Visual Image Authentication

Michael Stephen Fiske
Aemea Institute
San Francisco, CA, 94129, USA

ABSTRACT

Malware plays a critical role in breaching computer systems. Cybersecurity research has primarily focused on malware detection. Detection methods are currently up against fundamental limits in theoretical computer science. The purpose of visual image authentication is to create Artificial Intelligence (AI) problems that are easy for humans to understand, yet also hinder malware from understanding or capturing a human user's credentials or keys. Human visual intelligence still surpasses AI visual recognition systems, particularly when a problem is undefined, or visual occlusions or distracting objects occur in the image. Our technology can exploit AI weaknesses because overlapping, outlining and distracting can help hinder AI recognition of visual images.

Keywords: artificial intelligence · authentication · cybersecurity · human-machine interaction · human visual intelligence · key exchange · malware · visual images

Human Interaction & Emerging Technologies (IHIET-AI 2022)
Artificial Intelligence & Future Applications
https://openaccess.cms-conferences.org/#/publications/book/978-1-7923-8989-4

INTRODUCTION

Malware plays a critical role in breaching computer systems. Cybersecurity research has primarily focused on malware detection (Idika and Mathur, 2007). It seems unlikely that malware detection methods can solely provide an adequate solution to the malware problem because there does not exist an algorithm that can detect all malware (Filiol, 2005). Moreover, some recent malware implementations use NP problems (Cook, 2000) to encrypt and hide the malware (Filiol, 2012). Overall, detection methods are currently up against fundamental limits in theoretical computer science.

With these limitations in mind, the purpose of visual image authentication (Fiske, 2020) is to create Artificial Intelligence (AI) problems that are easy for humans to understand, yet also hinder malware from understanding or capturing a human user's credentials or cryptography keys. Human visual intelligence still surpasses AI visual recognition systems (Kunda, 2020), particularly when a problem is undefined, or visual occlusions and distracting objects occur in the image. Our technology can exploit AI weaknesses because overlapping, outlining and distracting can help hinder AI recognition of visual images.

VISUAL IMAGE ONE-TIME PASSCODES

An embodiment of our technology can secure financial accounts and transactions when there is malware operating on Alice's computer or smartphone. When Alice is requested to verify her identity so that she can login to her financial account, she is sent (via a distinct channel) an image that acts as a visual one-time passcode (OTP).

During authentication, Alice is prompted with four display screens where she must select the correct image, corresponding to each one (see Figure 1).

Figure 1: A visual image OTP: elephant, piano, horse, trumpet.

On display screen 1 (see Figure 2), Alice must select the elephant to correctly authenticate the first visual image of her visual image one-time passcode.

Figure 2: Display Screen 1

On display screen 2, Alice should select the piano, located in row 4 and column 1.

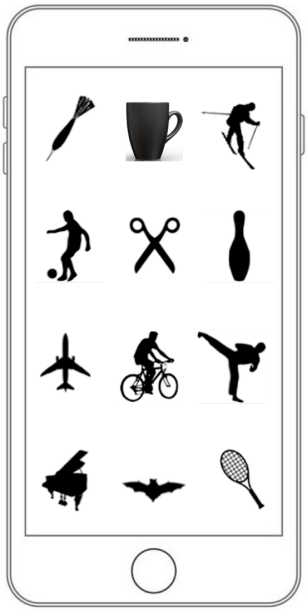

Figure 3: Display Screen 2

According to Figure 1, on display screen 3, Alice should select a horse. On display screen 4, Alice should select a trumpet. During subsequent authentications, the images and the object categories change, creating a visual image OTP.

Figure 4: A visual image OTP: truck, camera, helicopter, cowboy

Figure 4 shows a distinct visual image OTP, used during a separate authentication and with a different image format. This image format illustrates the use of overlapping, outlining and distracting via rotation to hinder malware understanding of the visual image OTP.

In some applications, the visual image OTP can be sent to Alice via email. Email eliminates the cost of SMS; offers a global infrastructure; and provides privacy for users who don't wish to disclose their mobile phone number.

HINDERING SIM CLONING ATTACKS

When Alice authenticates on a smartphone, our technology can verify that she chose the correct image with her finger(s). This human-computer interaction helps address SIM cloning attacks, when SMS is used to send the visual image passcode to Alice, because cloning SIM information on a distinct smartphone is not sufficient for a hacker to breach our authentication system. This human-screen interaction with human fingers also helps assure that a bot using machine learning is not attempting to gain unauthorized access or execute a fraudulent transaction.

DETECTING AN EVE-IN-THE-MIDDLE

Our visual image technology has another useful application in computer security. During a public key exchange, Alice and Bob may not want to trust a central authority who holds a root certificate. In a trustless, decentralized cryptographic system, Alice and Bob can detect an Eve-in-the-middle-attack (Geary, 2009) on their public key exchange, using an enhancement of short authentication strings (Vaudenay, 2005). Visual images can act as an out-of-channel transmission medium, so Alice and Bob can verify that an Eve-in-the-middle attack has not compromised their public key exchange (Fiske, 2021).

Figure 5 shows an Eve-in-the-middle-attack on Alice and Bob's public key exchange. Eve intercepts Alice's public key g^a, which was intended for Bob.[1] Eve sends Alice her public key g^d, pretending she is Bob. Eve and Alice compute shared secret g^{ad} without Alice knowing that her shared secret is actually with Eve.

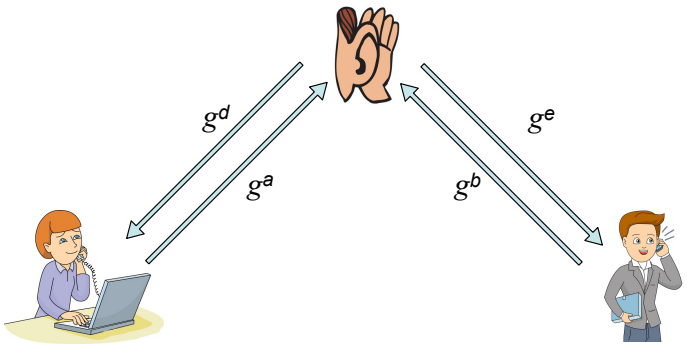

Figure 5: Eve-in-the-middle-attack on Alice and Bob's Public Key Exchange

Figure 5 also shows Eve intercepting Bob's public key g^b. Eve also sends Bob her public key g^e, pretending she is Alice. Eve and Bob compute shared secret g^{be} without Bob knowing that his shared secret is with Eve.

Figure 6: Visual Image Encoding

[1] g is an element of the abelian group, used by the key exchange. a and b are Alice and Bob's private keys, respectively. In (Montgomery, 1987), an abelian group is described, based on elliptic curves of the form $y^2 = x^3 + A\,x + B$, where A and B lie in a field of characteristic other than 2 or 3, and $4A^3 + 27B^2 \neq 0$. In these groups, g is a point on the elliptic curve.

We describe a detection algorithm, based on a visual image encoding (see Figure 6) scheme that can be used by Alice and Bob to detect an Eve-in-the-middle attack. Alice and Bob can use this visual image encoding to check that they have the same shared secret without disclosing their shared secret.

When the key exchange is completed, Alice and Bob each apply a one-way hash function ϕ such as SHA-512 (NIST, 2012) to their shared secret, followed by a visual image encoding computation, denoted as f. If Alice and Bob compute the same shared secret g^{ab}, then $\phi(g^{ab}) = \phi(g^{ba})$ and f can transform $\phi(g^{ab})$ to 5 visual images as follows. Let $b_1b_2b_3b_4b_5$ denote the first five bytes of $\phi(g^{ab})$. The first visual image is b_1 modulo 16, using the encoding. If b_1 modulo $16 = 7$, then the first visual image is a horse. The ith visual image is b_i modulo 16, according to the visual encoding in Figure 6.

After the key exchange, Alice and Bob compare their sequences of five visual images without relying on any centralized authority. If Alice and Bob's sequence of visual images are the same and all shared secrets are equally probable and ϕ probabilistically behaves like a random oracle (Bellare and Rogaway 1993, Koblitz and Menezes 2015), then with probability $\frac{16^5 - 1}{16^5} \approx 0.999999$ Alice and Bob can be assured that their public key exchange was not hijacked by Eve.

When an Eve-in-the-middle-attack occurs on their public exchange, then the probability that Alice's shared secret with Eve g^{ad} is not equal to Bob's shared secret g^{be} is extremely close to 1. For example, when Alice or Bob's shared secret is representable with 256 bits and each shared secret is equally likely, then the probability that g^{ad} equals g^{be} is 2^{-256}.

When Eve hijacks Alice and Bob's public key exchange, Figure 7 shows an example sequence of 5 images that Alice derives from her shared secret g^{ad} with Eve: skier, airplane, bike, scissors, baseball player.

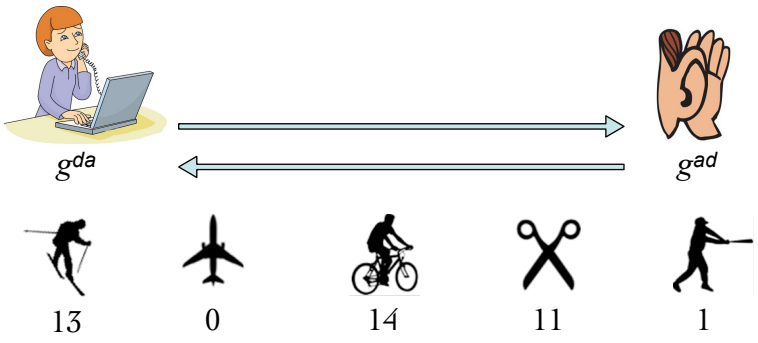

Figure 7 Visual Images Derived from Alice and Eve's shared secret g^{ad}

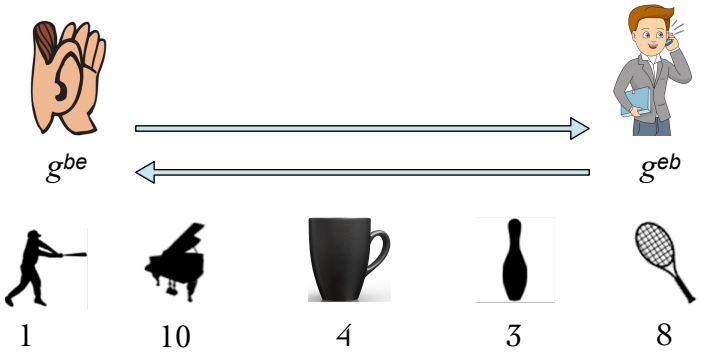

Figure 8: Visual Images Derived from Bob and Eve's shared secret g^{be}

Figure 8 shows a sequence of 5 images that Bob derives from his shared secret g^{be} with Eve: baseball player, piano, cup, bowling pin, tennis racket.

When Alice and Bob compare their derived images, and communicate on a separate channel, they can immediately detect that an Eve-in-the-middle-attack was launched on their key exchange.

CRYPTOGRAPHIC LIBRARY

Our visual authentication system is protected with a cryptographic library, coded in ANSI C (Kernighan and Ritchie, 1988). Our library uses the following cryptographic primitives: elliptic curve 25519 for our public key exchange (Bernstein, 2006) and digital signatures (Bernstein et al. 2012), AES-256 (NIST, 2001), and one-way hash function SHA-512 (NIST, 2012).

ACKNOWLEDGMENTS

We thank Darren Sullivan for his outstanding software development efforts. We thank Mark Tuttle for sponsoring Darren's software development efforts.

REFERENCES

Bellare, Mihir and Rogaway, Phillip. (November 1993) Random oracles are practical: a paradigm for designing efficient protocols. Proc. First Annual Conf. Computer and Communications Security, ACM. pp. 62–73

Bernstein, Daniel. (April 2006) Curve25519: new Diffie–Hellman speed records. Public Key Cryptography. LNCS 3958. New York, Springer. pp. 207–228

Bernstein, Daniel. (August 14, 2012) Duif, Neils; Lange, Tanja; Schwabe, Peter; Yang, Bo-Yin. High-speed high-security signatures. Journal of Cryptographic Engineering. Volume 2, pp. 77-89

Cook, Stephen. (2000) The P versus NP Problem. Clay Math Institute.

Filiol, Eric. (2005) Computer Viruses: From Theory to Applications. Springer.

Filiol, Eric. (March 7, 2012) Malicious Cryptology and Mathematics. Cryptography and Security in Computing. Chapter 2. Intech.

Fiske, Michael S. (March 17, 2020) Visual Image Authentication. US Patent 10,592,651.

Fiske, Michael S. (September 21, 2021) Visual Image Authentication. US Patent 11,128,453.

Geary, Aaron C. (2009) Analysis of a Man-In-The-Middle-Attack on the Diffie-Hellman Key Exchange Protocol. Naval Postgraduate School.

Idika, Nwokedi and Mathur, Aditya P. (February 2007) A Survey of Malware Detection Techniques. Technical Report. Department of Computer Science. Purdue University.

Kernighan, Brian and Ritchie, Dennis. (1988) The C Programming Language. Prentice Hall.

Koblitz, Neal and Menezes, Alfred. (April 2015) The Random Oracle Model: A Twenty-Year Retrospective. Cryptology ePrint Archive. https://eprint.iacr.org/2015/140.pdf

Kunda, Maithilee. (November 24, 2020) AI, visual imager, and a case study on the challenges posed by human intelligence tests. PNAS. Volume 117. No. 47. pp. 29390–29397

Montgomery, Peter. (January 1987) Speeding the Pollard and Elliptic Curve Methods of Factorization. Mathematics of Computation. Volume 48. No. 177. pp. 243–264

Vaudenay, Serge. (August 2005) Secure Communications over Insecure Channels Based on Short Authenticated Strings. Advances in Cryptology. LNCS 3621. Springer. pp. 309–326

NIST. (November 2001) Advanced Encryption Standard (AES). FIPS 197.

NIST. (March 2012) FIPS-180-4. Secure Hash Standard. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf