УДК 004.77

Yelizaveta Khovrenko, student
Svitlana Shcherbyna, Candidate of Pedagogical Sciences
Central Ukrainian National Technical University, Kropyvnytskyi

THE IMPORTANCE OF CAPTCHA FOR CYBERSECURITY IN INTERNET

The article is devoted to the role and importance of the Completely Automated Public Turing test to tell Computers and Humans Apart. A variety of the tests and its history are presented in the article. The importance of using CAPTCHA in the Internet to protect users and information. Highlighting the problem of vulnerability this technology before modern artificial intelligence.

captcha, cybersecurity, artificial intelligence, security in the Internet

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a computer test used to determine whether a user of a system is a human or a computer. It is a security measure known as challenge-response authentication. A verification code is employed to protect against spam and password theft. To verify, users must complete a simple test to confirm that the actions are being performed by a human rather than a computer program attempting to gain access to a password-protected account [1].

CAPTCHA was originally developed by researchers from Carnegie Mellon University and was primarily associated with a technique where a person identifies distorted characters in a raster image and then enters those characters into a web form. This approach is widely recognized by Internet users. Eran Reshef, Jili Raanan, and Eilon Solan, who worked at Sanctum on the Application Security Firewall, first patented CAPTCHA in 1997. Their patent application states that “the invention is based on the application of human strengths in the application of sensory and cognitive skills to solve simple problems that are too difficult for computer software. Such skills include, but are not limited to, the processing of sensory information such as the identification of objects and letters in a noisy graphic environment.” One of the earliest commercial applications of CAPTCHA was in the Hausebeck-Levchin test. In 2000, idrive.com began protecting its registration page with CAPTCHA and prepared to file a patent. In 2001, PayPal used such tests as part of a fraud prevention strategy in which they asked people to "re-enter garbled text that programs have difficulty recognizing". PayPal co-founder and CTO Max Levchin helped commercialize this use. [2].

In the most common variant of the test, users are required to enter characters typically displayed in a distorted manner on a provided image, sometimes with added noise or transparency. This is done to prevent computer programs from recognizing the content of the image. Images can contain objects, such as cars or streets, and the user needs to select all images that contain a certain object. The user is required to select certain images from a set that meet certain criteria. Text type CAPTCHA is widely used, it requires active intervention and understanding of the text from the user. The user is prompted to enter the text displayed on the image or audio file. It can print garbled text or garble characters, making it difficult for bots to automatically recognize. Less frequently, tests based on speech recognition (primarily as an alternative for people with visual impairments). Audio CAPTCHA uses audio files to verify that the user is a real person. The user is prompted to listen to an audio and enter its duration or content. The audio may contain voice instructions or codes that the user must enter to verify their human nature.

The most widely used service is Google's reCAPTCHA, originally developed at Carnegie Mellon University and used for helping recognize words during text digitization, as well as

protecting websites from bot access to restricted resources. On September 16, 2009, Google acquired this technology. In 2013, reCAPTCHA began implementing behavioral analysis of browser interactions to predict whether the user was a human or a bot. The following year, Google introduced a new "invisible" reCAPTCHA, where the verification happens in the background, and challenges are not displayed if the user is considered low risk [1].

CAPTCHAs based on text reading or other visual perception tasks prevent access for blind or visually impaired individuals to protected resources. Since websites may use CAPTCHAs as part of the registration process or even for every login, this issue can block access. Thus, the use of CAPTCHAs excludes a small percentage of users from utilizing significant subsets of popular web services such as PayPal, Gmail, Orkut, Yahoo!, many forum systems, and weblogs. An alternative method involves displaying users a simple math equation and requiring them to enter the solution as a verification. While these are much easier to overcome with software, they provide much higher accessibility for blind users than image-based test. However, they may be challenging for users with cognitive impairments, such as dyscalculia [1,2].

CAPTCHA plays a significant role in cybersecurity and protecting information on the Internet. It is used to prevent various bots and automated programs from impersonating humans and carrying out actions such as automated account creation, subscription to offers, harvesting email addresses, creating email accounts, breaching privacy, password cracking attempts, spam emails, or messages. The importance of using CAPTCHA from the point of view of cyber security is that it allows you to protect sites and services from various types of cyber-attacks that can harm their functioning, reputation, and users. For example, attacks on passwords, where automated programs are used to select or crack passwords to user accounts on websites or services. A CAPTCHA helps prevent such attacks because it limits the number of password attempts and requires an additional human presence check. Also, this test helps prevent DDoS attacks, as it limits the number of requests or downloads from a single IP address and requires confirmation of human presence.

Unfortunately, this technology also has vulnerabilities and bugs that allow CAPTCHAs to be bypassed or recognized without human intervention. This can lead to a breach in the security of sites and services that use this technology to protect against automated attacks. For example, if the CAPTCHA image has low resolution, low distortion, a uniform background, or readable text, it can make it easier to recognize using artificial intelligence, machine vision, or image processing. Also, if a site or service uses a limited set of CAPTCHA images that are repeated with a certain frequency, this may allow creating a database of already recognized pictures and using them to bypass the check. This can allow clustering, classification, or nearest neighbor search techniques to be used to identify similar images. The use of weak generation algorithms can allow hacking or imitating such algorithms and predicting or simulating the results of their work. Also, it may allow using methods of reverse engineering, decoding or feature extraction to identify rules or regularities in the generation.

Artificial intelligence can pose a threat to CAPTCHA if they are able to bypass verification and mimic human behavior. This can lead to a breach in the security of sites and services that use this verification to protect against automated attacks. For example, artificial intelligence may use machine learning, machine vision, or natural language processing techniques to recognize the text, images, or sounds used in tests and enter the correct answer. AI can also use planning, search, or optimization techniques to perform more complex tasks that require logic, contextual understanding, or creativity. AI-powered bots are rapidly advancing and can now outsmart the reCAPTCHA methodology used to verify the authenticity of users on various websites. They do so by mimicking the human brain and visual recognition processes. Experts from Microsoft, the Swiss Federal Institute of Technology Zurich, the University of California, Irvine, and the Lawrence Livermore National Laboratory, involving 1,400 participants who tested websites using CAPTCHA puzzles, experimented. 120 out of the top 200 websites in the world used these puzzles. The accuracy of bots ranged from 85% to 100%, with most exceeding 96%. Some tests required humans 9 to 15 seconds to solve with an accuracy of approximately 50% to 84%, while bots could solve them in less than a second and do so almost perfectly [3].

Although CAPTCHA is used primarily for security reasons, it can also serve as a benchmark for artificial intelligence technologies. And this fuels the race to improve artificial intelligence and methods for distinguishing humans and machines. When a new program manages to solve a problem in an automated way, it is an improvement in artificial intelligence. But the method of distinguishing a person from a robot is losing its reliability and is also starting to improve.

In summary, CAPTCHA helps prevent automated attacks such as forum spam, server resource abuse, fake account creation, and safeguarding against attacks on registration, login, comments, surveys, and other malicious activities. This technology also helps protect user's confidential information from interception or theft. Furthermore, this technology can improve data quality and services by converting handwritten text into digital format or recognizing objects in images. However, it's important to note that some types of CAPTCHAs can be cumbersome for users, so it's important to use them cautiously and in a balanced manner to avoid creating excessive obstacles for legitimate users.

References

1. https://en.wikipedia.org/wiki/CAPTCHA
2. https://www.w3.org/TR/turingtest
3. https://qz.com/ai-bots-recaptcha-turing-test-websites-authenticity-1850734350