las cantidades monetarias (o lo que esté bajo comparación) $u = \{1, 2, ..., n\}$, con $a,\ b \in u$.

Protocolo 1 : Compara(a,b)

Entrada: $u = \{1, 2, ..., n\}$, con $a,\ b \in u$, y $n < q$.

Salida: $\{a \geq b, a < b\}$.

Protocolo:

1. B genera un valor elevado aleatorio $x \in \mathbb{Z}_q^*$.
2. B calcula $Enc_{KPriv_B}(Enc_{KPub_A}(x)) = k$.
3. B calcula $s = k - b$ y envía a A el valor doblemente cifrado $s' = Enc_{KPriv_B}(Enc_{KPub_A}(s))$.
4. A descifra el valor recibido obteniendo $Dec_{KPriv_A}(Dec_{KPub_B}(s')) = s$.
5. A realiza un conjunto de descifrados con su clave privada, tantos como elementos tenga el conjunto u, generando $y_u = Dec_{KPriv_A}(Dec_{KPub_B}(s + \{1, 2, ..., n\}))$.
6. A busca un número primo $p' < x$ (es suficiente que B le diga a A la cantidad de cifras de x) y calcula $z_u = y_u \ mod \ p'$.
7. A verifica que se cumple $\forall\ i, j \in u,\ i \neq j$, que $|z_i - z_j| \geq 2$. Si no se cumple, A debe buscar otro valor primo p' adecuado (por ejemplo, el siguiente).
8. A envía a B la secuencia $\{z_1, ..., z_a, (z_{a+1}+1), (z_{a+2}+1), ..., (z_n+1)\} = \{z'_1, ..., z'_n\}$, junto con el número p', por medio de $Enc_{KPriv_A}(Enc_{KPub_B}())$.
9. B descifra el valor recibido aplicando $Dec_{KPriv_B}(Dec_{KPub_A}())$ obteniendo la secuencia $\{z'_1, ..., z'_n\}$, junto con el número p'.
10. B comprueba si $z_b = x \ mod \ p'$.
11. Si coincide, $a \geq b$, y en caso contrario, $a < b$.

En nuestro sistema, A corresponderá al subsistema del usuario que ofrece la plantilla de datos y B sería el subsistema verificador / identificador. Además, se debe aplicar el protocolo anterior a todos los puntos (datos) que hay que comparar entre ambas plantillas, la nueva generada por el usuario A y la almacenada en el subsistema B. Así, si definimos

como N la cantidad total de valores a comparar, todos ellos normalizados al valor n, podemos volver a considerar el conjunto $u = \{1, 2, ..., n\}$, como se definió anteriormente.

El protocolo en todos los demás aspectos es similar al indicado anteriormente, aunque ahora el actor B comparará para cada valor del vector $\overrightarrow{j}$, $\{j_1, j_2, ..., j_N\}$, almacenado previamente, cada valor de la plantilla de A, el vector $\overrightarrow{i}$, $\{i_1, i_2, ..., i_N\}$.

Si se define un vector de error $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_N\}$, apropiado a las condiciones de precisión del sistema comparador, el protocolo comparará los valores i_m con $j_m - \varepsilon_m$, por un lado, y con $j_m + \varepsilon_m$ por otro, siendo $m \in \{1, ..., N\}$, ofreciendo como salida la cantidad de valores ($Total$) dentro de su respectivo intervalo (Protocolo general). Sin embargo, no siempre se ejecutarán ambos, ya que si resulta que $j_m - \varepsilon_m > i_m$ no sería necesario probar el valor $j_m + \varepsilon_m$. De esta forma sabremos si los diferentes valores de A, $\{i_1, i_2, ..., i_N\}$, están en el rango $\{j_m - \varepsilon_m, j_m + \varepsilon_m\}$ o no. Tras este proceso, y según un umbral determinado de coincidencias, podremos aceptar o rechazar al usuario A.

Protocolo 2 : Protocolo general

Entrada: $\overrightarrow{i}$, $\overrightarrow{j}$, $\overrightarrow{\varepsilon}$.

Salida: $Total$.

Protocolo:

1. $Total = 0$
2. **For** $m = 1$ **to** N **do**
   $Izquierda_m := Compara(i_m, j_m - \varepsilon_m)$
   If $(Izquierda_m \equiv \geq)$ **then**
   $Derecha_m := Compara(i_m, j_m + \varepsilon_m)$
   If $(Derecha_m \equiv <)$ **then**
   $Total = Total + 1$
   end if
   end if
   end for

III-B. Biometría de voz

La voz es una de las modalidades de reconocimiento, identificación y autenticación biométricas existentes. Sabemos que los seres humanos pueden reconocer voces familiares e identificar multitudes de personas solo por la voz, aunque siempre nos expresamos de manera diferente y ni siquiera decimos las mismas palabras de la misma forma [22]. Los diferentes tipos de sonidos son el resultado acústico de las ondas sonoras y, como tales, podemos analizarlos con tecnologías de procesamiento de señales [23][24][25].

El reconocimiento de voz tiene dos fases principales: en la primera fase, de enrolamiento, se toman muestras de la voz de los hablantes, extrayendo sus características principales y diferenciadoras, constituyendo la plantilla, la cual será almacenada en el sistema. En la fase posterior, de reconocimiento (verificación / identificación), se toma una muestra de voz, que se compara con los datos almacenados en el sistema.

Cuando hablamos de reconocimiento de voz debemos distinguir entre dos tipos: por un lado, la verificación (o autenticación), mediante la cual una persona afirma ser una persona concreta y específica, y cuya verdad o falsedad es verificada por el sistema, comparando una muestra con su plantilla o patrón, previamente generado en su proceso de inscripción o enrolamiento, almacenado en el sistema. La otra es la identificación, a través de la cual se pretende a través de una muestra de voz y comparando con un amplio (tanto como sea necesario) conjunto de plantillas de diversos usuarios, decir quién es exactamente dicho sujeto (siempre que esté en el sistema).

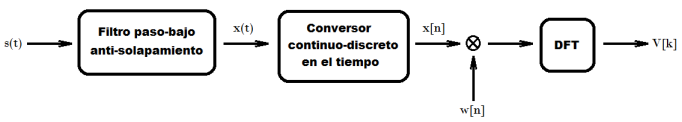

Figura 1. Esquema para el procesamiento de una señal desde su origen hasta su valor discretizado por la Transformada Discreta de Fourier (DFT).

En biometría de voz se distingue entre técnicas dependientes o independientes del texto, ya sea si las muestras de los hablantes son o no de la misma frase o texto común. Nuestro sistema implementado es de verificación e independiente del texto [26][27][28].

El discurso hablado puede modelarse como la respuesta de un sistema lineal invariante en el tiempo, el sistema de voz, a un tren de pulsos casi periódicos para sonidos sonoros y ruido de banda ancha para sonidos sordos. La articulación de todo el tracto vocal puede modelarse como un filtro que varía lentamente en el tiempo, dando así el rango de frecuencias del habla. Por tanto, el habla es una señal no estacionaria. Sus características suelen permanecer muy constantes en intervalos cortos, entre $30 - 40$ ms. Aunque la frecuencia de la señal de sonido humano alcanza un rango de unos 15 kHz o incluso superior, se puede filtrar hasta 3 kHz, siendo perfectamente inteligible. Con una ventana L adecuada podemos así recoger la invariancia de las propiedades de la señal, y mediante la DFT (Transformada Discreta de Fourier) visualizar las propiedades en el dominio de frecuencia de la señal en dicho intervalo (Fig. 1). La transformada de Fourier dependiente del tiempo proporciona una descripción muy útil de las propiedades de la señal a lo largo del tiempo [29].

Un estimador especialmente relevante en el reconocimiento de voz es el Espectro de Densidad de Potencia de la señal bajo DFT, obteniéndose así el periodograma, basado en la Transformada Directa de Fourier de segmentos de longitud finita de la señal [30][31].

Se define el periodograma $I(\omega)$ como $I(\omega) = \dfrac{S(\omega)}{LU}$ con

$$S(\omega) = \sum_{m=-(L-1)}^{L-1} \left(\sum_{n=0}^{L-1} x[n]w[n]x[n+m]w[n+m] \right) e^{-j\omega m},$$

donde $x[n]$ es la señal inicial discreta y $w[n]$ la discretización de la ventana que selecciona una cierta cantidad finita, siendo L el número de muestras del segmento de longitud finita, con U un factor de valor normalizador

$$U = \frac{1}{L} \sum_{n=0}^{L-1} (w[n])^2.$$

La expresión mejorada de los datos del periodograma (Fig. 2), por la aplicación de lo que se denomina el método de Welch (Fig. 3), se puede realizar mediante el promedio del periodograma, dividiendo la secuencia $x[n]$, $0 \leq n \leq (Q-1)$

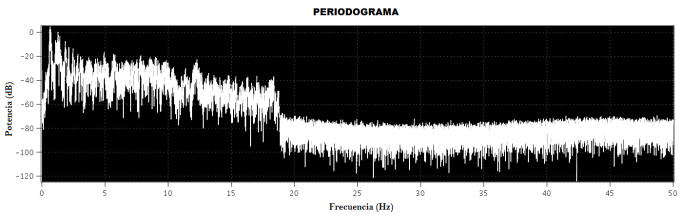

Figura 2. Periodograma de una señal vocal.

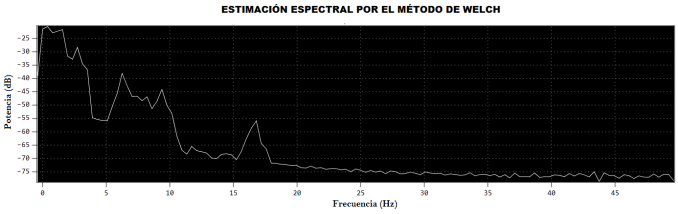

Figura 3. Estimación espectral promedio del periodograma anterior por el método de Welch.

(donde Q corresponde a la longitud de los datos disponibles) en segmentos de longitud un número L de muestras:

$$x_r[n] = x[rR + n]w[n],\ 0 \leq n \leq L - 1.$$

Para $R = L$ los segmentos serán continuos, quedando solapados en el caso de $R < L$. Dividiendo la señal en estudio en pequeños segmentos y promediando sus periodogramas respectivos se obtiene un estimador con un buen comportamiento [32][33].

Las fases del método de Welch son las siguientes:

1. División de la señal (con solapamiento de segmentos). Vamos a considerar dos escenarios, por un lado $2^7 = 128$ muestras, y por otro, uno más preciso de $2^9 = 512$ muestras, usando solapamiento mitad.
2. Enventanamiento y FFT. Usaremos la variante eficiente de la DFT, la Transformada de Fourier Rápida (FFT); tomaremos una ventana de Hamming con tamaño de $2^8 = 256$ y $2^{10} = 1024$, para cada uno de los casos, y constante de normalización en ambos casos de $U = 0.3970$.

$$w[n] = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{M}) & 0 \leq n \leq M \\ 0 & resto \end{cases}$$

3. Promediado. Los valores promedio y normalizados se calculan a partir de los valores vectorizados de los fragmentos superpuestos de la señal enventanada y procesada por la FFT.

III-C. Resultados experimentales

La base de datos de voz utilizada ha sido "Common Voice" de Mozilla [34]. Se han realizado un total de 35 (escenario 1) / 25 (escenario 2) experimentos diferentes para intentar verificar a una persona (con 50 fragmentos vocales de texto independiente por persona), además de considerar como impostores otras personas (950 (escenario 1) / 200 (escenario 2) usuarios diferentes y con independencia de texto), usando respectivamente en las fases de entrenamiento / verificación los porcentajes del $80\,\% - 20\,\%$ sobre el total de muestras. Los fragmentos de voz tienen una duración de 4000 ms. Los usuarios son hombres y mujeres con diversidad de edades,

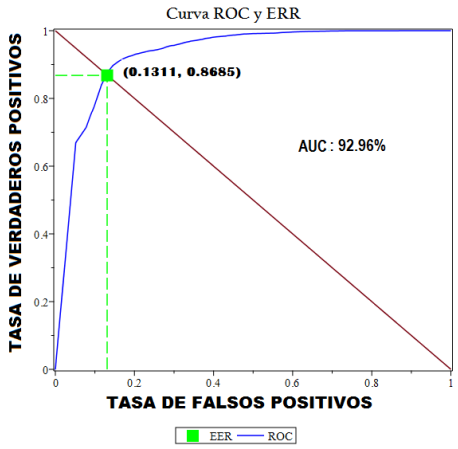

Figura 4. Curva ROC, punto EER y área bajo la curva ROC (AUC) de los experimentos realizados (escenario 1).

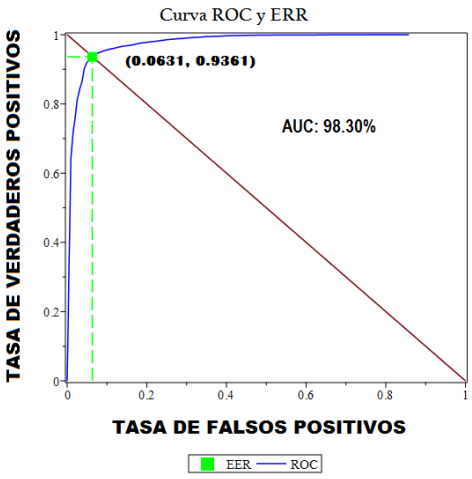

Figura 5. Curva ROC, punto EER y área bajo la curva ROC (AUC) de los experimentos realizados (escenario 2).

en su casi mayoría hablantes en español (de España y de Hispanoamérica), y algunos de ellos hablantes en lengua inglesa.

Mostramos en Fig. 4 y Fig. 5 los resultados experimentales a través de la curva ROC (Característica Operativa del Receptor), que ofrece gráficamente la relación entre Sensibilidad, o Tasa de Verdaderos Positivos, en el eje vertical, versus Especificidad (realmente, 1-Especificidad), o Tasa de Falsos Positivos, en el eje horizontal, para nuestro sistema de clasificador binario, conforme varía el umbral de discriminación. El valor del Área Bajo la Curva (AUC) es del 92.96 % (escenario 1) y del 98.30 % (escenario 2), y el punto de Igual Tasa de Error (EER), punto donde la Tasa de Falsos Positivos y la Tasa de Falsos Negativos son iguales es $(0.1311, 0.8685)$ y $(0.0631, 0.9361)$, respectivamente en cada escenario.

IV. Conclusiones

La biometría se utilizará cada vez más en entornos de verificación e identificación personal. Ante esta perspectiva es necesario, así creemos, dada la sensibilidad de dichos datos como inherentes a la persona, proteger al máximo las características personales.

Es en este escenario donde proponemos realizar el proceso de comparación entre la muestra del sujeto y la plantilla almacenada en el sistema sin revelar los datos reales. Entre las posibles opciones tecnológicas disponibles, como la Computación Segura Multiparte, el saneamiento o variantes intermedias, no parece existir una solución aceptable con alto rendimiento y capaz de implementarse sin alto costo computacional, y por lo tanto, sin alto costo de tiempo. Hemos tomado uno de los primeros protocolos de la Computación Segura Multiparte para adaptarlo a la situación biométrica. Partiendo del problema de los millonarios desconfiados hemos diseñado un sistema de comparación ciega aplicándolo a los extremos de los intervalos de cada punto o valor de los datos biométricos, basándonos en la media y tomando como umbrales la varianza, a partir de la que se busca un margen de error.

Nuestro sistema, utilizando cifrado RSA para las parejas ($KPub_A$, $KPriv_A$), ($KPub_B$, $KPriv_B$), siguiendo las recomendaciones de tamaño de los números primos actuales, utiliza valores de alrededor de 1400 bits para cada primo p, q, lo que da como resultado un valor de 2800 bits para el módulo $r = pq$ del cifrado RSA de cada usuario o actor.

En cuanto al análisis de complejidad computacional, en el protocolo 1, Compara(a,b), el orden viene limitado por las operaciones criptográficas, Crp. Esta operación se aplica una cantidad de veces regida por el cardinal del conjunto u, que denotaremos como $|u|$. Al aplicarse el protocolo 2, el bucle allí presente en el peor de los casos obliga a realizar $2N$ comparaciones del tipo Compara(a,b), resultando finalmente un orden lineal en las operaciones criptográficas de orden $O(2 \cdot N \cdot |u| \cdot Crp)$. Por todo ello, se trata de un sistema viable en cuanto a su implementación.

Como limitación indicar que la forma de realizar la comparación en el proceso de emparejamiento impide utilizar los datos de las muestras en su conjunto, relacionarlas o correlacionarlas, lo que estadísticamente permite un mejor discernimiento de similitudes y plausibilidad, debiéndonos conformar con hacer una comparación secuencial, valor por valor. No obstante, se pueden lograr resultados muy satisfactorios, como hemos visto en el caso biométrico mostrado, tras caracterizar el procesamiento de la señal bajo el periodograma de Welch, en dos escenarios con diversidad de precisión.

Se ha tomado como ejemplo de aplicación la biometría de voz, una modalidad biométrica generalmente aceptada por el público. Los altos valores de los resultados obtenidos, junto con el hecho de que el sistema podría actualizar los valores vocales y de habla de los sujetos, lo convierten en un sistema factible para ser utilizado en la práctica. Por otro lado, este mismo sistema de comparación ciega siempre se puede utilizar con otras modalidades biométricas, posiblemente con resultados aceptables, siempre que los datos biométricos puedan expresarse como valores ordenados y de igual cardinalidad para ser comparados entre sí.

En definitiva y finalmente, en los próximos tiempos deberíamos estar cada vez más acostumbrados a aplicar procedimientos criptográficos en el uso y manipulación de datos biométricos.

REFERENCIAS

[1] V. Jara-Vera y C. Sánchez-Ávila: “La Criptobiometría y la Redefinición de los Conceptos de Persona e Identidad como Claves para la Seguridad”, en *Proc. DESEi+d*, Madrid, España, pp. 583-590, 2013.

[2] A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen y S. Yadav: “Some Research Problems in Biometrics: The Future Beckons”, en *International Conference on Biometrics*, pp. 1-8, 2019.

[3] B. Schneier: “Applied Cryptography”, 2 ed., New York, NY, USA: John Wiley & Sons, pp. 134-137, 551-552, 1996.

[4] O. Goldreich, S. Micali y A. Wigderson: “How to Play any Mental Game or a Completeness Theorem for Protocols with Honest Majority”, en *ACM Symposium on Theory of Computing*, pp. 218-229, 1987.

[5] A. Yao: “How to Generate and Exchange Secrets”, en *IEEE Symposium on Foundations of Computer Science*, pp. 162-167, 1986.

[6] Y. Lindell: “Secure Multiparty Computation”, en *Communications of the ACM*, vol. 64 (1), pp. 86-96, 2020.

[7] T. Rabin y M. Ben-Or: “Verifiable Secret Sharing and Multiparty Protocols with Honest Majority”, en *ACM Symposium on Theory of Computing*, pp. 73-85, 1989.

[8] R. Cramer, I. Damgård y J. Nielsen: “Multiparty Computation from Threshold Homomorphic Encryption”, en *Advances in Cryptology, EUROCRYPT*, vol. 2045, pp. 280-300, 2001.

[9] M. J. Freedman, K. Nissim y B. Pinkas: “Efficient Private Matching and Set Intersection”, en *Advances in Cryptology, EUROCRYPT*, vol. 3027, pp. 1-19, 2004.

[10] A. Kulshrestha, A. Rampuria, M. Denton y A. Sreenivas: “Cryptographically Secure Multiparty Computation and Distributed Auctions Using Homomorphic Encryption”, en *Cryptography*, vol. 1 (25), pp. 1-18, 2017.

[11] R. Agrawal, A. Evfimievski y R.Srikant: “Information Sharing across Private Databases”, en *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 86-97, 2003.

[12] R. Agrawal y R. Srikant: “Privacy Preserving Data Mining”, en *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 439-450, 2000.

[13] L. Sweeney: “k-Anonymity: A Model for Protecting Privacy”, en *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, pp. 557-570, 2002.

[14] A. Inan, M. Kantarcioglu, G. Ghinita y E. Bertino: “Private Record Matching Using Differential Privacy”, en *Proceedings of the 13th International Conference on Extending Database Technology*, pp. 123-134, 2010.

[15] C. Dwork y A. Roth: “The Algorithmic Foundations of Differential Privacy”, en *Foundations and Trends in Theoretical Computer Science*, vol. 9 (3-4), pp. 211-407, 2014.

[16] M. U. Hassan, M. H. Rehmani y J. Chen: “Differential Privacy Techniques for Cyber Physical Systems: A Survey”, en *IEEE Communications Surveys & Tutorials*, vol. 22 (1), pp. 746-789, 2020.

[17] M. Blanton y P. Gasti: “Secure and Efficient Protocols for Iris and Fingerprint Identification”, en *Computer Security, ESORICS*, pp. 190-209, 2010.

[18] Y. Luo, S. S. Cheung, T. Pignata, R. Lazzeretti y M. Barni: “An Efficient Protocol for Private Iris-Code Matching by Means of Garbled Circuits”, en *IEEE International Conference on Image Processing*, pp. 2653-2656, 2012.

[19] T. Wang, Z. Zheng, A. K. Bashir, A. Jolfaei y Y. Xu: “Finprivacy: A Privacy-Preserving Mechanism for Fingerprint Identification”, en *Journal of Association for Computing Machinery*, vol. 37 (4-111), pp. 1-16, 2018.

[20] A. R. Sadeghi, T. Schneider y I. Wehrenberg: “Efficient Privacy-Preserving Face Recognition”, en *International Conference on Information Security and Cryptology*, pp. 229-244, 2009.

[21] A. C. Yao: “Protocols for Secure Computations”, en *Proceedings of the 23rd IEEE Symposium on the Foundations of Computer Science*, pp. 160-164, 1982.

[22] J. González-Rodríguez, D. Torre-Toledano y J. Ortega-García: “Voice Biometrics”, en *Handbook of Biometrics*, A. K. Jain, P. Flynn y A. A. Ross, Eds. New York, NY, USA: Springer Science & Business Media, 2007.

[23] T. F. Quatieri: “Discrete-Time Speech Signal Processing: Principles and Practice”, MIT Lincoln Laboratory, Lexington, MA, USA: Pearson Education, 2008.

[24] G. B. Varile, R. Cole, R. A. Cole, A. Zampolli, J. Mariani, H. Uszkoreit y A. Zaenen: “Survey of the State of the Art in Human Language Technology”, Cambridge, UK: Cambridge University Press, 1997.

[25] H. Beigi: “Fundamentals of Speaker Recognition”, New York, NY, USA: Springer Science & Business Media, 2011.

[26] Z. Saquib, N. Salam, R. P. Nair, N. Pandey y A. Joshi: "A Survey on Automatic Speaker Recognition Systems", en *Communications in Computer and Information Science*, vol. 123, pp. 134-145, 2010.
[27] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal y R. Wang: "Speaker Identification Features Extraction Methods: a Systematic Review", en *Expert Systems with Applications*, vol. 90, pp. 250-271, 2017.
[28] Z. Meng, M. U. B. Altaf y B. H. F. Juang: "Active Voice Authentication", en *Digital Signal Processing*, vol. 101, pp. 1-39, 2020.
[29] P. Stoica y R. Moses: "Spectral Analysis of Signals", Upper Saddle River, NJ, USA: Prentice Hall, 2005.
[30] S. L. Marple: "Digital Spectral Analysis with Applications", Englewood Cliffs, NJ, USA: Prentice Hall, 1987.
[31] A. V. Oppenheim y R. W. Schafer: "Discrete-Time Signal Processing", Englewood Cliffs, NJ, USA: Prentice Hall, 1989.
[32] P. D. Welch: "The Use of Fast Fourier Transform for the Estimation of Power Spectra: a Method Based on Time Averaging over Short, Modified Periodograms", en *IEEE Transactions on Audio and Electroacoustics*, AU-15(2), pp. 70-73, 1967.
[33] P. Stoica y R. Moses: "Introduction to Spectral Analysis", Englewood Cliffs, NJ, USA: Prentice Hall, pp. 52-54, 1989.
[34] Mozilla.org: "Common Voice", https://voice.mozilla.org.