Ciencias Matemáticas, Vol. 33, No. 2, Pag. 119-128, 2019
Recibido 11-2018, Aceptado 09-2019

Nuevo modelo probabilístico en autenticación gráfica
New probabilistic model on graphical authentication

Carlos Miguel Legón [1]*, Pedro Navarro[1], Raisa Socorro[2], Osviel Rodríguez[3], Ernesto Borrego[1]

Resumen La autenticación es esencial en la seguridad de los modernos servicios digitales de procesamiento de información. Las contraseñas alfanuméricas son las más empleadas, pero poseen debilidades que las hacen vulnerables a diversos ataques basados en modelos probabilísticos. Una de las alternativas es la autenticación gráfica. Los modelos probabilísticos en autenticación gráfica se aplican para estimar la clave más probable a seleccionar, en cada imagen, por el usuario que se va a registrar. En este trabajo se propone un nuevo modelo probabilístico de autenticación gráfica, su principal aporte consiste en que permite cuantificar el grado de autenticidad de cada usuario. Se confirma experimentalmente que el modelo propuesto es efectivo y permite medir en la práctica el nivel de autenticidad de los usarios autenticados.

Abstract Authentication is essential in the security of modern digital information processing services. Alphanumeric passwords are the most used, but they have weaknesses that make them vulnerable to various attacks based on probabilistic models. One of the alternatives is graphical authentication. The probabilistic models in graphical authentication are applied to estimate the most likely key to be selected in each image by the user to be registered. In this work a new probabilistic model of graphic authentication is proposed, its main contribution is that it allows to quantify the degree of authenticity of each user. It is confirmed experimentally that the proposed model is effective and allows to measure in practice the authenticity level of autenticated users.

Palabras Clave
Autenticación gráfica—modelo probabilístico—contraseñas

[1] *Instituto de Criptografía, Facultad de Matemática y Computación, Universidad de la Habana, Habana, clegon58@gmail.com, pedropepe3437@gmail.com, ernesto.borrego@matcom.uh.cu*
[2] *Facultad de Informática. Universidad Tecnológica de La Habana, raisa@ceis.cujae.edu.cu*
[3] *Facultad de Ciencias y tecnologías. Universidad de las ciencias Informáticas, osviel@uci.cu*
***Autor para Correspondencia**

Introducción

La autenticación de un usuario para concederle acceso a un sistema o recurso es un aspecto esencial para la seguridad de la información [9]. Según la información que utilizan, los sistemas de autenticación suelen clasificarse en: sistemas basados en conocimiento (¿qué sabes?), en Tokens (¿qué tienes?) y en información Biométrica (¿quién eres?). Los basados en conocimiento emplean contraseñas, las que pueden ser alfanuméricas o gráficas [28].

Las contraseñas alfanuméricas son las más empleadas, a pesar de que poseen una contradicción entre su seguridad y su usabilidad, pues para ser seguras deben ser aleatorias, largas y no predecibles mientras que para ser usables deben ser memorizables, esta contradicción suele denotarse como *the password problem* [35]. Estas debilidades reducen el espacio de búsqueda y las hacen predecibles y vulnerables a diversos ataques basados en modelos probabilísticos [22].

Un interesante estudio sobre la aplicación de modelos probabilísticos para realizar ataques de diccionarios a las contraseñas alfanumérica y también para evaluar su seguridad puede verse en [23]. Se profundiza en los dos modelos principales, los modelos de Márkov [10], [18] y las gramáticas libres de contexto [37], [19], [7]. Existen otros modelos menos populares como la ley de zip [25], [15], [36],[24], las redes neuronales [26] y técnicas de aprendizaje automático [33], [12].

Para las contraseñas gráficas el empleo de modelos probabilísticos ha sido menos investigado y existen menos modelos. En su mayoría estos se basan en técnicas de tratamiento de imágenes digitales y en las características de la imagen [16], [40], [14]. Estos modelos permiten pronosticar, para cada imagen, la clave más probable a seleccionar por el usuario que se va a registrar. Se usan para escoger las imágenes más adecuadas a emplear en este tipo de autenticación y también para desarrollar ataques de diccionario.

En este trabajo se propone un nuevo modelo probabilístico de contraseñas gráfica, la novedad principal de este modelo consiste en que después que un usuario es autenticado, este modelo permite cuantificar la autenticidad de este usuario, asignándole una probabilidad de ser el usuario legítimo. Hasta

donde saben los autores de este trabajo, no existen antecedentes de esto en los modelos de contraseñas gráficas. Se confirma experimentalmente que el modelo propuesto es efectivo y permite medir en la práctica el nivel de autenticidad de los usarios autenticados. Esta investigación se centra específicamente en los sistemas de autenticación gráfica del tipo *Cued Recal* [5],[31].

1. Preliminares

Modelos probabilísticos: contraseñas alfanuméricas

Modelos Probabilísticos de contraseñas: un modelo probabilístico de contraseñas [23], está determinado por cualquier función P, definida en el espacio de posibles contraseñas (S) en el intervalo $[0,1]$, que asigna una probabilidad $P(s)$ a cada contraseña de forma tal que:

- $P(s) \geq 0$, para toda contraseña $s \in S$.
- $\sum_{i=1}^{|S|} P(s_i) = 1$

Estos modelos constituyen una herramienta fundamental para investigar la seguridad de las contraseñas [23]. La definición e interpretación de $P(s)$, depende del modelo.

La existencia de grandes bases de datos de contraseñas alfanuméricas disponibles en internet [13], ha permitido capturar experimentalmente sus características, las que son utilizadas para la definición de $P(s)$. Las cadenas de Márkov y las gramáticas libres de contexto han sido los dos modelos más empleados para cuantificar mediante $P(s)$ la probabilidad de que esa contraseña s sea seleccionada, por el usuario, en la fase de registro.Estos valores $P(s)$ son la base de los ataques de diccionario y de algunas métricas de evaluación de la seguridad de las contraseñas([27], [34],[2], [38]).

Existen diversas herramientas (consideradas *software libre*) para atacar contraseñas [13], las cuales hacen uso de diferentes modelos y de la información obtenida de las bases de datos. Entre estas herramientas para investigar la seguridad de las contraseñas alfanuméricas se destaca PARS [32], una plataforma propuesta en 2015 que contiene 12 algoritmos para atacar contraseñas, 15 sitios sobre métricas de fortaleza de contraseñas, 8 métricas académicas de fortaleza de contraseñas y 15 métricas comerciales de fortaleza de contraseñas.

En [17] se propone una herramienta estadística, para clasificar en 3 clases los intentos de autenticación mediante contraseñas alfanuméricas, detectando los intentos sospechosos. Se emplean diferentes parámetros como: dirección IP, geolocalización, configuración del *browser*, hora, etc. En [23] se clasifican las investigaciones sobre contraseñas atendiendo a su objetivo, esta clasificación debe ser ampliada para incluir investigaciones como [17] y la propuesta en este trabajo.

Sistemas de autenticación gráfica *cued recall*

En [5] se presenta un resumen muy completo hasta 2011 de los sistemas de autenticación basados en contraseñas gráficas, [31] se enfoca en los sistemas del tipo *Cued click points*. Una descripción y evaluación critica actualizada de la seguridad y usabilidad de los diferentes sistemas de autenticación gráfica puede verse en [28]. Una propuesta de 2018 para implementar autenticación para computación en la nube mediante contraseñas gráficas, se propone en [30]. En [20] se comparan distintos sistemas de autenticación gráfica de acuerdo a los parámetros usabilidad, confiabilidad, funcionalidad mantenibilidad, eficiencia y portabilidad como se definen en la norma $ISO-9126$ [1].

En los Sistemas de Autenticación Gráfica del tipo *Cued Recall*, la contraseña del usuario consiste en k puntos (pixeles) que este selecciona en la fase de registro, de una (o varias) imágenes, dada por el sistema o escogida por el usuario. Se espera que el usuario legítimo recuerde aproximadamente el orden y la posición de los k pixeles seleccionados en la fase de registro, pero realmente es muy poco probable que logre recordar de forma exacta la posición de cada pixel. Por esta razón la imagen se **discretiza**, definiendóse una región de tolerancia alrededor de cada punto. En la fase de registro, el sistema por cuestiones de seguridad, no almacena los k puntos, ni sus regiones de tolerancia, sino el valor del *hash* de la concatenación ordenada de las k regiones de tolerancia determinadas por la contraseña.

En estos sistemas se han empleado varios métodos de discretización para definir la región de tolerancia, la discretización robusta [6], la centrada [11] y la óptima [4]. La discretización robusta requiere una región de tolerancia mayor que las centrada y óptima. Una descripción detallada de estas tres discretizaciones y una discusión de sus limitaciones puede verse en [8].

Para **autenticarse**, el usuario debe escoger en el orden correcto, los mismos k puntos aproximadamente. Será autenticado sí y solo sí los puntos que escoge determinan un *hash* igual al que fue guardado por el sistema, es decir si los k puntos que escogió están dentro de las regiones de tolerancia (definidas por el método de discretización) de su correspondiente punto de la contraseña. Entre los sistemas de este tipo, destaca por sus ventajas el sistema *Pass Point* [39].

El espacio de claves queda determinado por 3 parámetros, el tamaño de la imagen, el tamaño de la región de tolerancia, y el número k de puntos de la contraseña. Para cada tamaño de imagen se pueden escoger el número de k puntos y el tamaño de la región de tolerancia de forma que la dimensión del espacio de contraseñas sea mayor que para una contraseña alfanumérica de k puntos

Limitaciones de los métodos de discretización

Durante la autenticación entre las principales limitaciones detectadas se encuentran:

Primera: La distancia entre el punto de la contraseña y el punto escogido para la autenticación se tiene en cuenta para autenticar o no al usuario, pero no se tiene en cuenta entre los usuarios autenticados. La autenticación no hace distinción entre los puntos dentro de la región de tolerancia, les da el mismo tratamiento a todos los puntos dentro de esta región. Este enfoque tiene una

limitación pues contradice el comportamiento esperado para un usuario legítimo, del cual intuitivamente se espera que escoja con mayor frecuencia a los puntos más cercanos al punto legítimo de la contraseña. Sorpresivamente, no se han encontrado reportes donde se discuta esta limitación de la autenticación gráfica, la cual se investigará usando el modelo propuesto en este trabajo.

Segunda: Existen algunas parejas de puntos situados ambos a la misma distancia del punto de la contraseña, sin embargo unos quedan dentro de la región de tolerancia y otros quedan fuera. Estos puntos serian igualmente aceptables para el usuario legítimo, sin embargo, la discretización les da un tratamiento diferente.

Tercera: Existen algunas parejas de puntos, tales que ambos están situados a diferentes distancias del punto de la contraseña, pero uno es interno a la región de tolerancia y el otro es externo. La limitación de la discretización consiste en que el punto que queda fuera de la región de tolerancia está más cerca del punto de la contraseña que el que queda adentro.
Las limitaciones segunda y tercera se deben a que la región de tolerancia es cuadrada mientras la distancia define un círculo, una solución podría ser definir una región de tolerancia circular [8],[21].

Cuarta: Cada una de estas discretizaciones conserva cierta información necesaria para repetir la discretización de la imagen en la fase de autenticación. Esa información es aprovechada para aumentar la efectividad de los ataques de diccionario [3], [29].

Modelos probabilísticos de contraseñas gráficas

En contraseñas gráficas, no existen bases de datos de contraseñas disponibles en internet, pero aun así se han aplicado modelos probabilísticos [16],[40],[14].
Las características de la función $P(s)$ han sido extraídas de tres fuentes principales, en primer lugar, de la propia imagen propuesta para la autenticación usando técnicas de segmentación de imágenes para detectar las regiones más probables. En segundo lugar, de la información aportada por el método de discretización y tercero de las características personales del usuario.
Estos modelos son aplicados en ataques de diccionario, pero hasta donde sabemos, en ningún caso se aplica un modelo para cuantificar el grado de autenticidad del usuario.
En [16] se propone un modelo que para cada punto de la imagen calcula la probabilidad de que, en la fase de registro, ese punto sea seleccionado por el usuario legítimo como punto de la contraseña. Emplean segmentación de imágenes para detectar las regiones más probables. Es aplicable en **ataques de diccionario** y también para evaluar si una imagen es apropiada para ser usada en este tipo de autenticación.
En [40] a partir de la información en claro que el sistema guarda en el proceso de discretización, se construyen diccionarios de contraseñas más probables que permiten realizar ataques de diccionarios altamente efectivos.
En [14] se demuestra que existe correlación estadística entre las características personales del usuario (edad, sexo, etnia, educación, etc) y los patrones existentes en la contraseña que el selecciona. Se propone una métrica para medir la fortaleza de las contraseñas, a partir de los patrones que ella contiene, sin conocimiento previo de la imagen, ni estudios estadísticos de contraseñas anteiores.
En conclusión, los sistemas de autenticación gráfica del tipo *cued recall*, clasifican a los usuarios que tratan de autenticarse en dos clases: **usuario legítimo o usuario ilegítimo**, pero no son capaces de diferenciar a los usuarios dentro de una de estas clases. El resultado de [17] para contraseñas alfanuméricas nos motivó a investigar la forma de definir para contraseñas gráficas una función $P(S)$ que sea capaz de separar a los usuarios según su grado de legitimidad. Los resultados se muestran en el siguiente epígrafe y constituyen el aporte principal de este trabajo.

2. Resultados y discusión

Notaciones, hipótesis y transformación del estadígrafo de verosimilitud.

Se denotará por $I_{m\times n}$ a la imagen, por $S^* = (s_1^*, \ldots, s_k^*) \in I_{m\times n}^k$ a la contraseña, formada por k puntos/pixeles, seleccionada por el usuario legítimo en la fase de registro y por $S = (s_1, \ldots, s_k) \in I_{m\times n}^k$ a la contraseña, seleccionada por un usuario que trata de autenticarse y por $P(S) = P(S = S^*)$ a la probabilidad de que $S = S^*$.
Para simplificar el modelo, se asume que los k puntos de la contraseña $S = (s_1, s_2, \ldots, s_k)$ son seleccionados de forma independiente. Bajo esta hipótesis de independencia el estadígrafo $L(S) = P(S) = P(S = S^*)$ de verosimilitud para S será:

$$L(S) = \pi_{i=1}^k P(s_i) = \pi_{i=1}^k P(s_i = s_i^*) \tag{1}$$

A la contraseña $S = S^*$ más probable le corresponde el máximo valor de $L(S)$($\max_{S\in I_{m\times n}^k} L(S)$) en toda la imagen. Aplicando el logaritmo a la ecuación (1) el estadígrafo quedaría como:

$$L_1(S) = \log L(S) = \log P(S) = \sum_{i=1}^{k} \log P(s_i) \tag{2}$$

Por propiedades del logaritmo $L1(S)$ toma valores negativos. Por razones de implementación, para trabajar con valores enteros positivos, se selecciona una constante $C_L \in N^*$ y se aplicará la siguiente transformación lineal:

$$L_2(S) = \left(\frac{\max_{S_0\in I_{m\times n}} L_1(S_0) - L_1(S)}{\max_{S_0\in I_{m\times n}^k} L_1(S_0) - \min_{S_0\in I_{m\times n}^k} L_1(S_0)} \right) * C_L \tag{3}$$

La transformación $L_2(S)$ es una línea recta con pendiente negativa, por tanto $L_2(S)$ es función decreciente de $L_1(S)$ tal como se presenta en (4):

$$L_2(S) = \left(\frac{-C_L}{\text{máx}_{S_0 \in I^k_{m\times n}} L_1(S_0) - \text{mín}_{S_0 \in I^k_{m\times n}} L_1(S_0)} \right) * L_1(S) + \left(\frac{C_L * \text{máx}_{S_0 \in I_{m\times n}} L_1(S_0)}{\text{máx}_{S_0 \in I^k_{m\times n}} L_1(S_0) - \text{mín}_{S_0 \in I^k_{m\times n}} L_1(S_0)} \right) \quad (4)$$

Para $S \in I^k_{m\times n}$ se obtiene $L_2(S) \in 0, \ldots, C_L$. A la contraseña S más probable le corresponde el mínimo valor $L_2(S) = 0$ y a la menos probable el máximo valor $L_2(S) = C_L > 0$.
El problema al calcular $L_2(S)$ para una contraseña $S \in I^k_{m\times n}$ se reduce a definir la forma de calcular la probabilidad $P(s_i)$ de cada punto s_i. Para simplificar las notaciones se omitirán los subíndices y se denotara por s^* a un solo punto de la contraseña legítima, por s el correspondiente punto escogido por el usuario que intenta autenticarse, por $d(s,s^*)$ a la distancia entre s y s^*, y por $P(s) = P(s = s^*)$ la probabilidad $P(s)$ de que s sea el punto s^* de la contraseña legítima.
La distancia $d(s,s^*)$ es un parámetro decisivo en el proceso de autenticación, pero debido a que no se conocen bases de datos de contraseñas, ni se conoce el comportamiento experimental de esas distancias, se aplicará un enfoque axiomático. A partir de la forma de discretización y autenticación, se determinan las condiciones esenciales que debe cumplir $P(s) = P(s = s^*)$ y se buscarán analíticamente funciones $P(s)$ que satisfagan estas condiciones.

Modelación probabilística de la autenticación en los sistemas de autenticación gráfica *cued recall*

Denotando por $RT = RT(s^*)$ a la región de tolerancia alrededor del punto s^* y por $|RT(s^*)|$ a su cardinal, entonces el proceso de autenticación en estos sistemas puede modelarse mediante una distribución uniforme dentro de la región de tolerancia. La probabilidad $P_1(s)$ de que el usuario legítimo seleccione el punto s durante la autenticación puede definirse como:

$$P_1(s) = P(s = s^*) = \begin{cases} \frac{1}{|RT(s^*)|}, & para\, s \in RT(s^*) \\ 0, & para\, s \notin RT(s^*) \end{cases} \quad (5)$$

Este modelo refleja mediante la región $RT(s^*)$ el conocimiento del usuario legítimo sobre s^*:

- Si el usuario escoge un punto $s \notin RT(s^*)$, este modelo le asigna $P_1(s = s^*) = 0$, lo cual refleja la decisión del sistema de rechazar la autenticación.
- El modelo asigna la misma probabilidad $|RT(s^*)|^{-1}$ a cada punto dentro de esa región, sin distinguir entre ellos tal como hacen los sistemas de autenticación que representa.

Cada una de las $|RT(s^*)|^k$ contraseñas formadas por k puntos seleccionados dentro de la región de tolerancia (admitiendo repetición de puntos) tienen probabilidad $P_1(S) = \frac{1}{|RT(s^*)|^k}$.
La probabilidad de la región de tolerancia será:

$$P_1(RT(s^*)) = \sum_{S \in [RT(s^*)]^k} P_1(S) = 1 \quad (6)$$

En resumen, una característica de la autenticación en estos sistemas y de este modelo asociado es que no tienen en cuenta las diferencias de los valores $d(s,s^*)$ dentro de $RT(s^*)$. Esta característica constituye una limitación pues no refleja adecuadamente el comportamiento real del usuario legítimo, que se modela en el siguiente epígrafe.

Modelación del usuario legítimo

Condiciones sobre $P(s)$. Las condiciones 1 y 2 son las comunes a cualquier modelo probabilístico, para garantizar que $P(s)$ sea una distribución de probabilidades sobre todos los puntos de la imagen.

Condición 1: $P(s) \geq 0$, para todo punto s de la imagen.

Condición 2: $\sum_s P(s) = 1$ sumando sobre todos los puntos s de la imagen.

Se espera que el usuario legítimo escoja puntos s tales que $s \in RT(s^*)$ y que los valores de $d(s,s^*)$ sean pequeños. Para reflejar ese comportamiento se impone la condición 3.

Condición 3: $P(s) = P(s = s^*) = P(d(s,s^*))$ debe ser una función estrictamente decreciente de la distancia $d(s,s^*)$, por lo cual

- $P(s)$ alcanza su máximo en $d(s,s^*) = 0$, para $s = s^*$
- $P(s)$ alcanza su mínimo dentro de $RT(s^*)$ en aquellos s tal que $d(s,s^*) = \text{máx}_{s \in RT(s^*)} d(s,s^*)$.
- $P(s)$ alcanza su mínimo dentro de $I_{m\times n}$ en aquellos s tales que $d(s,s^*) = max_{s \in I_{m\times n}} d(s,s^*)$.

En los sistemas actuales un usuario es autenticado si y solo si todos los puntos escogidos por ese usuario caen dentro de sus respectivas zonas de tolerancia, por eso $P_1(RT(s^*)) = 1$. En la práctica un usuario legítimo pueda escoger por error un punto s cercano s^*, pero $s \notin RT(s^*)$ (fuera de la región de tolerancia). Para flexibilizar el proceso de autenticación y admitir esta posibilidad se introduce en el modelo un parámetro ε tal que $P(RT(s^*)) = 1 - \varepsilon$.
Por otra parte, para que el sistema de autenticación sea efectivo el usuario legítimo debe ser autenticado con alta probabilidad, por tanto, la suma de las probabilidades de todos los puntos que pertenecen a la zona de tolerancia debe ser "muy alta", por lo cual se impone la condición 4:

Condición 4: $P(RT(s^*)) = 1 - \varepsilon$, pero con $\varepsilon \approx 0$, $\varepsilon > 0$. Teniendo en cuenta los k puntos de la contraseña la probabilidad de que el usuario legítimo sea autenticado será:

$$P[s \in RT(s^*)] = [1 - \varepsilon]^k \quad (7)$$

La probabilidad de error de autenticación para el usuario legítimo será igual a: $1-[1-\varepsilon]^k$, lo cual reafirma que el valor ε debe ser "suficientemente pequeño". Se tuvo en cuenta la hipótesis de independencia entre los puntos y además se asumió que $P(RT(s^*)) = 1-\varepsilon$ es la misma para los k puntos.

Modelo probabilístico de autenticación gráfica

Se propone una nueva función $P_2(s) = P_2(s = s^*)$ que tendrá en cuenta no solo la región $RT(s^*)$, sino además la distancia $d(s,s^*)$ dentro de esa región. Para cada punto s de la imagen se calcula utilizando la ecuación (8):

$$P_2(s) = \begin{cases} P(d(s,s^*)), & si\, s \in RT(s^*) \\ \varepsilon_{d(s,s^*)}, & si\, s \notin RT(s^*) \end{cases} \tag{8}$$

Tal que $\varepsilon_{d(s,s^*)}$ es función decreciente de $d(s,s^*)$ y además se cumple que $\sum_{s \notin RT(s^*)} \varepsilon_{d(s,s^*)} = \varepsilon = 1 - P(RT(s^*))$. Basta hallar una función $P(d(s,s^*))$ que cumpla las condiciones 3 y 4. Estas condiciones establecen que un pequeño grupo de las menores distancias $d(s,s^*)$ debe tener asignada muy alta probabilidad.

Interpretación geométrica de $P_2(s)$. En la región correspondiente a los menores valores $d(s,s^*)$, tales que $s \in RT(s^*)$, el área bajo la curva de la función buscada $P_2(s)$ debe ser igual a $P(RT(s^*)) = 1-\varepsilon \approx 1$, mientras que en la región $s \notin RT(s^*)$ que contiene a la mayoría de los puntos $s \in I_{m\times n}$, la distribución $P_2(s)$ debe tener una cola estrecha y muy larga a la derecha, el área bajo esa cola debe ser $\varepsilon \approx 0$. Estas propiedades de $P_2(s)$ orientarán su búsqueda.

Propuesta de $P(d(s,s^*))$. Se propone utilizar en este caso la función $P(d(s,s^*)) = -\ln X$ con $X \in]0,1]$ pues es conocido que es una función cóncava, estrictamente decreciente en el valor de X tomando muy altos valores para cuando $X \to 0$ y valor 0 para $X = 1$.

Para aplicar esta función se debe definir su argumento X como función de $d(s,s^*)$. Como $-\ln 1 = 0$, y $lim_{X\to 0} -\ln X = \infty$, el valor $X = 1$ debe asignársele a el punto s de la imagen más alejado de s^*. Es decir, por la condición 3 y las propiedades de $-\ln X$ el argumento $X = X(d(s,s^*))$ debe definirse de forma tal que:

$$X(d(s,s^*) = \begin{cases} 1, & si \quad d(s,s^*) = MaxD \\ 0, & si \quad d(s,s^*) = 0 \end{cases}$$

Donde $MaxD = \text{máx}_{s_M \in I_{m\times n}} d(s_M, s^*)$ es la distancia entre s^* y el punto s_M más lejano a s^* en toda la imagen (uno de los 4 vértices). Esta exigencia sobre $X(d(s,s^*))$ sugiere de inmediato tomar $X = \frac{d(s,s^*)}{MaxD}$ y garantiza que a la distancia máxima se le asigna el argumento $X = 1$ donde el logaritmo toma valor cero, mientras a las distancias más pequeñas, cercanas a cero se le asigna un alto valor de $-\ln X$, cumpliéndose así las condiciones 3 y 4 exigidas al modelo.

El resultado anterior permite definir $P(d(s,s^*))$ como:

$$P(d(s,s^*)) = -\ln \frac{d(s,s^*)}{MaxD} \tag{9}$$

Esta elección de $P(s)$ cumple que:

- $P(s_M) = P(s_M, s^*) = -\ln \frac{d(s_M,s^*)}{MaxD} = -\ln \frac{MaxD}{MaxD} = 0$, lo cual se corresponde con la condición 3.
- Si $s = s^*$, queda $d(s,s^*) = 0$ y $P(s^*,s^*) = -\ln \frac{0}{MaxD}$, por lo que el logaritmo se indefine. Este punto $s = s^*$ se debe ver como un punto de discontinuidad, al cual debe asignársele un alto valor de $P(s)$, este puede tomarse cercano o igual al de los puntos, $s \neq s^*$ más cercanos a s^*.

Para garantizar la condición $P(RT(s^*)) = 1-\varepsilon$ es necesario ser capaces de aumentar a conveniencia la probabilidad de la región de tolerancia y para lograrlo se introduce en el modelo un parámetro $C_\varepsilon = C(\varepsilon) \gg 1$, cuya selección apropiada contribuye cumplir la condición 4. Para los puntos dentro de la región de tolerancia se multiplicará la ecuación 9 por este valor C_ε. Finalmente, se definirá $P_2(s)$ como se plantea en (10):

$$P_2(s) = P_2(d(s,s^*)) = \begin{cases} -\ln \frac{1}{MaxD} * C_\varepsilon, & si\, s = s^* \\ -\ln \frac{d(s,s^*)}{MaxD} * C_\varepsilon, & si\, s \in RT(s^*)\ y\ s \neq s^* \\ -\ln \frac{d(s,s^*)}{MaxD}, & si\, s \notin RT(s^*) \end{cases} \tag{10}$$

Para garantizar las condiciones 1 y 2 basta con dividir entre la suma de los valores de $P_2(s)$ en toda la imagen:

$$P_3(s) = \frac{P_2(s)}{\sum_{s \in I_{n\times m}} P_2(s)} \tag{11}$$

Comparación de $P_3(s)$ **y** $P_1(s)$. La autenticación de estos sistemas se modela por $P_1(s)$ (ecuación 5), mientras que $P_3(s)$ (ecuación 11) por su forma de construcción modela con más exactitud el comportamiento esperado intuitivamente para el usuario legítimo, por esta razón se propone introducir $P_3(s)$ en los sistemas de autenticación, lo cual se discutirá en el epígrafe 4.

Sobre el parámetro C_ε.Para comprender la influencia del parámetro C_ε, debe tenerse en cuenta que

$$P_2(RT(s^*)) = \sum_{s \in RT(s^*)} P_2(s) \tag{12}$$

Sutituyendo (10) en (12) se obtiene:

$$P_2(s) = -\ln \frac{1}{MaxD} * C_\varepsilon + \sum_{s \in RT(s^*)\ y\ s \neq s^*} -\ln \frac{d(s,s^*)}{MaxD} * C_\varepsilon \tag{13}$$

Por lo tanto, si analizamos el comportamiento del parámetro c_ε en la ecuación (13), si se aumenta el valor de C_ε, aumentará la probabilidad $P_3(RT(s^*))$ (ecuación 11) de la región de tolerancia y disminuye ε como se presenta en la ecuación (14).

$$\varepsilon = 1 - P_3(RT(s^*)) \tag{14}$$

Una forma de aumentar la influencia del parámetro C_ε, en la diferencia de las probabilidades dentro de la región de tolerancia es tomar también a C_ε como una función decreciente de la distancia $d(s,s^*)$. Se propone emplear la ecuación (15).

$$C_\varepsilon = C(d(s,s^*),\alpha) = \frac{10^\alpha}{d(s,s^*)^2} \tag{15}$$

Al aumentar el valor del parámetro α aumentan $C(d(s,s^*),\alpha)$, $P_2(RT(s^*))$, $P_3(RT(s^*))$ y disminuye el valor de ε. Esto permitirá aumentar a conveniencia en el modelo el valor de $P_3(RT(s^*))$ para ajustarlo a los observados experimentalmente. Esta opción se aplicará en los experimentos del siguiente epígrafe.

3. Validación experimental del Modelo

Se comprobará que la expresión $P_3(s)$, satisface las condiciones 3 y 4 exigidas al modelo. La dimensión de la imagen empleada es 441×331. Se utilizaron 2 tamaños de región de tolerancia 9×9 y 21×21 pixeles, por ser la más pequeña y mayor respectivamente, encontradas en la bibliografía consultada. Se utilizó la ecuación (15) para calcular C_ε y se halló experimentalmente que para $\alpha = 5$ se cumple que $C(d(s,s^*),\alpha) \gg 1$ y además $\varepsilon \approx 0,\ \varepsilon > 0$.

Las probabilidades como función decreciente de las distancias. Para $s^* = (50,50)$, se calculó $P_3(s)$ para todos los puntos s de la imagen. La figura 1 muestra la distribución $P_3(s)$ obtenida, se observa que $P_3(s) = P_3(d(s,s^*))$ decrece al aumentar $d(s,s*)$, con una cola muy larga a la derecha. Ese comportamiento se mantiene al cambiar s^*, lo que confirma que $P_3(s^*)$ satisface la condición 3.

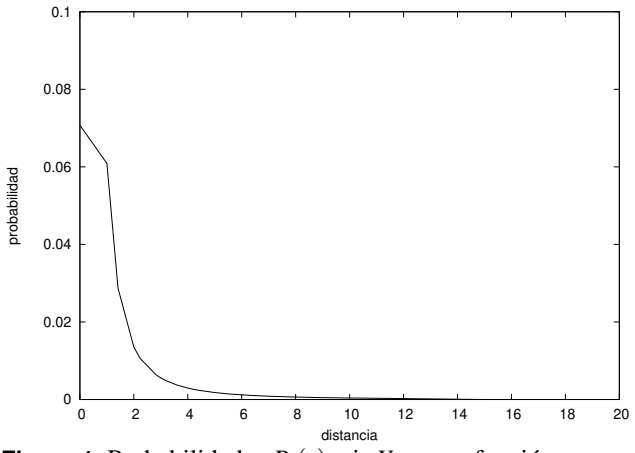

Figura 1. Probabilidades $P_3(s)$, eje Y, como función decreciente de las distancias $d(s,s^*)$, eje X , en toda la imagen.

Alta Probabilidad de la pequeña región de tolerancia. En una imagen de 441×331, las regiones de tolerancia de 9×9 y 21×21 tienen según $P_3(RT(s^{)})$ una alta probabilidad $P_3(RT(s^*)) = 1 - \varepsilon$ igual a 0.95 y 0.99 respectivamente, a pesar de que contienen solo un pequeño porcentaje de los puntos de la imagen). Los resultados se presentan en la tabla 1, confirmando así que $P_3(s^*)$ cumple la condición 4. En trabajos futuros se debe comparar estos valores teóricos de la probabilidad de RT, con valores observados en la práctica y modificar si es necesario las probabilidades teóricas (mediante los parámetros del modelo) para lograr un buen ajuste.

Tabla 1. Porcentaje de puntos en la región de tolerancia y su probabilidad en una imagen de 441×331

$\lvert\mathbf{RT}(\mathbf{s}^*)\rvert$	$\mathbf{9 \times 9}$	$\mathbf{21 \times 21}$
$\frac{\lvert\mathbf{RT}(\mathbf{s}^*)\rvert}{\lvert\mathbf{I_{m\times n}}\rvert}$	0,05	0,302
$\mathbf{P_3(RT(s^*))}$	$0,95 = 1 - 0,5$	$0,99 = 1 - 0,01$

Aplicación del modelo como métrica de autenticidad del usuario

A partir de las probabilidades $P_3(s)$ calculadas mediante el modelo propuesto, se empleará el estadígrafo $L_2(S)$ de verosimilitud para medir el nivel de autenticidad del usuario que intenta autenticarse con la contraseña S.

Diseño del experimento. Se empleó una imagen de dimensión 441×331 pixeles, una región de tolerancia de 21×21 pixeles. Con contraseña $S^* = (s_1^*,\ldots,s_5^*)$ [1]. Se realizaron en total 4000 intentos de autenticación divididos en 4 grupos (1000 intentos en cada grupo). Los grupos intentan modelar tipos de usuarios tales que el nivel de autenticidad es homogéneo dentro de cada grupo, pero difiere entre los grupos.

El objetivo del experimento es comprobar si los valores de $L_2(s)$ son homogéneos dentro de los grupos (pequeña varianza) y difieren entre ellos (diferentes valores esperados) pues en ese caso $L_2(s)$ podría emplearse para reconocer a que grupo pertenece un usuario.

Grupo $\mathbf{G_1}$: La contraseña $S = (s_1,\ldots,s_5)$ propuesta por el usuario que intenta autenticarse está formada por 5 puntos dentro de la región de tolerancia que además estarán todos muy cercanos al punto correspondiente de la contraseña S^*, en particular $d(s,s^*) \leq 5$. Este grupo simula a un usuario legítimo, que recuerda muy bien su contraseña.

Grupo $\mathbf{G_2}$: Los 5 puntos $S = (s_1,\ldots,s_5)$ estarán dentro de la región de tolerancia pero lejos del punto correspondiente de la contraseña S^* , a una distancia $5 \leq d(s,s^*) \leq 10$. Simula a un usuario legítimo, que no recuerda muy bien su contraseña.

Grupo 3 : Modelación de un usuario ilegítimo con más información, al menos uno de los puntos está dentro de la región de tolerancia, con una distancia tal que

[1] $S^* = ((50;50),(100;00),(150;150),(200;200),(250;250))$

$5 < d(s,s^*) \leq 10$, y los restantes puntos fuera de la región de tolerancia, pero cercana a ella, a una distancia $15 \leq d(s,s^*) \leq 20$.

Grupo G_4 : Los 5 puntos seleccionados por el usuario que intenta autenticarse están siempre fuera de la zona de tolerancia de 21×21 pixeles, pero muy cercanos a ella, a una distancia $15 \leq d(s,s^*) \leq 20$. Usuario ilegítimo, que posee alguna información sobre la contraseña.

En cada grupo, los pixeles de la contraseña se escogieron aleatoriamente dentro del rango de distancias que caracterizan al grupo. Los métodos actuales de autenticación gráfica, autenticarán a los usuarios de los grupos 1 y 2, pero sin distinguir la diferencia entre ellos, mientras rechazan a usuarios de grupos 3 y 4.

Se espera que para las contraseñas $S = (s_1, \ldots, s_k)$, de los 4 grupos, anteriores, el modelo propuesto les asigne valores distinguibles del estadígrafo $L_2(S) = P(S) = \pi_i^k P(s_i)$ lo cual justificará su empleo como una métrica que permite cuantificar el nivel de autenticidad del usuario, reconociendo a cuál grupo pertenece.

Para que el modelo sea efectivo, se espera que los valores de $L_2(S)$ tengan un comportamiento creciente entre los grupos $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, $\mathbf{G_4}$ en ese orden.Los resultados del experimento se muestran en la figura 3 y la tabla 2.

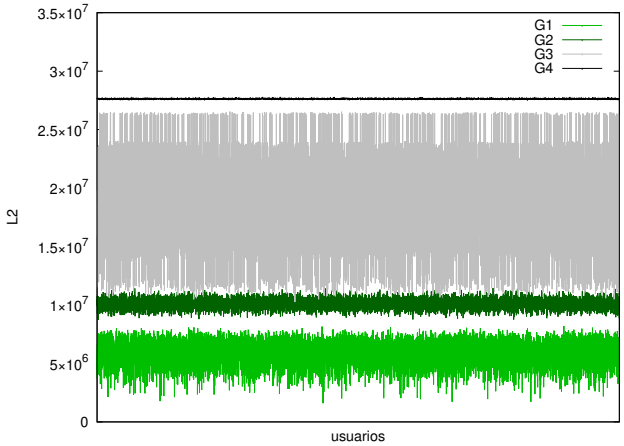

Figura 2. Valores de $L_2(s)$ en los grupos $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, $\mathbf{G_4}$

Discusión de los resultados. Se observa que el estadígrafo $L_2(s)$ aporta información sobre la autenticidad del usuario. A menor valor de $L_2(s)$, más cerca están los puntos propuestos por el usuario de los puntos de la contraseña y por tanto más confiabilidad existe sobre la autenticidad del usuario.

El valor de $L_2(S)$ logra distinguir claramente los 4 tipos de usuarios correspondientes a los grupos $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, $\mathbf{G_4}$. Como se esperaba, los usuarios legítimos que recuerdan muy bien su contraseña($\mathbf{G_1}$), muestran los menores valores de $L_2(S)$. Los grupos 3 y 4 de usuarios ilegítimos se distinguen claramente de los grupos 1 y 2 de usuarios legítimos.

El Grupo $\mathbf{G_3}$, a pesar de tener algunos puntos dentro de la región de tolerancia, es reconocido correctamente por el modelo la casi totalidad de las veces como un usuario ilegítimo, se observan algunos caso en que se confunde con un usuario del grupo $\mathbf{G_2}$.

Lo más importante a destacar es que los valores de $L_2(S)$ se diferencian entre los grupos $\mathbf{G_1}$ y $\mathbf{G_2}$, a pesar de que ambos son igualmente autenticados e indistinguibles para los sistemas actuales de autenticación gráfica.

Estos resultados experimentales validan la efectividad práctica del modelo para cuantificar el nivel de autenticidad de los usuarios, significa que el modelo propuesto permite diferenciar a los usuarios que son autenticados por el sistema, asignándoles diferentes grados de autenticidad y sugiere que tal vez el modelo pueda ser empleado para detectar los ataques de diccionario que realizan un pronóstico de la contraseña, lo cual es un aspecto que debe ser investigado.

Los resultados ilustrados en la figura 3 se resumen en la figura 3 y en la tabla 2.

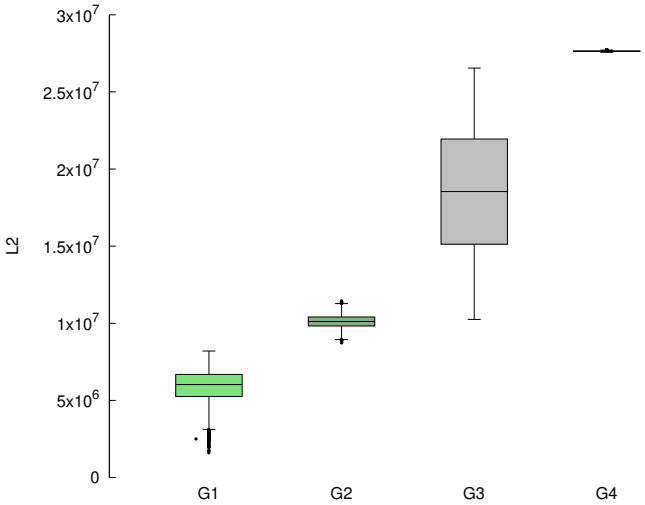

Figura 3. Comportamiento de $L_2(s)$ en los grupos $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, $\mathbf{G_4}$

Tabla 2. Valores máximo y mínimo de $L_2(S)$, con sus frecuencias por grupos.

Grupo	Máximo de $L_2(S)$)	Frecuencia	Mínimo de $L_2(S)$)	Frecuencia
$\mathbf{G_1}$	8201503	1	1602158	1
$\mathbf{G_2}$	11453131	1	8735939	1
$\mathbf{G_3}$	27743039	1	27568825	1
$\mathbf{G_4}$	26546809	1	10252333	1
$\cup_i^4 \mathbf{G_i}$	27743039	1	1602158	1

En la figura 3 se observa que los valores esperados de $L_2(s)$ difieren entre los grupos, además se aprecian diferencias notables entre la magnitud de las varianzas, siendo el grupo $\mathbf{G_3}$ el de mayor varianza, ya que corresponde a un usuario con cierta información sobre la clave legítima, que escoje

algunos puntos dentro y otros fuera de la región de tolerancia. Los valores de media y varianza permiten distinguir el comportamiento medio en cada grupo, lo cual justifica el empleo de $L_2(s)$ para reconocer a que grupo pertenece un usuario, es decir como una métrica para evaluar su nivel de autenticidad. Estos resultados sugieren definir 3 umbrales y 4 regiones en las que se puede clasificar a los usuarios autenticados según su nivel de autenticidad, los cuales se presentan en la tabla 3.

Tabla 3. Umbrales para los valores de $L_2(s)$

Región	$L_2(S))$	Nivel de Autenticidad
$\mathbf{R_1}$	$0 \leq L_2(s) < 8201503$	Alto
$\mathbf{R_2}$	$8201503 \leq L_2(s) < 11453131$	Medio
$\mathbf{R_3}$	$11453131 \leq L_2(s) < 26546809$	Bajo
$\mathbf{R_4}$	$26546809 \leq L_2(s)$	Muy bajo

Los usuarios de la región $\mathbf{R_4}$ corresponden a los no autenticados y las regiones $\mathbf{R_1}$, $\mathbf{R_2}$ y $\mathbf{R_3}$ a los autenticados, se propone investigar la forma de aprovechar este nuevo conocimiento sobre su nivel de autenticidad, por ejemplo, a los de la región $\mathbf{R_3}$ se les puede exigir alguna información adicional antes de autenticarlo.
Se requieren nuevos experimentos con mayor número de muestras y diversidad de usuarios para evaluar con más exactitud el nivel de precisión alcanzado por el modelo y la conveniencia de redefinir los umbrales. Esta es una dirección de trabajo futuro.

4. Introducción práctica del modelo en los sistemas de autenticación gráfica

Para obtener $P_3(s) = P_3(s = s^*) = P_3(d(s,s^*))$, es importante destacar que se requiere conocer $S^* = (s_1^*, \ldots, s_k^*)$ y $S = (s_1, \ldots, s_k)$ para calcular $d(s,s^*)$. La contraseña $S = (s_1, \ldots, s_k)$ es introducida por el usuario, pero los sistemas de autenticación gráfica no guardan directamente el valor $S^* = (s_1^*, \ldots, s_k^*)$, su desconocimiento parece impedir la introducción práctica del modelo.
La dificultad anterior puede resolverse, en algunos casos, teniendo en cuenta las propiedades de los métodos de discretización. En discretización centrada y optima, una vez que un usuario es autenticado, se puede calcular, a cada uno de los puntos s_i de la contraseña $S = (s_1, \ldots, s_k)$ propuesta por este usuario, su celda de discretización.
Por la forma en que se realiza la discretización, el centro de esta celda es el punto s_i^* de la contraseña $S^* = (s_1^*, \ldots, s_k^*)$. Conociendo s_i^* y el punto s_i seleccionado por el usuario, se puede calcular $d(s_i, s_i^*) \forall i = \overline{1, \ldots, k}$ y se calcula la probabilidad $P_3(s_i)$ asociada a esa distancia. Con las k probabilidades se calcula el estadígrafo $L_2(s)$ que permite valorar el nivel de autenticidad del usuario.
El enfoque anterior no es aplicable en discretización robusta pues no se pueden recobrar los k puntos de la contraseña. Se propone para futuros trabajos, desarrollar una aplicación para incluir este criterio en sistemas de autenticación gráfica que emplean discretización óptima o centrada.

Conclusiones y trabajo futuro

Se obtuvo un nuevo modelo probabilístico que permite distinguir/clasificar a los usuarios asignándoles una probabilidad $P_3(S)$ de que su contraseña S sea seleccionada por el usuario legítimo.Ademas se comprobó experimentalmente que el modelo cumple las condiciones esperadas para el usuario legítimo y que el estadígrafo propuesto es capaz de distinguir a los usuarios según su nivel de autenticidad.
A los usuarios autenticados con baja probabilidad de ser legítimos, se le puede solicitar una autenticación adicional lo cual aumentaría la seguridad del sistema de autenticación. Este modelo es aplicable on-line en los sistemas de autenticación que emplean discretización centrada y óptima. No puede ser empleado en los sistemas que emplean discretización robusta pues la información que requiere no está disponible y es necesario hallar la forma de hacerla accesible sin comprometer la seguridad del sistema.
Algunas direcciones de trabajos futuros son:

- Desarrollar una aplicación que implemente el modelo y pueda incorporarse en sistemas de autenticación gráfica aumentando su seguridad.
- Obtener valoraciones sobre la eficiencia de esta aplicación para evaluar la afectación a la usabilidad de los sistemas en que se utilice.
- Evaluar el modelo en escenarios más complejos, por ejemplo, simulando a usuarios ilegítimos que tratan de pronosticar la contraseña por ataques de diccionarios, los cuales pueden llegar a tener varios puntos dentro de la región de tolerancia y hasta cerca del punto de la contraseña.
- Caracterizar el comportamiento de $L_2(s)$ en esos nuevos escenarios y evaluar si puede detectar esos ataques.
- Usar el modelo para calcular k distribuciones de probabilidades, una por cada punto de la contraseña y cada una definidas sobre todos los puntos de la imagen, considerando la dependencia entre los k puntos.
- Explorar el empleo de otros tipos de funciones $P(s)$ que cumplan las condiciones impuestas.
- Cambiar el enfoque axiomático, para estimar la distribución a partir de muestras de las distancias $d(s,s_k^*)$, en particular investigar el ajuste de la distribución de las distancias $d(s,s_k^*)$ para los usuarios legítimos, mediante alguna ley de potencia como la Ley de *Zipf-Mandelbrot*.

.

Referencias

[1] Iso-9126 software product evaluation – quality characteristics and guidelines for their use.

[2] S Aggarwal, S Houshmand, and R Flood. Probabilistic password cracking system, 2016.

[3] Bin. B., D.W. Zhu, M. Yang, and J.. Yan. Security implications of password discretization for click-based graphical passwords. In *WWW*, 2013.

[4] K. Bicakci. *Optimal Discretization for High-Entropy Graphical Passwords. T.* PhD thesis, OBB University of Economics and Technology, Ankara, Turkey, 2007.

[5] R. Biddle, S. Chiasson, and P.C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computer Survey*, 44(4):19:1–19:41, September 2012.

[6] J.C. Birget, D. Hong, and N Memon. Graphical passwords based on robust discretization. *IEEE Transactions on Information Forensics and Security*, 1(3), Sep 2006.

[7] J. Blocki, H. Benjamin, and Samson Z. On the economics of offline password cracking. In *IEEE Symposium on Security and Privacy*, volume 1, 2018.

[8] E. B Borrego, P.E. Navarro, and C.M. Legón. Debilidades de los métodos de discretización para contraseñas gráficas. In Instituto de Criptografía. Sociedad Cubana de Matemática y Computación., editor, *IV Seminario Científico Nacional de Criptografía. Universidad de la Habana*, 2018.

[9] W. E. Burr, D. F. Dodson, E. M. Newton, R. A. Perlner, W. T. Polk, S. Gupta, and E. A. Nabbus. Electronic authentication guideline: Recommendations of the national institute of standards and technology -. Technical report, U.S. Department of Commerce and National Institute of Standards and Technology, USA, 2012.

[10] C. Castelluccia, M. DÃŒrmuth, and D. Perito. Adaptive password-strength meters from markov models. In *19th Annual Network & Distributed System Security Symposium*, San Diego, United States, Feb 2012.

[11] S. Chiasson, J. Srinivasan, R. Biddle, and P. C. van Oorschot. Centered discretization with application to graphical passwords. In *Usability, Psychology, and Security*, 2008.

[12] H.C Chou, H.C. Lee, H. J Yu, F.P. Lai, K.H. Huang, and C.W. Hsueh. Password cracking based on learned patterns from disclosed passwords. *International Journal of Innovative Computing, Information and Control*, 9(2):821–839, 2013.

[13] X.C. de Carnavalet and M. Mannan. From very weak to very strong: Analyzing password-strength meters. In *21st Annual Network and Distributed System Security Symposium*. The Internet Society, 2014.

[14] M. Devlin, J.R. Nurse, D. Hodges, M. Goldsmith, and S. Creese. Predicting graphical passwords. In *Proceedings of the Third International Conference on Human Aspects of Information Security, Privacy, and Trust -*, volume 9190, pages 23–35, New York, NY, USA, 2015. Springer-Verlag New York, Inc.

[15] W. Ding, Haibom C., P. Wang, X. Huang, and G. Jian. Zipfs law in passwords. *IEEE Transactions on Information Forensics and Security*, 12(11):2776 – 2791, 2017.

[16] A. E. Dirik, L. F. Cranor, and J.C Birget. Modeling user choice in the passpoints graphical password scheme. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, pages 20–28, New York, NY, USA, 2007. ACM.

[17] D. Freeman, S. Jain, M. Dürmuth, B. Biggio, and G. Giacinto. Who are you? A statistical approach to measuring user authenticity. In *NDSS*. The Internet Society, 2016.

[18] M. Golla and M. DÃŒrmuth. On the accuracy of password strength meters. In *Conference on Computer and Communications Security*, Toronto, Canada, Oct 2018.

[19] S. Houshmand and S. Aggarwal. Building better passwords using probabilistic techniques. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pages 109–118, New York, USA, 2012. ACM.

[20] L. Jue Min, L. Yong Hao, N. Huey Wen, T. Soon Guan, Y. Li Ho, A. Azman, and L. Siong Hoe. Comparison of graphical password using iso 9126. *Advanced Science Letters*, 4:400–407, 2016.

[21] P. Karmajit, N. Bhushan, D. Prasad-Mishra, and P. Priyadarsini-Satapathy. Cued-click point graphical password using circular tolerance to increase password space and persuasive features. In *Proceedings of International Conference on Communication, Computing and Virtualization*, volume 79, pages 561 – 568, 2016.

[22] C.M. Legón, P.E. Navarro, E.A. Borrego, O. Rodríguez, and R. Socorro. Modelos probabilísticos de contraseñas alfanuméricas. In *IV Seminario Científico Nacional de Criptografía*. Universidad de la Habana, Noviembre 2018.

[23] J. Ma, W. Yang, M. Luo, and N. Li. A study of probabilistic password models. In *IEEE Symposium on Security and Privacy*, volume 1, pages 689–704, 2014.

[24] D. Malone and K. Maher. Investigating the distribution of password choices. *Cryptography and Security*, 2011.

[25] D. Malone and K. Maher. Investigating the distribution of password choices. In *Proceedings of the 21st International Conference on World Wide Web*, pages 301–310, New York, NY, USA, 2012. ACM.

[26] W. Melicher, B. Ur, S.M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F Cranor. Fast, lean, and accurate: Modeling password guessability using neural networks. In *Proceedings of the 25th USENIX Security Symposium*, pages 10–12, 2016.

[27] L.L. Morales and C.M. Legón. Estimación de la fortaleza de las contraseñas. Technical report, Facultad de Ingeniería Informática. Universidad Tecnológica de la Habana, 2014.

[28] O. Rodríguez, C.M. Legón, and R. Socorro. Seguridad y usabilidad de los esquemas y técnicas de autenticación gráfica. *Revista Cubana de Ciencias Informáticas*, 12(Especial UCIENCIA):13–27, Sep 2018.

[29] S. Salehi-Abari, J. Thorpe, and P.C. van Oorschot. On purely automated attacks and click-based graphical password. In *Computer Security Applications Conference*, 2008.

[30] A. Shaikh, R. Pathan, R. Patel, and S. Rukaiya. Implementation of authentication using graphical password cloud computing. *International Research Journal of Engineering and Technology*, 5(5), 2018.

[31] Sunil Shendage Swapnil, Prakash Dhainje, and Shivaji Yevale Ramesh. Cued click points: Graphical password authentication technique for security. *International Journal of ComputerScience and Information Technologies*, 5(2), 2014.

[32] J. Shouling, Y. Shukun, W. Ting, L. Changchang, L. Wei-Han, and B. Raheem. Pars: A uniform and open-source password analysis and research system. In *Proceedings of the 31st Annual Computer Security Applications Conference*, pages 321–330, New York, USA, 2015. ACM.

[33] M. N. Todd. An investigation of machine learning for password evaluation. Master's thesis, Arizona State University, 2016.

[34] A Toledo, M García, C. M. Legón, and J. L. Morales. Caracterización de un atacante a sistemas de autenticación por contraseña. In *Segurmática*, 2014.

[35] E. Walkup. The password problem. Technical report, Sandia National Laboratories, Albuquerque, United States, 2016.

[36] D. Wang and P. Wang. On the implications of zipfs law in passwords. In *European Symposium on Research in Computer Security*, pages 11–131, 2016.

[37] M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek. Password cracking using probabilistic contex-free grammars. In *IEEE Symposium on Security and Privacy*, pages 391–405, 2009.

[38] D. L. Wheeler. zxcvbn: Low-budget password strength estimation. In *Proceedings of the 25th USENIX Security Symposium*, SEC'16, pages 157–173, Berkeley, CA, USA, 2016. USENIX Association.

[39] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. Passpoints: design and longitudinal evaluation of a graphical password system. *International Journal of Human Computers Studies*, 63(1):102–127, 2005.

[40] Bi. B. Zhu, D. Wei, M. Yang, and J. Yan. Security implications of password discretization for click-based graphical passwords. In *Proceedings of the 22Nd International Conference on World Wide Web*, pages 1581–1591, New York, USA, 2013. ACM.