UNIVERSIDADE FEEVALE

JARDEL KRONBAUER KAPPAUN

# PROPOSTA DE CAPTCHA ADAPTATIVO BASEADO EM IMAGENS GERADAS ATRAVÉS DA TRANSFERÊNCIA DE ESTILOS

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2021

JARDEL KRONBAUER KAPPAUN

# PROPOSTA DE CAPTCHA ADAPTATIVO BASEADO EM IMAGENS GERADAS ATRAVÉS DA TRANSFERÊNCIA DE ESTILOS

Anteprojeto de Trabalho de Conclusão de Curso, apresentado como requisito parcial à obtenção do grau de Bacharel em Ciência da Computação pela Universidade Feevale

Orientador: Paulo Ricardo Muniz Barros

Novo Hamburgo
2021

# RESUMO

O *Completely Automated Public Turing Test to Tell Computers and Humans Apart* (CAPTCHA) trata-se de um mecanismo de segurança que realiza um teste com o objetivo de diferenciar entre usuários humanos e programas autônomos como *bots*. Como um mecanismo de segurança é amplamente utilizado em diversos sites e plataformas web, como de grandes companhias, a exemplo Microsoft, Google e eBay. Um CAPTCHA pode auxiliar a prevenir que programas autônomos (*bots*) sejam capazes adquirir acesso a contas online, dessa forma, evitando operações mal-intencionadas, como o envio e-mails de *spam*, influenciar resultados em votações de pesquisas, quebra de privacidade, coleta de dados, entre outros. Dentre os tipos CAPTCHAs, os baseados em texto são a técnica de segurança mais utilizada em sites na internet, além destes os CAPTCHAs baseados em imagens também são bastante utilizados. Recentes pesquisas na área de reconhecimento de imagens, e aprendizado de máquina tem explorado falhas de design e segurança destes mecanismos, a fim de avaliar o grau de segurança de cada CAPTCHA. O reconhecimento de imagens comparado ao reconhecimento de texto, é visto como um problema muito mais difícil para os programas de computador resolverem, mas a utilização de *Deep Learning* e outros métodos de aprendizado de máquina, tem possibilitado que CAPTCHAS sejam resolvidos com certa facilidade. Dessa forma esse trabalho tem o objetivo de apresentar um protótipo de Software de CAPTCHA, atendendo a requisitos como, alta e fácil usabilidade para humanos, e alta complexidade contra computadores. O protótipo de CAPTCHA adaptativo a ser proposto, deve ser capaz de identificar através do conjunto de respostas submetidas, um possível atacante, e a partir disso, ajustar sua complexidade utilizando técnicas de transferência de estilo entre imagens, e contra-atacar o possível *bot*, sem afetar a sua complexidade para humanos.

Palavras-chave: CAPTCHA. Bots. Segurança. Reconhecimento de Imagens. Usabilidade.

# SUMÁRIO

# MOTIVAÇÃO

O *Completely Automated Public Turing Test to Tell Computers and Humans Apart* (CAPTCHA) trata-se de um mecanismo de segurança que realiza um teste com o objetivo de diferenciar entre usuários humanos e programas autônomos (VON AHN; BLUM; LANGFORD, 2004). Esse mecanismo de segurança serve para com proteção a ataques autômatos, *spam*, ou mesmo de atividades fraudulentas. Esse mecanismo de segurança é amplamente utilizado em diversos sites e plataformas web, como de grandes companhias, a exemplo Microsoft, Google e eBay (LI *et al.*, 2021).

Existem vários tipos de CAPTCHAs, que podem ser categorizados como baseados em texto, áudio, vídeo e imagens. CAPTCHAs baseados em textos possuem como característica a utilização de um dicionário de palavras ou a geração de uma cadeia de caracteres, e então aplicado uma técnica de degradação, deformação, ou mesmo a aleatorização das fontes utilizadas (ALQAHTANI; ALSULAIMAN, 2020). Os CAPTCHAs de áudio servem como uma alternativa para usuários que possuem alguma imparidade visual, e consiste de sons sobrepostos, e um sintetizador de voz que lê a informação (ALQAHTANI; ALSULAIMAN, 2020). Os CAPTCHAs de vídeo possuem *tags* definidas pelo usuário no momento do upload do vídeo, onde o usuário é então questionado sobre que *tags* melhor representam vídeo, a resposta será então validada mediante as *tags* definidas para o vídeo no momento do *upload* (ALQAHTANI; ALSULAIMAN, 2020). CAPTCHAs baseados em imagens, funcionam através do reconhecimento de elementos em uma ou mais imagens, fazendo com que o usuário selecione o que foi solicitado, ou indique onde um objeto está presente em uma imagem (ALQAHTANI; ALSULAIMAN, 2020), a exemplo o mais conhecido é o reCAPTCHA fornecido pelo Google.

Os CAPTCHAs baseados em texto são a técnica de segurança mais utilizada em sites na internet (LI *et al.*, 2021), em razão disso pesquisadores têm conduzido pesquisas e desenvolvido abordagens de métodos de ataque a fim de examinar a complexidade desses CAPTCHAs, e definir o grau de segurança que eles proporcionam. Dessa forma novas pesquisas têm atingido melhores resultados com utilização de mecanismos de reconhecimento de texto. Além da utilização de CAPTCHAS textuais, do mesmo modo, os CAPTCHAs baseados em imagens, onde, da mesma forma que ocorre com os baseados em textos, estudos veem sendo realizados com a utilização de diversas técnicas de reconhecimento de imagens

existentes, bem como o desenvolvimento de novos métodos, que veem possibilitando avaliar o segurança fornecida por esses CAPTCHAs.

É importante ressaltar que, por design um CAPTCHA não é capaz de diferenciar entre um usuário legítimo e ataques que utilizam pessoas, dessa forma sites ficam vulneráveis a esse tipo de ataque (ALSUHIBANY; ALRESHOODI, 2021). O CAPTCHA é uma tecnologia projetada para prevenir que programas autônomos (*bots*) sejam capazes adquirir acesso a contas online, pois a partir daí é possível fazer com que eles passem a realizar todo tipo de operação mal-intencionada, como o envio e-mails de *spam*, e manipular, ou influenciar resultados em votações de pesquisas (ABURADA *et al.*, 2019).

Em redes sociais, os *bots* sociais tem se tornado uma rápida e seria questão de segurança, pois são utilizados como ferramentas de software capazes de realizar atividades maliciosas em redes sociais, de maneira autônoma, como, realizar o compartilhamento de postagens e mensagens, realizar o envio de solicitações de amizades a diversos usuários, e do mesmo modo, também são capazes de coletar informações como endereços de e-mail, telefones, ou dados pessoais que possam ter valor monetário (BOSHMAF *et al.*, 2011). Há evidências de que os *bots* sociais desempenham um papel crucial na invasão da privacidade e segurança dessas redes, logo essas ferramentas de software maliciosas representam um grande desafio aos serviços de redes sociais (MOHAMED; ALI; HANI, 2019).

O tráfego *Web* em sites de *e-commerce* tem sido dominados por agentes autônomos que põe em risco a sua segurança, a sua privacidade e a sua performance (ROVETTA; SUCHACKA; MASULLI, 2020). Em razão disso esse tráfego tem levantado séria preocupação entre os administradores desses sites, pois o tráfego de *bots* ocasiona o consumo considerável dos recursos em servidores *web*, resultando em altas cargas de trabalho e longos tempos de resposta, sem que haja lucro por parte da empresa de *e-commerce.* E no pior dos casos, a navegação no conteúdo das páginas pode ser utilizada mais tarde em outros tipos de atividades fraudulentas (XU *et al.*, 2018).

A fim de trazer segurança aos sites e evitar que recursos em servidores *web* sejam desperdiçados, a utilização de CAPTCHAs surge como uma solução, porém CAPTCHAs possuem problemas de design e segurança, como demonstrado no trabalho de  Mori e Malik, CAPTCHAs baseados em modelos de linguagem são suscetíveis a ataques, pois uma base de dados de palavras pode ser construída para resolver um CAPTCHA. Do mesmo modo

modelos de CAPTCHA que utilizam distorção também puderam ser resolvidos (MOY *et al.*, 2004).

O Time de pesquisa da Microsoft demonstrou que a aplicação de aprendizado de maquina é capaz de resolver certa variedade de CAPTCHAs, em seu estudo foi demonstrado que um computador é capaz de reconhecer um caractere de maneira mais rápida que um ser humano em CAPTCHAS compostos de caracteres aleatórios, mesmo fazendo uso de técnicas para aplicação de distorção ou desordem (CHELLAPILLA; SIMARD, 2005) (CHELLAPILLA *et al.*, 2005).

CAPTCHAs baseados em texto são fortemente suscetíveis a técnicas de segmentação, utilizadas para separar os caracteres e reconhecê-los um a um. Desta forma, foram desenvolvidos métodos a fim de fortalecer a resistência a segmentação como, ruído de fundo, utilização de linhas e colapso de palavras (Chow YW et al.). Em resposta às técnicas utilizadas para prevenir segmentação, métodos novos de segmentação foram sendo criados, como, algoritmos para eliminação de ruído, segmentação baseada em histograma, segmentação baseada em padrões e formatos, dentre outras (Chow YW et al.).

Comparado com o reconhecimento de texto, o reconhecimento de imagem é visto como um problema muito mais difícil para os programas de computador resolver. Isso ocorre porque ainda há um grande domínio de problemas de percepção e interpretação de imagens por computadores (ZHU *et al.*, 2010). Desta forma o reconhecimento de imagens aplicadas a esses CAPTCHAs fornece uma oportunidade de estudo para o seu reconhecimento e resolução. Há propostas para resolução do reCAPTCHA fornecido pelo Google, através da utilização de *Deep Learning*, atingindo alta taxa de sucesso (WANG; MOH; MOH, 2020) e um estudo de (ALQAHTANI; ALSULAIMAN, 2020) onde o método de ataque proposto no estudo, utilizando aprendizado de máquina, apresentou uma taxa de precisão de 85.32% para 56.29% dos desafios propostos.

A fim de melhorar o desempenho desses CAPTCHAs devemos primeiramente atender alguns requisitos como 1) possibilitar que o CAPTCHA seja de alta e fácil usabilidade para humanos e 2) seja ao mesmo tempo resiliente a ataques autônomos, que fazem uso de aprendizado de máquina e outros métodos de reconhecimento de imagens. Tendo como base o estudo apresentado por SHI *et al.*(2020), que desenvolveu um protótipo de CAPTCHA capaz de identificar através do conjunto de respostas submetidas a ele, um possível atacante, e a

partir desses dados aumentar sua complexidade, permitindo realizar um contra-ataque, sem que seja afetada a complexidade para humanos.

O ajuste de complexidade das imagens pode ser realizado através de técnicas de transferência de estilos entre imagens, como aplicado nos trabalhos de (KWON; YOON; PARK, 2020) e (CHENG *et al.*, 2019). O uso deste método apresenta-se efetivo contra classificação de imagens (CHENG *et al.*, 2019), sendo capaz de reduzir o reconhecimento de imagens por computadores, sem afetar a taxa de acerto por humanos (KWON; YOON; PARK, 2020), permitindo que o grau de complexidade da imagem seja manipulado de acordo com o possível atacante, dessa forma fornecendo mais segurança e adaptabilidade contra aos métodos de reconhecimentos de CAPTCHAs. Sendo assim, este trabalho se propõe a desenvolver e validar um protótipo de Software de CAPTCHA baseado em imagens, geradas através de transferência de estilos, que atenda a requisitos como, alta e fácil usabilidade para humanos, e alta complexidade contra computadores, sendo capaz de adaptar o grau de complexidade mediante ataques de origem autônoma.

# OBJETIVOS

Objetivo geral

Desenvolver e validar um protótipo de CAPTCHA baseado em imagens, capaz de adaptar o grau de complexidade mediante ataques de origem autônoma, dificultando o reconhecimento de imagens por computadores, sem impactar na experiência de um usuário.

Objetivos específicos

- Estudar a literatura sobre o estado da arte de métodos de reconhecimento de imagens aplicadas a CAPTCHAS;
- Buscar na literatura o estado da arte, sobre a transferência de estilos entre imagens, bem como;
- Estudar a literatura sobre o estado da arte no desenvolvimento de ferramentas para criação de *bots*;
- Identificar possíveis *datasets* para utilização na manipulação de imagens;
- Estudar métodos de avaliação utilizados em estudos similares, para avaliação do protótipo;
- Desenvolver um protótipo de CAPTCHA adaptativo;
- Avaliar o protótipo desenvolvido, no contexto da segurança;
- Avaliar o protótipo desenvolvido, no contexto da usabilidade;

# METODOLOGIA

A metodologia a ser utilizada para o desenvolvimento desta pesquisa será a *Design Science Research* (DSR). Esta metodologia estabelece um processo sistemático que tem como objetivo projetar e desenvolver artefatos que tenham condições de resolver problemas, com foco em garantir que, o que foi desenvolvido, atinge os objetivos inicialmente propostos, demostrando assim sua alta relevância para aplicação em campo prático (DRESCH; LACERDA; MIGUEL, 2015).

Essa metodologia consiste de seis elementos essenciais (DRESCH; LACERDA; MIGUEL, 2015) (PEFFERS *et al.*, 2006), descritos abaixo, tendo juntamente com a aplicação no presente estudo de desenvolvimento de um protótipo de CAPTCHA adaptativo com uso de transferência de estilos em imagens:

1. **Identificação do problema:** buscar na literatura os principais métodos existentes para identificação de *bots*, buscar os métodos mais comuns para reconhecimento de imagens, aplicado a CAPTCHAS, e identificar métodos utilizados para realização de transferência de estilos entre imagens. Como motor de busca será usado a *Web of Science*, que consolida artigos e materiais provenientes de várias origens de dados;
2. **Definição dos objetivos para a solução:** definir a linguagem de programação na qual o software para o CAPTCHA será desenvolvido e quais requisitos o software terá. Avaliar e definir uma ou mais bases de dados com imagens diversificadas que possibilitem diversas iterações de estilos;
3. **Projeto e desenvolvimento:** desenvolver o protótipo de CAPTCHA em si, capaz de 1) identificar possíveis atacantes e 2) adaptar o nível de complexidade na geração das imagens;
4. **Demonstração da Solução:** aplicar o protótipo de CAPTCHA dentro de um contexto de site de e-commerce, simulando ataques, a fim de demonstrar a identificação de *bots*, e utilizando de uma ou mais bases de dados de imagens para a geração de CAPTCHAs;
5. **Avaliação:** avaliar através dos métodos mais comuns, utilizados em estudos por outros pesquisadores, o CAPTCHA desenvolvido na presente pesquisa. Será avaliada a precisão na identificação e distinção entre pessoas e computadores, e o desempenho

dos diferentes graus de complexidade que podem ser atingidos através da técnica de manipulação de imagem;

6. **Comunicação dos resultados:** Apresentação dos dados ao final do TCC;

# CRONOGRAMA

## Trabalho de Conclusão I

| Etapa | Meses | | | |
|---|---|---|---|---|
| | Mar | Abr | Mai | Jun |
| Buscar na literatura os principais métodos existentes para identificação de *bots*. | | X | X | |
| Buscar os métodos mais comuns de reconhecimento de imagens em CAPTCHAS. | | X | X | |
| Buscar técnicas utilizadas para transferência de estilo entre imagens. | | X | X | |
| Escrita do TCC I | X | X | X | X |
| Definição linguagem de programação. | | | X | |
| Buscar base de imagens. | | | X | |
| Desenvolvimento do modelo para o CAPTCHA. | | | X | X |

## Trabalho de Conclusão II

| Etapa | Meses | | | |
|---|---|---|---|---|
| | Ago | Set | Out | Nov |
| Desenvolvimento do CAPTCHA. | X | X | | |
| Escrita do TCC II | X | X | X | X |
| Simular o CAPTCHA dentro de um contexto de site de *e-commerce*. | X | X | | |
| Aplicar a geração de CAPTCHAs utilizando bases de imagens. | X | X | | |
| Avaliar a precisão na identificação e distinção entre | X | X | X | |

| | | | | |
|---|---|---|---|---|
| pessoas e computadores. | | | | |
| Avaliar grau de complexidade das imagens geradas. | X | X | X | |
| Escrever e apresentar resultados no documento do TCC. | | X | X | |
| Apresentação dos resultados | | | | X |

# BIBLIOGRAFIA

ABURADA, Kentaro *et al.* Implementation of CAPTCHA suitable for mobile devices. **IEICE COMMUNICATIONS EXPRESS**, KIKAI-SHINKO-KAIKAN BLDG, 3-5-8, SHIBA-KOEN, MINATO-KU, TOKYO, 105-0011, JAPAN, v. 8, n. 12, p. 601–605, 2019. Disponível em: https://doi.org/10.1587/comex.2019GCL0060

ALQAHTANI, Fatmah H.; ALSULAIMAN, Fawaz A. Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. **Computers and Security**, [*s. l.*], v. 88, p. 101635, 2020. Disponível em: https://doi.org/10.1016/j.cose.2019.101635

ALSUHIBANY, Suliman A; ALRESHOODI, Latifah A. Detecting human attacks on text-based {CAPTCHAs} using the keystroke dynamic approach. **{IET} Information Security**, [*s. l.*], v. 15, n. 2, p. 191–204, 2021. Disponível em: https://doi.org/10.1049/ise2.12018

BOSHMAF, Yazan *et al.* The Socialbot Network: When bots socialize for fame and money. **ACM International Conference Proceeding Series**, [*s. l.*], p. 93–102, 2011. Disponível em: https://doi.org/10.1145/2076732.2076746

CHELLAPILLA, Kumar *et al.* Computers beat humans at single character recognition in reading based human interaction proofs (HIPs). **2nd Conference on Email and Anti-Spam**, [*s. l.*], v. 98053, 2005.

CHELLAPILLA, Kumar; SIMARD, Patrice Y. Using machine learning to break visual human interaction proofs (HIPs). **Advances in Neural Information Processing Systems**, [*s. l.*], 2005.

CHENG, Zhouhang *et al.* Image-based CAPTCHAs based on neural style transfer. **IET Information Security**, [*s. l.*], v. 13, n. 6, p. 519–529, 2019. Disponível em: https://doi.org/10.1049/iet-ifs.2018.5036

DRESCH, Aline; LACERDA, Daniel Pacheco; MIGUEL, Paulo Augusto Cauchick. Uma análise distintiva entre o estudo de caso, a pesquisa-ação e a design science research. **Revista Brasileira de Gestao de Negocios**, [*s. l.*], v. 17, n. 56, p. 1116–1133, 2015. Disponível em:

https://doi.org/10.7819/rbgn.v17i56.2069

KWON, Hyun; YOON, Hyunsoo; PARK, Ki Woong. CAPTCHA image generation: Two-step style-transfer learning in deep neural networks. **Sensors (Switzerland)**, [*s. l.*], v. 20, n. 5, p. 1–14, 2020. Disponível em: https://doi.org/10.3390/s20051495

LI, Chunhui *et al.* End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network. **Neurocomputing**, [*s. l.*], v. 433, p. 223–236, 2021. Disponível em: https://doi.org/https://doi.org/10.1016/j.neucom.2020.11.057

MOHAMED, Torky; ALI, Meligy; HANI, Ibrahim. A challenge-response mechanism for securing online social networks against social bots. **International Journal of Ad Hoc and Ubiquitous Computing**, [*s. l.*], v. 32, n. 1, p. 1–13, 2019. Disponível em: https://doi.org/10.1504/IJAHUC.2019.101819

MOY, Gabriel *et al.* Distortion estimation techniques in solving visual CAPTCHAs. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, [*s. l.*], v. 2, 2004. Disponível em: https://doi.org/10.1109/cvpr.2004.1315140

PEFFERS, Ken *et al.* The design science research process: A model for producing and presenting information systems research. **Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST**, [*s. l.*], 2006.

ROVETTA, Stefano; SUCHACKA, Grażyna; MASULLI, Francesco. Bot recognition in a Web store: An approach based on unsupervised learning. **Journal of Network and Computer Applications**, [*s. l.*], v. 157, n. October 2019, 2020. Disponível em: https://doi.org/10.1016/j.jnca.2020.102577

SHI, Chenghui *et al.* Text Captcha Is Dead? A Large Scale Deployment and Empirical Study. **Proceedings of the ACM Conference on Computer and Communications Security**, [*s. l.*], p. 1391–1406, 2020. Disponível em: https://doi.org/10.1145/3372297.3417258

VON AHN, Luis; BLUM, Maunei; LANGFORD, John. Telling humans and computers apart automatically. **Communications of the ACM**, [*s. l.*], v. 47, n. 2, p. 56–60, 2004. Disponível em: https://doi.org/10.1145/966389.966390

WANG, Dylan; MOH, Melody; MOH, Teng Sheng. Using Deep Learning to Solve Google reCAPTCHA v2's Image Challenges. **Proceedings of the 2020 14th International Conference on Ubiquitous Information Management and Communication, IMCOM 2020**, [*s. l.*], p. 4–8, 2020. Disponível em: https://doi.org/10.1109/IMCOM48794.2020.9001774

XU, Haitao *et al.* **Detecting and characterizing web bot traffic in a large e-commerce marketplace**. [*S. l.*]: Springer International Publishing, 2018. ISSN 16113349.v. 11099 LNCS Disponível em: https://doi.org/10.1007/978-3-319-98989-1_8

ZHU, Bin B. *et al.* Attacks and design of image recognition CAPTCHAs. **Proceedings of the ACM Conference on Computer and Communications Security**, [*s. l.*], p. 187–200, 2010. Disponível em: https://doi.org/10.1145/1866307.1866329