# CAPTCHA para la Seguridad de la Información en las Aplicaciones Web

Zenón Amado Callisaya Uchani
*Postgrado en Informática*
*Universidad Mayor de San Andrés*
La Paz - Bolivia
amado113@gmail.com

***Resumen*—Los CAPTCHA, las pruebas humanas interactivas más utilizadas, aparecen como un método de diferenciación entre usuarios humanos y máquinas para brindar seguridad a la información en internet y evitar ataques DDoS y DoS. Por ello, se hace una presentación introductoria de las pruebas humanas interactivas, para luego profundizar en los CAPTCHA, donde se trata su evolución, su clasificación, especificando sus técnicas de generación, algunas de sus aplicaciones y los tipos de ataques a los que se ven sometidos. Después, se tratan algunas repercusiones que han tenido sobre la usabilidad y accesibilidad de servicios de internet de empresas públicas para usuarios humanos, lo que lleva a plantearse la inquietud que si los CAPTCHA, son una solución para las aplicaciones web.**

***Palabras clave—captcha, seguridad de información, pruebas humanas interactivas, usabilidad y accesibilidad***

## I. INTRODUCCIÓN

En nuestro país con el crecimiento de la tecnología cada vez los usuarios buscan el acceso a internet es así que se debe pensar en los ataques automatizados llamados bots o robots que se hacen pasar por humanos. Las pruebas humanas interactivas conocidas también como CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) por sus siglas en inglés, basado en la inteligencia artificial proponen retos basados en texto o imágenes, sonidos para personas con discapacidad; estos retos son fáciles de resolver para los humanos y difícil para las computadoras pero también implica la limitación de accesibilidad y la usabilidad a las páginas web para usuarios que no cuentan con conocimientos o que no se los puede resolver rápidamente [1].

Es por ello que se pretende realizar un acercamiento a esta estrategia de seguridad de la información, y algunas de sus repercusiones y uso en sitios web, no desde una posición terminante, sino como una incitación sobre la seguridad en las aplicaciones web que tienen información importante.

El presente estudio se plantea para realizar las pruebas humanas, clasificación, aplicaciones y ataques a los que están expuestos; las aplicaciones web debido al aumento de dificultad en su lectura para enfrentar los avances de los bots.

### *A. Pruebas Humanas Interactivas*

En el marco de la utilización masiva de las tecnologías de la información y las comunicaciones (TIC), y debido a la cantidad de información personal e institucional en internet que requiere ser protegida, se han establecido técnicas para probar que se es humano al momento de ingresar o utilizar cierta información. Para ello, la biometría, en conjunto con otras disciplinas, tiene un interesante reto.

### *B. CAPTCHA*

Los métodos más utilizados para diferenciar automáticamente máquinas y humanos son los CAPTCHA (por sus siglas en inglés, *Completely Automated Public Turing test to tell Computers and Humans Apart*), los cuales están basados en inteligencia artificial. Son similares a la prueba de Turing, con la diferencia de que el juez es un equipo cuyo objetivo es hacer preguntas que los usuarios humanos podrán responder, pero las máquinas actuales no [4].

Los CAPTCHA son unos "*puzzles*" o rompecabezas que los *webmasters* incluyen en su sitio web para asegurarse de que los visitantes que quieren interactuar con el contenido son personas, y no robots que tratan de registrarse en el sitio web, incluir comentarios en blogs, etc. [2].

### *C. Tipos de CAPTCHA*

Los CAPTCHA, se clasifican fundamentalmente en métodos basados y no basados en OCR (*Optical Character Recognition*), por sus siglas en inglés. En los primeros, los más conocidos y usados en la actualidad, se presenta la imagen de una palabra con una distorsión de diversos efectos, la cual debe ser escrita por el usuario y, debido a los efectos pictóricos, no podrá ser reconocida por el equipo. Para su creación, se acostumbra llevar a cabo un procedimiento general (elegir una palabra de un diccionario predefinido, aplicar un formato a la palabra y convertirla en imagen y degradar la composición mediante distorsión), en el que se diferencian los métodos de generación por medio de los algoritmos de elección de palabras y diccionarios, el formato aplicado a los caracteres y las degradaciones realizadas sobre las imágenes [5].

En el segundo grupo se presentan imágenes cuyos retos implican, para resolverlos, dar clic en una zona específica de la imagen, identificar una serie en las imágenes, mover algún componente de ella, o incluso formar cadenas de caracteres con las iniciales de los objetos representados. Una de las principales razones por las que los CAPTCHA basados en imágenes son vulnerables frente a ataques, es que en casi ninguna de las técnicas existentes estas son distorsionadas para evitar el reconocimiento de una máquina. Adicionalmente, existen los sistemas basados en sonido (o los sistemas de audio), los cuales, por lo general, requieren de reconocimiento de voz para resolver una tarea [6].



**Para referenciar este artículo (IEEE):**
[N] Z. Callisaya, «CAPTCHA para la Seguridad de la Información en las Aplicaciones Web,» *Revista PGI. Investigación, Ciencia y Tecnología en Informática*, nº 7, pp. 112-115, 2020.

En el Tabla I se presenta una comparación realizada a los principales CAPTCHA.

TABLA I. COMPARACIÓN DE LOS DIEZ MEJORES CAPTCHA

| Nombre | CAPTCHA | | | |
|---|---|---|---|---|
| | Seguridad alta | Tecnología soportada | Sonido | Imagen |
| TheCAPTCHA | X | PHP | | X |
| Securimage | X | PHP | X | X |
| ReCAPTCHA | X | PHP | X | X |
| JCAPTCHA | X | JAVA | X | X |
| NuCAPTCHA | X | JavaScript, HTML5, Flash | X | X |
| ProtectedWebForm.com | X | PHP | X | X |
| Free CAPTCHA-Service | X | Multi-platform | X | X |
| CAPTCHA Confident | X | Multi-platform | X | X |
| IronClad CAPTCHA | X | PHP | X | X |
| BotDetect CAPTCHA | X | ASP, PHP, Perl | X | X |

### *D. Aplicaciones de los CAPTCHA*

Los CAPTCHA, como sistemas que diferencian personas de máquinas, tienen multiplicidad de aplicaciones relacionadas con la seguridad en diferentes sectores de la sociedad actual. Algunas de ellas son:

- Encuestas por internet.
- Servicios gratuitos de correo electrónico.
- Motores de búsqueda.
- Correo electrónico en cadena y spam.
- Prevenir ataques de diccionario.

### *E. Ataques a los CAPTCHA*

Tal y como sucede con cualquier tipo de aplicación usada para prevenir el uso ilícito de un servicio, los CAPTCHA son susceptibles a ataques, y dado su uso principal en la protección en el registro a correo web mediante formularios, la mayor cantidad de ataques proviene de los *spammers*, quienes sostienen una constante búsqueda de registro y uso gratuito de múltiples cuentas de correo. Aunque aquí se mencionarán algunos de los ataques más conocidos, es importante tener en cuenta que estos se mantienen en constante desarrollo, al igual que los esfuerzos por prevenirlos [3].

- Romper un CAPTCHA resolviéndolo automáticamente. Existen múltiples programas destinados a solucionar los caracteres que aparecen en un CAPTCHA de forma automática.
- Método de resolución semiautomático. Este introduce la clave y completa el proceso de registro, permitiendo así a los creadores de spam el acceso gratuito a las cuentas de correo, desde las cuales comienzan a distribuir los mensajes entre miles de cuentas de correo legítimas [3].
- Vulnerar el algoritmo de generación de CAPTCHA. La probabilidad de la vulneración se incrementa cuando un proveedor de servicio usa un sistema de generación de CAPTCHA muy antiguo, a diferencia de los recaptchas, muy seguros en este aspecto.
- Resolver el CAPTCHA manualmente. Si los CAPTCHA son creados para ser resueltos por humanos, la forma más sencilla de solucionarlo es por un humano.

### *F. CAPTCHA sonoro*

Luego de realizar un estudio experimental sobre los CAPTCHA sonoros para descubrir la diferencia entre el nivel de comprensión del ser humano en distintas pruebas con sonidos que los usuarios deben descifrar, y el de un reconocedor de voz automático [7]. Se mencionan los siguientes:

- La tecnología actual de reconocimiento de voz está bastante lejos de poder resolver CAPTCHA sonoros con garantías
- Los CAPTCHA sonoros son algo más complicados de resolver que los gráficos, aunque, a la vez, son bastante más seguros.
- Como los CAPTCHA son un sistema seguro que resuelve la mayoría de los usuarios, se presenta gran cantidad de errores en las pruebas con usuarios.
- Los parámetros que provocan en mayor medida los fallos son la combinación de una menor distancia entre la pronunciación de los números y la velocidad de pronunciación.

## II. MÉTODOS

El presente estudio es de tipo analítico, prospectivo, cuasi experimental. Se ha creado un grupo experimental compuesto por 20 personas a los que se les pidió resolver CAPTCHA. Y luego se entrevistó a cada participante para obtener su valoración de la usabilidad y utilidad del sistema CAPTCHA.

Se buscó aplicaciones web que posean información importante al usuario, gran parte de los ciudadanos con el avance de la tecnología, la información importante la tienen en la instituciones públicas, es así que se escogió a dos instituciones importantes que tienes información en sus aplicaciones web, las que son de Servicio de Impuestos Nacionales (SIN) y Servicio Nacional del Sistema de Reparto (SENASIR), en estos sitios web podemos encontrar servicios online u otro tipo de servicios que le ayudan a los usuarios con la información personal que tienen.

También se tomó un grupo de 10 personas para el SIN y 10 personas para SENASIR que sus datos personales y acceso a la información en sus bases de datos. El experimento se realizó en los formularios de ingreso a la información que las aplicaciones web ofrecen a sus usuarios. Navegando por la página del Servicio de Impuestos Nacionales se pudo observar que tienen CAPTCHA para el ingreso a los servicios web “La oficina Virtual”, Mientras que el Servicio Nacional de Reparte no tiene algún CAPTCHA para la búsqueda de información de sus usuarios.

Para encontrar resultados se basó en la facilidad de ingreso a sus aplicativos en la web, el tiempo con el que resuelve el CAPTCHA cada uno de los participantes y la importancia de la información que poseen en los servicios que ofrecen.

## III. Resultados

Para verificar los resultados obtenidos se realizó una encuesta donde se registra el uso de la información de los datos personales de usuarios del SIN y SENASIR que están en la web, posteriormente se tabula y resume con un "SI" o con un "NO" el promedio de cada respuesta que el participante da a las preguntas indicadas y la facilidad de acceso rápido con la que cuentan.

En el Tabla II se presenta una comparación de la facilidad de acceso a la información.

Tabla II. Acceso de la información según la dificultad de acceso

| | Servicio de acceso | Acceso rápido a la información | Información relevante de uso personal | Uso de CAPTCHA |
|---|---|---|---|---|
| SIN | SI | NO | SI | SI |
| SENASIR | SI | SI | SI | NO |

El cuadro la dificultad de acceso a la información muestra lo dificultoso el ingreso al SIN, pero no es tanto así en el SENASIR, Entonces comprobamos porque demoran al ingreso de su información, se mide en base al tiempo que demora en la solución del CAPTCHA, donde se mide la dificultad de ingreso.

En el Tabla III se presenta el tiempo para medir el acceso a la solución del CAPTCHA.

Tabla III. Tabla comparativa de tiempos de acceso del captcha

| Tiempo para medir la dificultad de acceso por CAPTCHA | |
|---|---|
| Fácil Acceso | De 1 a 3 minutos |
| Dificultad Media | De 3 a 5 minutos |
| Complicado | De 5 a 10 minutos |

Cuando definimos la forma de medir, procedemos con experimento de obtención del tiempo que demoran los usuarios al ingresar a la información.

En la Tabla IV se presenta una comparación de la facilidad de acceso a la información.

Tabla IV. Resultados obtenidos para el ingreso a la información

| Tiempo de solución de los CAPTCHA | | | |
|---|---|---|---|
| Usuarios | De 1 a 3 minutos | De 3 a 5 minutos | De 5 a 10 minutos |
| SIN | 0 | 0 | 5 |
| SENASIR | 5 | 0 | 0 |

Las causas que dificultaron el acceso rápido, según indican los evaluados son la poca visibilidad de la imágenes o símbolos que aparecen en los CAPTCHA esto dificulta obtener de forma rápida la información que se hace molestoso para el usuario, pero los del SENASIR se ven más tranquilos porque pudieron obtener la información de forma rápida sin necesidad de resolver los CAPTCHA.

La accesibilidad web permite que un sitio web pueda ser visitado y utilizado de forma satisfactoria por el mayor número posible de personas. Es difícil separar la usabilidad (facilidad de uso) de la accesibilidad (facilidad de acceso); y no sólo es difícil, sino en muchos casos, innecesario. Y en sentido general, los principios de usabilidad y accesibilidad tienen como objetivo que el diseño de un sitio web permita que éste pueda ser accedido y usado por el mayor número posible de personas, independientemente de las limitaciones propias del individuo o de las derivadas del contexto de uso Es innegable que, a medida que se usan técnicas en los CAPTCHA para disminuir la posibilidad de que métodos automatizados puedan fungir como humanos, aumentando el ruido y las deformaciones de la imagen, como efecto colateral aumenta la dificultad para los usuarios, se conduce a altas tasas de error y, con ello, a mucha frustración [7].

En la prueba realizada en plataforma web del SIN y SENASIR se pudo obtener mediante CAPTCHA, demoran entre 5 a 10 minutos mientras que los usuarios del SIN es decir que el acceso es complicado y de 1 a 3 minutos los usuarios de SENASIR para ingresar que sería siendo de fácil acceso a sus aplicaciones web.

Los usuarios no siempre entienden qué acción se debe acometer. Les resulta difícil o imposible visualizar el CAPTCHA, como en el caso de los re-CAPTCHA, que proponen al usuario que transcriba una serie de letras y/o números, y en ocasiones están tan distorsionados que es muy complicado, y a menudo desconocen su utilidad.

Según el experimento realizado se resume lo siguiente:

- Los usuarios inicialmente perciben como más fáciles aquellos CAPTCHA con los que están más familiarizados.
- Los CAPTCHA pueden resueltos cuando estos sean familiarizados en otros sistemas.
- Las mayores puntuaciones en percepción de eficacia y facilidad, son los que no tenían implementados CAPTCHA.
- El CAPTCHA semiótico (interpretación de imagen: manos), resuelto por todos los participantes, debería de explorarse más, dado el alto grado de eficacia que presenta.

Pese a la dificultad que presentan los CAPTCHA se puede garantizar que la información está segura o que por lo menos tiene un medio que evita que los *bots* o robot puedan obtener a nuestra a la información.

## V. Discusión

Analizando la parte teoría se puede decir que la necesidad de proporcionar seguridad a la información personal e institucional alojada en el internet, se incrementa día a día, ante la gran cantidad y constante posibilidad de ataques informáticos. Sin embargo, con soluciones como los CAPTCHA se han logrado resultados óptimos al acceso de las aplicaciones web, aunque de manera paralela, los CAPTCHA dificultan el acceso a los usuarios en los diversos servicios en el marco de la sociedad de la información y el conocimiento, situación que también se puede lograr buenos resultados en todas las aplicaciones que requieran seguridad y estos sean implementados en todos los sistemas que manejan información importante. Lo anterior, antes que desvirtuar tal solución, pretende servir de eficiente para el debate académico sobre su efectividad integral.

Según la información obtenida sobre la usabilidad en la implementación de sitios web con CAPTCHA podría tener resultados poco óptimos por la falta de conocimientos que este causaría a los usuarios en específico con problemas visuales, pero también los resultados muestran la necesidad de poder implementar herramientas informáticas para que los sitios web no sean vulnerables a cualquier ataque, en este sentido la implementación de CAPTCHA ayudaría a mejorar la seguridad de atacantes mal intencionados como de la fuerza bruta, robots que llegarían a encontrar vulnerabilidades en los *login* o módulos de acceso que tienen como medios de seguridad los sistemas de información web y esto causaría un impacto que dañe la imagen de la institución la perdida de información.

Es cierto que los CAPTCHA no son del todo seguros, pero logran disminuir el riesgo a ser atacado por algún robot, y aunque se diga que es complicado y moroso resolverlos es solución que en un futuro todas las instituciones deberían implementar a sus sistemas web que como también se vio poseen información personal.

## REFERENCIAS

[1] Areitio, J. y Areitio, T. (2007). Análisis en torno a la tecnología biométrica para los sistemas electrónicos de identificación y autenticación. Revista española de electrónica, (630), 52-67.

[2] Cabezas, V., Sabaté, A., Vendrell, A. y Marcos, M. C. (2014). Experiencia de usuario y captchas, explorando la semiótica visual. No Solo Usabilidad, 13. Recuperado de www.nosolousabilidad.com/articulos/usabilidad captchas.htm

[3] Martínez, D. y Prieto, O. (2009). Servicios Accesibles de Acceso Exclusivamente Humano. Recuperado de http://orff.uc3m.es/bitstream/10016/10622/1/servicios_accesibles_CIBD_09.pdf

[4] Hernández, C. J. y Ribagorda, A. (2010). Pitfalls in CAPTCHA design and implementation [Errores en el diseño e implementación CAPTCHA: El CAPTCHA de Matemáticas, un estudio de caso]. Computers & Security, 29(1), 141-157.

[5] Shirali-Shahreza, M. y Shirali-Shahreza, S. (2008). Encouraging persons with hearing problem to learn sign language by Internet websites [Alentar las personas con problemas de audición para aprender el lenguaje de signos por los sitios web de Internet]. Recuperado de http://ftp.informatik.rwth- aachen.de/Publications/CEUR-WS/Vol-345/paper8.pdf

[6] Yan, J. y Salah El Ahmad, A. (2008). Usability of CAPTCHAs or usability issues in CAPTCHA design [Usabilidad de CAPTCHAs o problemas de usabilidad en el diseño de CAPTCHA].

[7] García, F.J. (2013). Experimentación con CAPTCHA sonoros (Tesis de grado). Universidad de Valladolid, Valladolid, España.