インターネット多要素・複数段階認証の安全性評価と限界解明

Evaluating Security and Unraveling the limits of multi-factor/multi-step authentication on the Internet

櫻井　幸一（Kouichi SAKURAI, Ph. D.）

九州大学　大学院システム情報学研究院　教授

(Professor, Kyushu University, Faculty of Information Science and Electrical Engineering)

電子情報通信学会　人工知能学会　応用数理学会　IEEE　ACM　他

受賞：IEEE Transactions on Evolutionary Computation Outstanding 論文賞(2022) 他

著書:数論アルゴリズムと楕円暗号理論入門　N.コブリッツ (著), 櫻井 幸一 (翻訳) 丸善出版 (1997 年)　他

研究専門分野：暗号とサイバーセキュリティ　プライバシー AI

あらまし

本研究はスマホ認証サービスへのなりすまし・不正送金をサイバー認証学の立場から解明し、その対策を研究するものである。安全・安心な支払いインターネットを利用した決済システムの構築を目指して、２段階認証はじめ、{知識・所持・生体}情報を複数組み合わせる多要素認証の現状と課題を論じる。

１．研究の目的

2020 年９月に露呈したスマホ認証サービスへのなりすまし・不正送金をサイバー認証学の立場から解明し、その対策を目指す研究である

スマートフォンをはじめとするモバイルデバイスの普及と高性能化の恩恵を受けて、QR コードを利用した電子決済システムの導入が進んでいる。しかし、近年、ハッカーによる攻撃と不正利用が深刻な社会問題となっている。対策としては、SMS などを利用した２段階認証が基本とされ、事実、単一段認証のみに頼ったサービスシステムは攻撃されている。また、２段階認証でも、パスワードの再設定や、２要素認証の扱いミスによる攻撃も報告されている。

本研究では、安全・安心な支払いインターネットを利用した決済システムの構築を目指して、２段階認証はじめ、{知識・所持・生体}情報を複数組み合わせる多要素認証の設計と安全性解析を研究する。

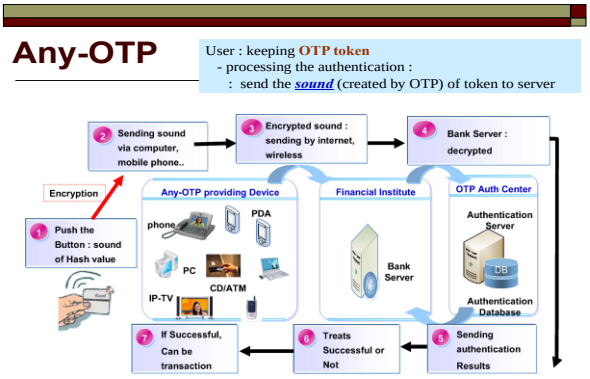
２．研究の背景

社会的背景：2018 年 7 月に経済産業省は、産学官からなるキャスレスサービス委員会を立ち上げ、サービス導入の促進を宣言した。しかし、すでに現場でのサービスに対する攻撃が発生している。IT インフラも、店頭にリーダを設置した、耐タンパ IC カードシステムから、導入コストのかからないスマホ決済システムに移りつつある。

電子認証標準の国際/国内動向：米国政府標準技術研究所 NIST は、2017 年 6 月に、電子認証ガイドライン NIST SP 800-63-3 を発表した。国内でも、NIST を参照し、「行政手続きにおけるオンラインによる本人認証の手法に関するガイドライン」が、各府省情報化統括責任者(CIO)連絡会議で、2019 年 2 月に決定している。政府の CIO ガイドラインでは、行政向けに、マイナンバーに基づく耐タンパな IC カードシステムを対象としている。しかし、より一層の普及を目指して、スマート機器の利活用も議論され始めたところである。

認証の進化と変化：

本研究代表者は、2006 年コンピュータセキュリティシンポジウム（情報処理学会）で、対話型認証の必要性を説く研究発表をしていた「ハッキングに対して安全な電子銀行サービスの一提案」(Kwak, Jangz, 今本, 櫻井)。：ここでは、キーロガーなどを利用したハッキングに対して安全な電子銀行を実現するためのワンタイムパスワード認証方式を紹介し，既存方式との安全性



比較を行った。この 2006 年時点では、まだ現在のようなスマートフォンも普及していなかったが、韓国では、先駆的に２段階認証を銀行サービスに導入していた。Google の 2018 年の発表では、Google 利用者の９割が、２段階認証を利用していない、という事実。反面、多くのセキュリティ警告は、２段認証の導入や活用をすすめている。また、マイクロソフトの新 OS の"CALL"は、アンドロイドスマホからの電話を、Win-PC からかけたり受けたりできる機能を追加したりと、信頼できるデバイスインフラの整備が進んでいる。国内では yahoo ジャパンが、自社でユーザーパスワードを管理することをやめて、ユーザーの信頼できるデバイスにコードを送り、動的認証を行うパスワード無しの認証方式を導入している。

学術的背景：

敵対的機械学習理論の応用と理論展開： CAPTCHA の安全性強化版の１つに DeepCAPTCHA が提案された [Osadchy, M. et al. D. No Bot Expects the DeepCAPTCHA! Introducing Immutable Adversarial Examples, With Applications to CAPTCHA Generation. IEEE Trans. Information Forensics and Security (2017)]。ここで導入されている学習フィルターは、通常のニューラルネットでの非線形関数と違い、非連続である。この性質は、最近、暗号解読の大家である Shamir(2002 年チューリング賞)も注視しており、敵対的機械学習の理論的限界を超越する事例として、期待されている。

人間対 AI：CAPTCHA を見破る自動認識プログラムとそれに対抗できる CAPTCHA の改良、その中で、人間でも判読困難なものや誤読するほど極端なものもあり、ユーザーの利便性も重要な設計条件である。この問題は、人間が計算できる対象で、非対称性・一方向性を持つパスワード、という普遍的なテーマで、CAPTCHA の創始者の一人である Manuel BLUM(1995 年チューリング賞)により研究されている[Blocki, Blum, Datta, Vempala"Towards Human Computable Passwords" proc. ITCS 2017]。

暗号解読 AI: AI 自体を暗号の安全性評価に応用する研究も活性化しつある。共通鍵暗号には、差分解読や線形解読など統計的な性質を利用する安全性評価が現在の主流である。この延長として、ニューラルネットワークを識別器に用いる差分ニューロ解読が,最高峰の国際暗号会議で発表された[Improving Attacks on Round-Reduced Speck32/64 Using Deep Learning (CRYPTO2019)]。これをきっかけとして、この方面の研究が急激に活性化している。

3. 研究課題

CAPTCHA 認証: 画像に描かれた文字や数字など、人間には判読可能であるが、コンピュータでは困難なものを使い、操作している本人が、機械ではなく人間であることを認証する Completely Automated Public Turing Test To Tell Computers and Humans Apart（CAPTCHA 計算機と人間を区別する完全自動の公開チューリングテスト）は、無料電子メールサービスにおける大量のアカウント取得など、機械プログラムによる不正行為防止に、すでに実装され効果をあげている。しかし、CAPTCHA を見破る自動認識プログラムも生まれており、現在では、CAPTCHA の改良と攻撃のイタチごっこが続いている状況にある。最新の人工知能による攻撃の脅威解析とその対策を行い、多要素認証へのシステム応用を設計することが主題である。

人間計算可能な関数を用いたパスワード認証：従来のパスワードでは、覚えやすいパスワードは辞書型攻撃に脆弱である。その一方で、強度ある乱数パスワードは覚えにくいという問題がある。人間計算可能なパスワードは、これらの欠点を解決すると期待して、Blum らが提案した。ここでのパスワードは、人間が記憶できる範囲に限定し、さらに暗算できる程度の簡単な計算を行うものである:

$$f(x0, x1, \cdots, x13) = x\,j + x12 + x13 \bmod 10,$$
$$j = x10 + x11 \bmod 10$$

彼らの研究では、組み合わせ計算論的な解析により、安全性評価が行われていた。しかし、この方式の安全性評価に関して、深層学習を適用するという研究までは行われておらず、本研究の主題として取り組んだ。

４．研究成果

2021/R3

卒研の一環として、本研究で扱う人間計算可能なパスワードは、ユーザーが画像から数字へのマッピンを記憶するフェーズと、画像をもとに与えられた関数 f（人間計算可能（暗算可能）な関数）を用いて応答するフ

ェーズから構成される。本研究では、深層学習を用いて安全性の強度評価を試みた。
多層パーセプトロンを用いて検証を行った結果、パスワードを簡略化した場合の１つで、問題に対する応答の予測精度は 15 %、その他の場合は 10 %という自然な結果を得た。深層学習のモデルは多数存在するので、人間計算可能なパスワードの強度評価を,他の深層学習で行うことを次年度の課題とした。
櫻井自身の研究としては、二要素認証の一設計論 を展開した。Google は 2014 年に、それまでのワンタイムパスワードよりも堅固な安全性を実現できるとする二要素証方式 U2F(Universal 2nd Factor)を発表した。これは USB デバイス上で実装される。本調査では、ワンタイムパスワード方式や U2F 方式の安全性を、暗号プロトコルの設計と解析の立場から論じた。
Google では、Moti Yung 主導で、FIDO アライアンスなどとも連携し、Universal Second Factor 開発計画を遂行した。 Google の USB 型デバイスは製品化されておりブラックボックスで、詳しいアルゴリズムとその設計思想を知ることは容易ではない。本調査では、U2F の技術仕様を、暗号プロトコルの設計と解析の観点から論じ、成果発表を行なった。

2022/R4:研究室学生の卒論として取り組んでいた人間計算可能なパスワード認証に対する深層学習を用いた安全性の強度評価は、一定の成果を収めて、国際会議(査読あり)や国内シンポジムでも発表を行った。

本研究で扱う人間計算可能なパスワードは、ユーザーが画像から数字への対応を記憶するフェーズと、画像をもとに与えられた関数 f(人間計算可能(暗算可能)な関数)を用いて応答するフェーズから構成される。既存研究では、この安全性を組み合わせ論的手法で解析評価している。

また、昨年は、人間計算可能な関数として、原定案の難しい関数と、本研究で明らかにした易しい関数の2つを評価した。現在は、この2つの関数以外、例えば、中間程度の計算困難性をもつ関数の構成に挑戦している。(チューリング機械モデルでの計算複雑性理論においても、こうした関数の構成は、自明ではない研究課題である。本研究では、これを AI で展開しようという試みである。)

櫻井自身の研究としての、二要素認証の一設計論を展開を継続した。

Google は 2014 年に、それまでのワンタイムパスワードよりも堅固な安全性を実現できるとする二要素証方式 U2F(Universal 2nd Factor)を発表した。これは USB デバイス上で実装される。昨年の調査では、ワンタイムパスワード方式や U2F 方式の安全性を、暗号プロトコルの設計と解析の立場から論じ、国内シンポジムでも発表した: Google では、Moti Yung 主導で、FIDO アライアンスなどとも連携し、Universal Second Factor 開発計画を遂行した。 Google の USB 型デバイスは製品化されておりブラックボックスで、詳しいアルゴリズムとその設計思想を知ることは容易ではない。本調査では、U2F の技術仕様を、暗号プロトコルの設計と解析の観点から論じた。

現在も、FIDO をはじめ認証関連のアライアンスや標準化団体の動向を調査すると同時に、現場で実際に使われている二要素認証の攻撃事例の収集と、その原因解明と対策を研究し続けている。

得られた結果

深層学習を用いて、人間計算可能なパスワード方式の安全性強度評価を試みた。多層パーセプトロンを用いて検証を行った結果、パスワードを簡略化した場合の1つで、問題に対する応答の予測精度は 15 %、その他の場合は 10 %という結果を得た。また、人間可能な関数を、原提案よりも簡略すると、予測精度は 80%以上に向上し、人間にとって計算の簡単な関数は、機械でも推測しすい、という自然な結果を得た。また、人間計算可能なパスワードの強度評価を、 LSTM(Long Short Term Memory)学習モデルでも、行っているが、MLP とほぼ同様の性能が得られることは実験的に確認できた。
から” CSS2022 3A4-II-4 (2022.10.26)

2023/R5 の成果

[A] (a1) 初期研究では、一番基本的な深層学習モデルである多層パーセプトロン(MLP: Multilayer Perceptron)を用いた。この場合は、画像から数値への対応と暗算に用いる合成関数の出力が予測可能であるかについて実験を行い、予測精度が 10%前後であると評価した。

また、関数を線形にした場合、予測精度が約 80%に向上したことも観察した。(a2)続く研究では、深層学習モデルとして、別のニューラルネットワーク（LSTM:Long Short Term Memory）を用いて、同様の実験評価を行った。その結果、合成関数の予測精度は 10%前後となり、人間計算可能なパスワードにおいて用いられている合成関数は、安全であることが分かった。一方で関数を線形にした場合の予測精度は 90%を超え、MLP による予測精度より 10 から 20%ほど高くなった。これより、出力に関係する値が演算のたびに変化することが、予測精度の低下につながっていることが分かった。(a3) さらに、最終年度は、埋め込み層を用いた双方向超短期記憶ニューラルネットワークを用いて安全性評価を再度実施した[2023/卒論]。結果、ユーザーの覚える秘密が簡単な場合は 55%程度の精度でなりすましが可能であることがわかった。これは、それまでの組み合わせ論的な解析を超える重要な進展である。(2024 年 3 月の IPSJ/CSEC 研究会で、櫻井が発表し、現在は、国際会議への投稿準備を進めている)。

これまでの一連の結果から、人間計算可能なパスワードの強度は、深層学習解析に対して、モデルに依存しない安全性を確保できていることが検証できたと言える。

[B]櫻井自身の研究としての、二要素認証の一設計論は、FIDO 標準化や製品化動向も調査した。パスワードを用いない認証手段が重要になってきている背景から、前年に続いて、FIDO(Fast Identity Online)を中心に認証関連のアライアンスや標準化団体の動向を調査すると同時に、現場で実際に使われている二要素認証の事例や、安全性に関する学術先端研究も調査した。
FIDO は、Google 主導のアライアンス主体の標準化の成功もあって、Microsoft や Apple も積極的に採用・導入をはじめている。また、ドングルと呼ばれる外付けの専用デバイスも、海外製品をはじめ、ネット通販で個人でも、簡単に購入できる時代になっている。しかし、USB などの外付けデバイスは、その運用。・管理の観点から、障害になることも勘案し、Web 版も普及しつつある。
さらに、最新の次世代 FIDO としては、耐量子暗号を使ったシステムを Google が提案した。 この実装は、対象となる外付けデバイスの記憶容量などの制約もあり、今後の課題の1つと言える。このテーマに関しする最新動向はまだ未発表であるため、今後、調査した成果をサーベイかホワイトペーパーとして発表したいと考えている。

４．現状の課題と将来展望

AI 解読 reCAPTCHAv2：この報告書を作成している 2024 年 10 月の時点での最新のニュースは、スイス工科大の研究チームによる Google の reCAPTCHAv2 システムへの攻撃である。彼らは、それまでの攻撃が 68～71%しか成功しなかったのに対し、機械学習を駆使した新しい手法により 100%で reCATCHAv2 を突破できると報告している。

この論文は、2024 年 7 月に大阪で開催された国際会議 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)で発表され、論文自体は、2024 年 9 月 13 日に Arxiv からオープンアクセス可能である[Andreas Plesner, Tobias Vontobel, and Roger Wattenhofer. “Breaking reCAPTCHAv2"]。また、実験コードも、Github 上に 2024 年 4 月の時点で掲示されていた。

reCAPTCXA は、現在 version-3 まで進化しているが、今後も AI を駆使した突破攻撃に対する強度解析は、学術的にも研究が続くはずである。そして、reCAPTCHA 強化のさらなる改良と、攻撃突破の攻防は続くと予想される。

QR コード認証: QR コードの不正生成と不当表示は、以前から現実の脅威であるが、その対策は、未だ解決できていない挑戦課題である。ユーザーが注意を要するヒューマンセキュリティの課題の側面もあり、IT 技術がどこまで適用可能か、が課題である。また、コード決済はコンビニなど小型店でより普及した現在、新たな脅威として、スマホの高性能カメラを使った、他人のスマホ決済の際の QR コードの盗撮による、なりすましが課題として報告されている。しかし、ここでも AI の進化による利活用と脅威を議論する時期に行っている。

おわりに

AI が人間の能力を、いつどの程度までに超えることができるか、という問題は、人工知能の基本的問いかけであるは、暗号系認証における現状では、深層学習の進化によって、解析技術が格段に優位になってきた。攻撃から守る立場の我々は、AI の威力と限界の解明を今後も続ける必要がある。

関連文献（成果発表）

[S1] 櫻井幸一 “サイバー空間における認証の安全性に関する現状と課題” CSS2021 3C3-4
(2021.10.28)
[S2] 櫻井幸一”二要素.認証の一設計論: Google の FIDO/U2F を事例に”信学技法 ITE ISEC LOIS 2021-11-12
[S3]櫻井幸一 “人間計算可能なパスワード認証の現状と課題:AI 複雑性の観点から” CSS2022 3A4-II-4 (2022.10.26)
[S4] 池田 悠登, ○櫻井 幸一（九州大学）″人間計算可能なパスワード認証に対する埋め込み型の双方向 LSTM を用いた安全性実験評価″ IPSJ/第 104 回 CSEC 2024 年 3 月 19 日(火) 千葉工業大学
[S5] Issei Murata, Pengju He, Yujie Gu & Kouichi Sakurai “Towards Evaluating the Security of Human Computable Passwords Using Neural Networks” The 23rd world conference on Information Security Appliacations, 2022 Aug. 24-26, Korea/ Jeju and online, Lecture Notes in Computer Science book series (LNCS,volume 13720) First Online: 04 February 2023.

この研究は、令和２年度ＳＣＡＴ研究助成の対象として採用され、令和３～５年度に実施されたものです。