3W-02

検索エンジン攻撃に耐性のある画像認証方式の検討

林　夏生† 佐藤　敬‡

†北九州市立大学大学院国際環境工学研究科　‡北九州市立大学国際環境工学部

1　はじめに

近年ブログや SNS などの Web サービスが普及するとともに，BOT（コンピュータ）による広告の不正な書き込みやメールアドレスの自動取得によって SPAM が自動送信されるなどの不正な利用も増えている．この不正利用を防止するために人と BOT を識別する CAPTCHA(Completely Automated Public Turing test to Tell Computers and Humans Apart) と呼ばれるチューリングテストの利用が増えている．一方で CAPTCHA を攻撃する手法が高度化しており，それに対抗するべくテストの難易度も上昇し，人間でも正解できない問題が現れる事態となった．そこで Google 社の開発した reCAPTCHA2.0[1] ではマウスの挙動と共に画像に何が写っているか読み取らせる画像認証が使われている．

本稿では検索エンジン攻撃によって reCAPTCHA2.0 で用いられる画像認証が破られる例を示し，検索エンジン攻撃に耐えうる画像の加工例を挙げる．

2　既存技術とそれに対する脅威

2.1　reCAPTCHA2.0

reCAPTCHA2.0 は Google 社が提供している CAPTCHA である．この認証では「私はロボットではありません」というチェックボックスが現れ，それをクリックするまでのマウスの挙動により BOT か人間か判断する．もしそれで BOT の疑いがあるのであれば**図 1** のように問題文と解答候補の画像が 8〜16 枚表示される．この画像の中から正解となるものを複数枚選択し，確認ボタンをクリックすることで認証が行われる仕組みである．なお不正解の場合はさらに別の問題が出題され，正解するまで続けられる．ただし reCAPTCHA2.0 は日々改良されているため，**図 1** で示したもの（2016/09/24 時点での仕様）と今現在の出題形式とは異なることがある．

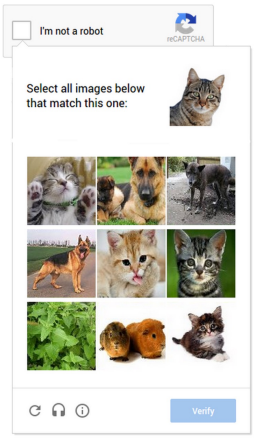

図 1: reCAPTCHA2.0 の例

2.2　検索エンジン攻撃

画像認証を行う際に表示される画像を読み込み，Google 社がサービスしている画像検索 [2] を行う．画像に写っているものをキーワードとして得ることができるので，画像に何が写っているかを把握することができる．得たキーワードと認証時の問題文を比較することで正解画像かどうかを判別することが可能と考えられる．このプロセスを自動化することで BOT を構成し，攻撃に利用できる可能性がある．しかし著者が確認したところでは得たキーワードが正しくなく攻撃は無効な例もあった．なお検索攻撃と呼ばれることもある．[3]

2.3　AI 攻撃

検索エンジン攻撃同様に表示される画像を読み込み，AI によって何が写っている画像か判別する．その結果と問題文を照会し正解画像を自動選択するといった攻撃が懸念される．検索攻撃と比べ，学習のための時間や教材となるデータを必要とする．

An Image Authentication Method Resistant to Search Engine Attack

Natsuo HAYASHI† and Takashi SATOH‡

†Graduate School of Environmental Engineering, The University of Kitakyushu

‡Faculty of Environmental Engineering, The University of Kitakyushu

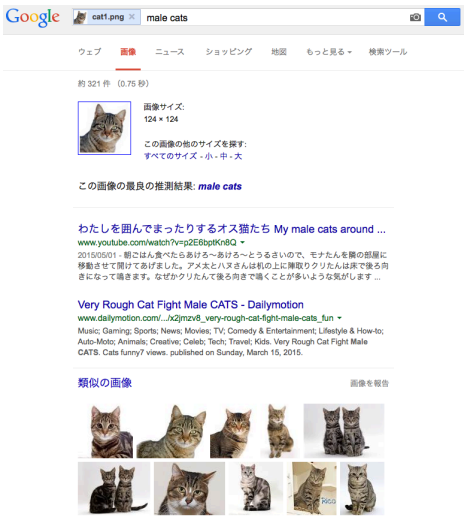

図 2: 攻撃例の検索結果

図 3: 加工の例

3　検索エンジン攻撃の例

現在 reCAPTCHA2.0 に対しての攻撃に関する懸念は著者が知る限り示されていない．そこで実際にどのようにして検索エンジン攻撃がなされるのかを示す．まず reCAPTCHA2.0 によって認証画面が表示される．操作しているのが BOT とすればチェックボックスのクリックによる認証が失敗し画像認証まで進む．ここでは**図 1** が表示されているものとする．解答候補の画像を読み込み，これを Google 画像検索に入力し検索を行う．すると**図 2** のような結果を得る．推測結果に書いてあるキーワードから，この画像に写っているものはオス猫であるとわかる．

4　検索エンジン攻撃に耐えうる加工例

検索エンジン攻撃は画像に何が写っているか BOT が判断できてしまうために脅威となっている．本章は**図 3** のように，認証時に提示する画像に**人間には何が写っているのかわかるが BOT にはわからない**加工を施すことで加工画像が検索エンジン攻撃に耐性があることを示す．

図 4: 加工例の検索結果

元画像の**図 3** 左を Google 画像検索で検索すると推測結果は **cat** となった．**図 3** 右を読み込み，同様にこれを画像検索すると**図 4** の結果が得られるので，

人間にはネコとわかるが BOT にはビーズと誤解を招く加工例となっている．この加工例では画像に格子模様が現れたためにビーズの手芸作品と Google 画像検索が見なしたと考えられる．この結果より本加工によって検索エンジン攻撃への耐性向上が期待できる．

5　考察と今後の課題

4 章で示した加工以外に他にも同様に有効な加工例を発見している．今後は多種多様な画像に加工を施しても検索エンジン攻撃に対して耐性が向上するのか調べ，同時に画像検索のアルゴリズムを探りさらに検索攻撃に対して強靭な加工や手法が無いかを検討する．また AI 攻撃に対しても耐性向上が見込まれるのか評価を行っていきたいと考えている．

参考文献

[1] “reCAPTCHA | Google Developers” https://developers.google.com/recaptcha/

[2] “Google 画像検索” https://www.google.co.jp/imghp

[3] 田村，久保田，油田，朴，岡崎: “文字認識攻撃に耐性を持つランダム妨害図形を用いた画像ベース CAPTCHA 方式の提案”，情報処理学会論文誌 Vol.56 No.3 808-818(Mar. 2015)