Segurança de *HIP*s além da obscuridade: *CAPTCHA*s, de volta ao básico

Introdução

Sistemas de Prova de Interação Humana (do inglês: *Human Interaction Proof - HIP*), conhecidos como *CAPTCHA*s (também do inglês: *Completely Automated Public Turing test to tell Computers and Humans Apart*[1]), se baseiam na apresentação de desafios que devem ser facilmente realizados por humanos e difíceis de serem realizados por máquinas.

São usados para restringir acesso a conteúdo sensível ou controlar atividades em um ambiente computacional. Observa-se, porém, que muitos dos *HIP*s propostos aparentam pautar-se em técnicas de segurança por obscuridade, na qual pressupõe-se o desconhecimento do atacante sobre o funcionamento do mesmo. Essas técnicas tendem a ser questionadas no contexto de segurança devido à impossibilidade de auditá-las e por apresentar um possível risco de ataque por àqueles que detém o acesso aos dados obscurecidos.

Objetivo

Atestar a segurança dos *HIP*s disponíveis na literatura dada uma avaliação sistemática pautada no princípio da Máxima de Shannon: "o inimigo conhece o sistema"[6], assim como o uso de técnicas recentes de reconhecimento de padrões que possam oferecer risco aos demais *HIP*s mesmo que partam desse pressuposto.

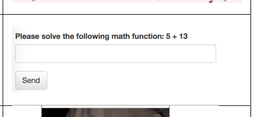
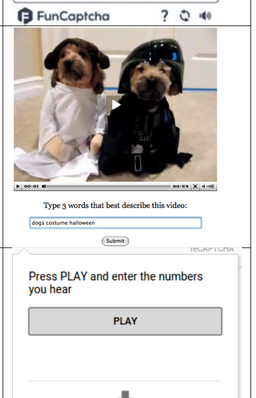
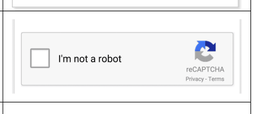
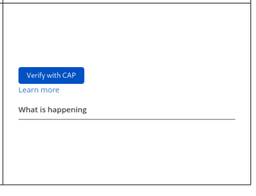
Metodologia

Avaliou-se o impacto da publicização dos dados utilizados na geração de *HIP*s sobre a vulnerabilidade dos mesmos às técnicas de busca reversa e reconhecimento de padrões, assim como técnicas de ataques específicos a determinados sistemas utilizando-se apenas processamento de imagem e de sinais.

Por fim, como um estudo de caso, usando modelos de geração de *CAPTCHA* textual (por ser o *HIP* mais difundido e verificadamente não-baseado em obscuridade), gerou-se conjuntos de dados com 1 milhão de elementos para o treinamento de redes neurais convolucionais genéricas realizando variações no modelo geracional, tamanho de alfabeto e dimensões das palavras, avaliando-se a generalidade dos modelos treinados às variações.

Desenvolvimento

Desenvolveu-se uma taxonomia possível pelo tipos de desafio [3]:

Categoria	Subcategoria	Desafio	Interface
Visual	Textual	Identificação de caracteres	
Visual	Textual	Resolução de problema lógico	
Visual	Imagético	Identificação de objetos em imagens	
Visual	Imagético	Reconhecimento de transformações em imagens	
Visual	Vídeo	Interpretação de conteúdo em vídeo	
Sonoro	Auditivo	Reconhecimento de caracteres ditados	
Comportamental	Telemétrico	Análise de dados de uso	
Token	Chave Criptográfica [4]	Autenticação de chave eletrônica por estrutura certificadora	

Resultados

Observou-se que a maioria dos sistemas da literatura partem de um pressuposto de inacessibilidade de um atacante à relação de etiquetas e desafios, pois, mesmo se aplicadas transformações ao desafio proposto, o mesmo poderá manter-se vulnerável a técnicas de detecção de características invariantes às mesmas.[3]

Para o estudo de caso usou-se 3 variações de modelos geracionais possibilitados pela biblioteca *captcha.py* utilizada: *wheezy* (um modelo depreciado com cores e taxas de deformação fixas), *default* (que busca aleatorizar os elementos fixos do modelo anterior) e *font* (que utiliza o mesmo conjunto de transformações que o *default*, mas, para cada caractere sorteia a fonte a ser utilizada na renderização do mesmo. Além disso, variou-se o alfabeto em letras maiúsculas sem diacríticos (A), letras minúsculas sem diacríticos (a) e números (n), treinando-se modelos de solução com a combinação de cada variação (*d_An*, por exemplo, refere-se ao conjunto de dados gerado com o algoritmo padrão e com o alfabeto composto de letras maiúsculas e números), também combinou-se essas variações com palavras de 2, 4, 8 e 16 caracteres de extensão, obtendo taxas de acerto acima de 60% em todas as variações até 8 caracteres.

Ao se generalizar a aplicação dos modelos treinados para conjuntos de dados distintos, observou-se uma maior capacidade de generalização dos modelos com maior variabilidade e uma razoável incompatibilidade entre o modelo depreciado com o padrão, como era esperado.

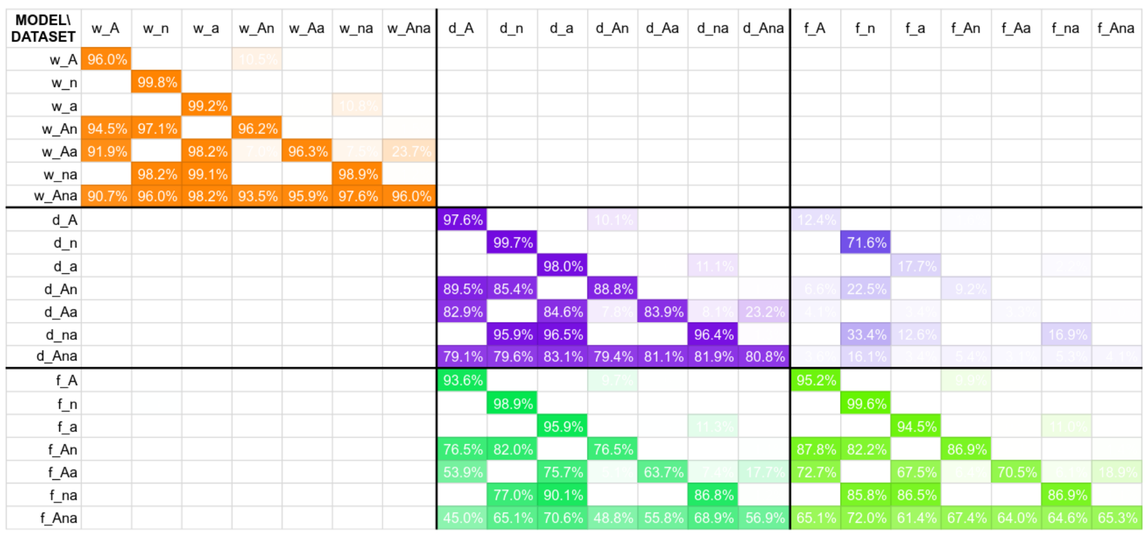


Figura 1: *Tabela de generalização de modelos de 8 caracteres com o modelo treinado no eixo y aplicado ao conjunto de dados gerado com o modelo no eixo x, a intensidade da cor refere-se ao percentual de acerto*

Conclusões

Concluiu-se que apesar de *CAPTCHA*s terem sido concebidos como um teste público, as alternativas propostas na literatura, assim como disponíveis no mercado, dificilmente partem desse pressuposto, falhando em apresentar sistemas comprobatoriamente seguros.

Os sistemas generativos, como o explorado no estudo de caso, são vulneráveis a ataques de sistemas de reconhecimento de padrões e, apesar de estratégias comumente usadas para aumentar a segurança de senhas em sistemas criptográficos (aumento de alfabeto e de extensão) contribuírem para a segurança do sistema, as mesmas diminuem a acessibilidade do mesmo.

Alternativas recentes baseadas em *token*s criptográficos apresentam o melhor equilíbrio entre segurança e acessibilidade impondo, porém, barreiras econômicas ao seu uso.

Referências

- Luis von Ahn et al. "CAPTCHA: Using hard AI problems for security". Em: *International conference on the theory and applications of cryptographic techniques*. Springer. 2003, pp. 294–311.
- Darko Brodić e Alessia Amelio. "The CAPTCHA: Perspectives and Challenges: Perspectives and Challenges in Artificial Intelligence". Em: (2019).
- Meriem Guerar et al. "Gotta CAPTCHA'Em all: a survey of 20 Years of the human-or-computer Dilemma". Em: *ACM Computing Surveys (CSUR)* 54.9 (2021), pp. 1–33.
- Thibault Meunier. *Humanity wastes about 500 years per day on CAPTCHAs. It's time to end this madness.* 2021. URL: https://blog.cloudflare.com/introducing-cryptographic-attestation-of-personhood/ (acesso em 10/10/2022).
- Fabien Petitcolas. "La cryptographie militaire". Em: *J. des Sci. Militaires* 9 (1883), pp. 161–191.
- Claude E Shannon. "Communication theory of secrecy systems". Em: *The Bell system technical journal* 28.4 (1949), pp. 656–715.
- Jing Wang et al. "CAPTCHA recognition based on deep convolutional neural network". Em: *Math. Biosci. Eng* 16.5 (2019), pp. 5851–5861.

Carybé Gonçalves Silva

Supervisor: Ronaldo Fumio Hashimoto

Departamento de Ciência da Computação - Universidade de São Paulo

IME-USP

USP