Enredándose

CAPTCHA

Fernando J. Elizondo Garza
FIME-UANL
fjelizond@hotmail.com

DE UNA DE TANTAS BATALLAS ENTRE EL BIEN Y EL MAL EN EL CIBERESPACIO

En el ciberespacio, esa realidad (virtual) que se encuentra dentro de los ordenadores y redes del mundo, se desarrollan luchas entre el bien y el mal, sea eso lo que cada quien conciba, en paralelo a las guerras físicas que la humanidad mantiene en el mundo, y al igual que el armamento de las batallas reales, las herramientas de lucha en el dominio de la informática día a día evolucionan.

La batalla de la que trataré en este artículo inició en 1994 cuando la firma de abogados Canter & Siegel especializada en inmigración, publica en un foro en Usenet un anuncio de su firma legal ofreciendo sus servicios para registrarse en una "Green Card Lottery", el cual les produjo muy buenos resultados por lo que en los siguiente 9 días enviaron 41 mensajes más, dando inició la era del SPAM.[1,2]

El Spam, fenómeno que actualmente considera diferentes tipos de medios de comunicación, es simplemente el envío masivo de mensajes no deseados a personas desconocidas, usualmente propaganda no solicitada. Este fenómeno debe su nombre al famoso jamón condimentado enlatado "Hormel´s Spiced Ham", que gracias a su popularidad se le empezó a llamar Spam,[3,4] y se volvió un nombre genérico del producto. Fue el grupo cómico de los Monty Python los que gracias a un sketch cantado en el que repetían la palabra spam hasta el cansancio, en algo así como dando a entender que para comer había papas con huevo, papa con tomate, papas, papas y más papas, volvieron la palabra spam como sinónimo de algo que se encuentra uno siempre y hasta el hastío, terminando este término aplicándose a los mensajes que actualmente nos abruman.

Los correos electrónicos no solicitados se volvieron un problema por el tiempo que invierten los receptores en separar lo bueno de lo malo y en depurar sus cuentas. Así que los usuarios de la red empezaron a usar programas que filtraran dichos correos, al tiempo que los spammers (los generadores de spam) desarrollaron ciber robots, en este caso "spambots" para hacer el trabajo de estar obteniendo constantemente nuevas direcciones de e-mail a través de las cuales enviar sus mensajes y principalmente robots que entren a las computadoras a robar su directorios de e-mail para acrecentar así su dominio publicitario, en el mejor de los casos, aunque los robots cibernéticos tienen varios usos más que no mencionaré.

Una vez que los malos se armaron de robots, los proveedores de servicios en Internet con información sensible, tuvieron que desarrollar mecanismos de defensa para pararlos y uno de ellos son los *CAPTCHA*.


Alan Mathison Turing [1912-1954].

DE HUMANOS Y MÁQUINAS INTELIGENTES

El matemático, criptógrafo y filósofo inglés, Alan Mathison Turing [1912-1954], condenado, a causa de su homosexualidad, a un tratamiento farmacéutico, equivalente a la castración, que lo llevaría al suicidio por envenenamiento con cianuro, dejando junto a sí una manzana mordisqueada, es considerado el padre de la inteligencia artificial.[5,6]

Turing, a partir de la hipótesis positivista de que, si una máquina se comporta en todos los aspectos como inteligente, entonces debe ser inteligente, establece, en sus trabajos pioneros, las bases conceptuales que han permitido la interacción hombre-máquina actual.

En un artículo publicado en la revista *Mind* de octubre de 1950, titulado "*Computing Machinery and Intelligence*", el cual inicia con "*I PROPOSE to consider the question, 'Can machines think?*", Turing presenta lo que actualmente se conoce como el Test de Turing, una secuencia de preguntas que permite identificar la existencia de inteligencia en una máquina, y por extensión, en caso de tener sólo flujo de datos, el identificar si se interactúa con una máquina.[7]

Estos conceptos se han adaptado a la problemática de identificación que presenta la batalla entre humanos y robots informáticos que actualmente se desarrolla en diferentes medios de comunicación, principalmente Internet, por supuesto que en sentido inverso, o sea: utilizando el Test de Turing para reconocer si se interactúa con humanos.

Una primera propuesta de usar la prueba de Turing para identificar humanos fue el trabajo "Verification of a human in the loop, or Identification via the Turing Test" elaborado por Moni Naor del Weizmann Institute of Science y fechado el 13 de septiembre de 1996.[8] Posteriormente, en 1997, Andrei Broader y sus colegas en AltaVista desarrollaron pruebas buscando imágenes resistentes a la interpretación por parte de programas identificadores de caracteres (OCR: *Optical Character Recognition*).[9]

CAPTCHAS

Acrónimo de "Completely Automated Public Turing test to tell Computers and Humans Apart", *CAPTCHA* (Prueba de Turing pública y automática para diferenciar a máquinas y humanos) es una prueba tipo desafío-respuesta utilizada en computación para determinar cuándo el usuario es o no humano. Dado que la prueba es controlada por una máquina, en lugar de un humano, como en la Prueba de Turing, también se denomina Prueba de Turing Inversa.[9]

El término Captcha, marca registrada de la Universidad Carnegie Mellon, fue acuñado en 2000 por Luis Von Ahn, Manuel Blum, Nicholas J. Hopper (de dicha universidad), y John Langford (entonces en IBM), durante el desarrollo de un proyecto para Yahoo que buscaba establecer un método que permitiera evitar que programas delincuentes invadieran sus sistemas de e-mails y chats.[10]

Este grupo trabajó sobre el desarrollo de acertijos cognitivos que pudieran ser generados y evaluados por computadoras pero que no pudieran ser resueltos por ellas. Uno que mostró ser adecuado fue el que se denomina Gimpy,[11] que presenta una serie de palabras empalmadas y distorsionadas tomadas al azar de un diccionario de las cuales se pide al usuario que las escriba en un lugar específico de la pantalla.

De ahí se llegó a la versión simplificada consistente en una sola secuencia de letras distorsionadas que inicialmente fue utilizada como captcha.[12]

Ejemplo de Gimpy.

Captcha primitivo, el cual actualmente es fácilmente resuelto por robots.

Actualmente se asigna que un captcha debe tener las siguientes características:

- Una computadora puede crear y revisar el acertijo.
- Las computadoras no pueden resolver el acertijo.
- Las personas deben poder percibir, entender y resolver el acertijo.
- El tipo de prueba depende del tipo de usuario humano que se desea la pase.
- El costo de engañar al sistema de captcha debe ser significativamente mayor que los beneficios obtenibles al violarlo.
- La sofisticación de un captcha debe ser proporcional al valor de la información a proteger, de tal manera que el porcentaje de error en la identificación de robots sea inversamente proporcional al valor de la información a proteger.
- Los captchas deben evolucionar a como la inteligencia artificial evolucione.

De los puntos anteriores hay aspectos que merecen discutirse y que se tratarán en los siguientes apartados.

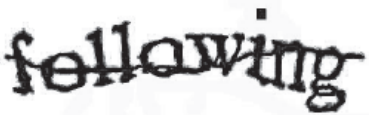
SOBRE LOS PROBLEMAS DE ACCESIBILIDAD Y SELECTIVIDAD

Desde que se crearon los captchas iniciales, resultó claro que el reconocimiento de imágenes de letras distorsionadas por parte de los humanos era muy superior al de las computadoras de ese momento y que, por supuesto, se dejaba fuera del universo de los que podían contestarlos correctamente no sólo a las computadoras sino también a los ciegos, que representa un pequeño porcentaje de la población.[13]

Lo anterior resultó inaceptable en algunas sociedades, pues se consideraron discriminatorios, e iban en contra de la tendencia de volver la computación cada vez más accesible para los diferentes tipos de minusválidos, por lo que se empezaron a desarrollar captchas auditivos que presenta una grabación de letras y números a los que se les sobreponen ruidos y/o se distorsionan.[14]

Muchos sistemas de verificación con captchas actualmente incluyen una versión sonora.

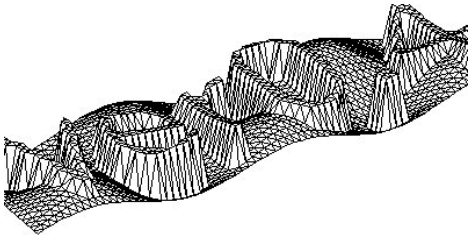
La investigación en este sentido continúa y a como los captchas se vuelven, en apariencia o de momento, más seguros también resulta más difíciles de resolver, al grado que algunos de ellos hay que intentarlos más de una vez para ser reconocido como humano. Esto ha llevado a que se busquen otros tipos de pruebas diferentes a las de letras.



Letras con línea encima. Uno de los intentos iniciales para dificultar a las máquinas la resolución de la prueba.

El empalmar las letras dificulta notablemente la identificación automática de caracteres.

El incluir fondos y sobreposiciones sin patrón ayuda a dificultar la resolución del captcha, incluso a los humanos.



Ejemplo de captcha usando el concepto de tridimensionalidad.

Para realizar tu comentario, por favor, responde a la pregunta siguiente...

la palabra en la coordenada B-2 es...

	A	B	C	D
1	uva	naranja	aceituna	caballo
2	coche	casa	cebolla	perro
3	manzana	guisante	lechuga	tigre
4	pera	fresa	tomate	tomate

Enviar comentario

Captcha de respuesta codificada.

¿Por qué no fotos?

Por otro lado la selectividad de la prueba, cuando se desea que sólo ciertos humanos pasen la prueba ha llevado a cosas parecidas a captcha que no dejan de ser buenas bromas, de hecho hay quienes se divierten creando captcha extremos, como el solicitar la resolución de un acertijo sobre un tema muy especifico y con gran dificultad, como ejemplo el siguiente.[15]

Qualifying question

Just to prove you are a human, please answer the following math challenge.

Q: Calculate:

$$\frac{\partial}{\partial x}\left[6 \cdot \sin\left(4 \cdot x - \frac{\pi}{2}\right) + 3 \cdot \cos(2 \cdot x)\right]\Big|_{x=2\pi}$$

A:

mandatory

Note: If you do not know the answer to this question, reload the page and you'll get another question.

SOBRE LA VULNERABILIDAD Y LOS SUEÑOS DE SEGURIDAD

En toda guerra los bandos constantemente están trabajando en mejorar sus técnicas de defensa y ataque, así que, en lo relativo a seguridad computacional, a tiempos se gana y a tiempos se pierde.

Realmente siempre ha sido claro para los creadores de captchas que estos tendrán una vida útil determinada y sólo un alto porcentaje (se espera que muy alto) de efectividad.

De hecho los captchas han sido un bonito reto para los malos, y para algunos buenos que trabajan en seguridad, y por otro lado una oportunidad para el desarrollo de la inteligencia artificial. El terreno del reconocimiento de caracteres (OCR) y de imágenes (visión artificial) ha mejorado significativamente en los últimos años.

Por supuesto que se han desarrollado estrategias para vencer a los captchas, como son:[16]

- Procesando la imagen buscando: filtrar colores, eliminar líneas, filtrar ruido, aumentar contraste, esto antes de aplicar OCR.
- Mejorando en reconocimiento de caracteres en lo relativo a distorsiones de letras.
- Y como era de esperarse, haciendo que los robots esclavicen a personas, programando que el robot tome la imagen del captcha, la ponga en una página de Internet que ofrezca algo gratis, por ejemplo pornografía, espere que un humano solucione el captcha, y dicha información utilizarla para pasar el captcha, todo esto por supuesto antes de que expire la sesión, o sea el tiempo que se da a la persona para solucionar el captcha.

Que no funcionó… bueno, un robot no está obligado a triunfar a la primera y como no se cansa

ni se aburre puede utilizar la probabilidad e intentar muchas veces hasta que lo logre.

Actualmente es posible conseguir captchas gratuitos para proteger información sensible en nuestras páginas de Internet u otros medios, pero es recomendable ser precavidos pues lo gratis pudiera tener sus costos, dado que estamos insertando un programa en nuestro sistema informático.[17]

Por otro lado hay personas que se anuncian para romper captchas, pero claro que como es ilegal, lo que ofrecen es probar la seguridad de un sistema de captcha. De hecho los programas para romper captchas ofrecen diferentes porcentajes de probabilidad de éxito según el tipo de imagen a descifrar, y por supuesto el costo es proporcional a la dificultad.[18, 19]

COMENTARIOS FINALES SOBRE EL NO FIN DE LAS GUERRAS

El párrafo final del citado artículo pionero de Turing: "*We can only see a short distance ahead, but we can see plenty there that needs to be done.*" resulta aún válido.[7]

Los usuarios de sistemas de información claman por mejores sistemas de seguridad que eviten la necesidad de tener que acreditarse como humano cada vez que se intente hacer un trámite de datos sensibles. Pero por otro lado los negocios a escala mundial se incrementan y por lo tanto los pagos virtuales y el flujo de información sensitiva también.

Deberemos seguir luchando, adaptándonos, bromeando y sobreviviendo en el ciberespacio… es *Our fate*.

El bien y el mal somos nosotros en persona y sociedad. La historia de la humanidad es una secuencia de guerras, no debe sorprendernos que nuestros espacios virtuales lo sean también, y dado que somos muy egocéntricos, y por añadidura antropocéntricos, se vislumbra que las máquinas nos aprenderán.

REFERENCIAS

1. Antonio Caravantes Spam, décimo aniversario. http://www.caravantes.com/04/spam10.htm
2. First Commercial Spam. http://www.mailmsg.com/SPAM_history_001.htm
3. SPAM, http://www.cpiicyl.org/ciudadanos/boletines/seguridad/Spam.pdf
4. Qué es el Spam. http://www.geocities.com/siliconvalley/way/4302/spam.html
5. Andrew Hodges, Alan Turing: a short biography. http://www.turing.org.uk/bio/part1.html
6. Alan Turing (1912-1954), http://etsiit.ugr.es/alumnos/mlii/Alan%20Turing.htm
7. A. M. Turing. Computing machinery and intelligence. Mind:Vol. LIX. No.236, October, 1950, p.433-460. http://www.abelard.org/turpap/turpap.htm
8. Moni Naor, No publicado, Verification of a human in the loop or Identification via the Turing Test. Weizmann Institute of Science. 1996. http://www.wisdom.weizmann.ac.il/~naor/topic.html
9. Captcha. Wikipedia. http://en.wikipedia.org/wiki/Captcha
10. Brad Stone. A Dog or a Cat? New Tests to Fool Automated Spammers. The New York Times. Junio 11 de 2007
11. The captcha proyect. Gimpy. Carnegie Mellon University, School of Computer Science. http://www.captcha.net/captchas/gimpy/
12. Sara Robinson. Human or Computer? Take this test. The New York Times. Diciembre 10 de 2002.
13. W3C. Inaccessibility of captcha. Alternatives to Visual Turing Tests on the Web. W3C Working Group Note 23 November 2005. http://www.w3.org/TR/turingtest/
14. Nelson Rodríguez-Peña. Captchas y Accesibilidad. http://www.webstudio.cl/blog/captchas-y-accesibilidad/
15. Can you handle this, you little spambot? http://bolsanegra.com/2007/09/17/can-you-handle-this-you-little-spambot/
16. How To Crack Captchas June 5th, 2007. http://www.apathysketchpad.com/blog/2007/06/05/how-to-crack-captchas/.
17. Felix Holderied & Sebastian Wilhelmi. Free CAPTCHA-Service. http://captchas.net/
18. Has CAPTCHA Been "Broken"? http://www.codinghorror.com/blog/archives/001001.html
19. OCR Research team. http://www.ocr-research.org.ua/