2017年度 SWAN Working Group 活動報告

宮本大輔 (daisu-mi@is.naist.jp)　　門林雄基 (youki-k@is.naist.jp)

2017 年 12 月 31 日

目 次

1 はじめに

SWAN (Security for Web 2.0 Application) WG では，悪意あるウェブサイトの動向を観測し検討している．ウェブを介した攻撃にはその攻撃空間が広いという特徴があり，本研究グループはその広い特性に対応した研究を行なっている．これまでの活動としては，エンドユーザの認知能力に合わせたフィッシングサイト解析や脆弱性を持つウェブ 2.0 のアプリケーションを WIDE メンバーに提供する試み，悪意あるウェブサイトによく見られる難読化された JavaScript の構造に着目した解析，PC だけではなく Android などで動くマルウェアの解析技術のハンズオンなどが挙げられる．また，この所は認知心理学に着目したサイバー防御システムの開発と評価を行ってきた.「人間の観測される情報から，人間の精神状態を推測する」という領域の知見を用い，サイバーセキュリティにおける意思決定を認知の側面から解析しようという試みである．

今年度は，イギリスで心理学とサイバーセキュリティの学際領域を研究している，ボーンマス大学の John McAnaley 上級講師らとワークショップを行った．McAlaney 氏の研究グループはオンラインにおいて以下に安全に行動するべきか，あるいは従業員に安全な行動をとらせるべきかといった教材開発を集団心理学の観点から行っている．今年度は 4 月に日本，10 月にワークショップを交流ワークショップを開催した．また，SWAN WG の内容にも関連する試みとして,「サイバーセキュリティ」という概念の見直しを行い，本報告書ではこの紹介を行う．さらに，フィッシング対策についてもモバイルデバイス等でも動作する軽量の対策手法である UnPhishMe の研究開発を行った．これについて概要を説明する．

2 サイバーセキュリティワークと心理学ワークショップ

近年，サイバーリスクに対応する人材・組織・システム・技術を生み出すことが重要視されている．サイバーリスクは技術と人間の複合リスクであり，欠点のない技術であっても人間の判断ミスにより大事故につながる可能性もあり，また人間がミスのない運用をしたとしても技術的な欠陥により大事故につながる可能性もある．特にシステムが社会インフラ・産業基盤にまたがる場合は，経験豊富なオペレータが正確な判断

を下すことが重要である．しかしながら，高度に複雑化した今日の社会インフラにおいて，特に IT の文脈において人間の正確な意思決定支援を支援するための学際的な取り組みは全くといっていいほど不十分である．このため同じ問題意識を持つ異分野融合型の専門家ネットワークを形成する必要がある．

このような背景から，2017 年 4 月 6 日に，サイバーセキュリティと心理学的に関するワークショップ (Workshop on Psychological Factors of Cyber Security) を開催した．本ワークショップは，グレイトブリテン・ササカワ財団の日欧研究助成プログラムからの助成（助成番号 5084）の枠組みで行われた．

3　サイバーセキュリティに対する心理学的な見解

人間の行動が，サイバーセキュリティのシステムに弱点を生み出すことはよく知られている．心理学は，人的要因を狙って攻撃してくる攻撃者の手口を理解することに役立てられている．一方で，サイバー攻撃に対する防御を検討する者にも重要である．例えば，人間は，となりの席の人の行動に影響をうけやすい．会社で隣に座った人が，パスワードをシールに貼って共有していれば，本人も同様の行動を行うようになりやすい．これらの行動を変容させるために心理学的な側面から以下の研究を行っている．

- 意思決定のプロセスの理解
- 行動時における社会的影響の理解
- 行動変容の効果的な通知

先程の例では，どのように，何がきっかけで行動が変容するかを理論的に説明し，なおかつ実際に行われる行動を予測する．さらに，防御方針を検討する者を教育する際には,「人間のミスは避けられないこと」を教え,「人間が起こしうるミスはどのようなものであるか」を理解させ,「どのように人間の行動を変容させるか」について検討させる．

3.1　社会心理学とは

社会心理学とは「他者が実際に存在したり，創造の中で存在したり，或いは存在することが仄めかされていることによって，個人の思考，感情および行動が，どのような影響をうけるのかを理解し，説明する試み」と定義される．たとえ周りに誰もいない状況であったとしても，人間の行動は他者によって影響を受ける．この分野では，グループプロセス，印象管理，動機づけといった観点から，実際の行動の関わりが検討されている．

3.1.1　グループプロセス

グループ内での感情，コミュニケーション，影響のあり方といった組織内の情報と行動の影響である．社会心理学及び経営学の分野からは以下に要約される知見が報告されている．

- 人間は社会的認知を通して陽性強化される．すなわち，好ましい行動を褒めることにより，同じように好ましい行動を取り続けるようになる．これらは，グループメンバーに意図的に褒める人材を入れることによっても，陽性強化が促進される．
- 自分たちのグループは他のグループより優秀であると考える内集団バイアスがある．この傾向は「私の成功はスキルによるもので，彼らの成功は幸運によるものだ」と考えさせる．よって，他者の成功事例を持ち出すことが最適解とは限らない．
- カテゴリー差別化モデル，すなわち他のグループを認識することが，グループの結束を強化するという結果を生み出す．例えば，自分たちがハッカー集団におり，他のハッカー集団がメディアに報告されると，その認識によって自分たちのグループをより活発に行動させる動機になる．
- グループのメンバーシップがあるという状況が，グループの同一性や，自尊心の源泉であるハッキンググループへの参加は，社会的アイデンティティ理論においては，自分と自分の集団を同一化し，自分自身を集団の一部として自覚し行動する意味を持つ．
- チーム内の衝突の観測は重要である．大きなグループであれば，チーム内部での衝突や意見の不一致は経験している．このような衝突は，集団疑集性（集団が構成員をひきつけて，その構成員を集団の

一員になるように動機づける度合い）や，グループの有効性（組織におけるチームの目標達成能力）のインジケータとなる．こうした衝突は，外部のメンバーによって意図的につくることができる．

3.1.2　印象管理

人は，他者の印象を制御しようとし，どのような印象をもたれるかを恣意的に決めようとする．例えば，能力を誇張しようとする傾向にある．また，他の著名人と関連付けし，社会的認知を得ようとする場合もある．ハッカー集団の場合，DEFCON に友人が参加していた，あるいは（著名なハッカー集団である） ANONYMOUS と関連がある，などがこれにあたる．

その一報で,「晒し」（doxing）については脅威に思いがちである．印象管理の観点からサイバー教育を考えれば,「そのような行動をすると，サイバー犯罪の対象と識別される」ことを強く認識させることで教育効果が高まる．

3.1.3　動機づけ

ハッカーの行動の動機については，心理学では，以下のように考察の上で研究が進められている．

威信（prestige）のための行動としてトップコーダーを目指すハッカーもいれば，気晴らし（recreation）として活動するハッカーもいる．復讐（revenge） や収益（profit）はサイバー犯罪の原因となり，イデオロギー（ideology）もハッカー行為の源泉となり得る [1]．

3.2　サイバーセキュリティ教育の事例

McAlaney 氏は，心理学に基づいた教育手法の研究を行っている．例えば，ゲーミフィケーションが教材開発に有効であることは経験的に知られているが，一方でゲームによる教育のアプローチの基礎的な研究部分は進んでいない．

この研究では，ゲームのプレイと道徳（モラル）の関係があること，性別や年齢がモラルの重要な因子であることが知見として得られている．例えば，青少年の感情統制がうまくいかない場合に，モラルから外れた行動をとりがちだという傾向がある．どのようにモラルから外れた行動をとっていることを通知するべきか，ということはサイバー犯罪を防ぐための教育手法にほかならない．ここでは「おかしな行動をとった」を心理学の行動変容であると位置づけ，以下のような対象を相手に研究を行っている．

- 一般大衆に対する教育としては，実際の行動変容の観察及び防ぎ方について，英国心理学会と共同研究している．例えば「他の人はやっていない」という同調圧力は，一般大衆に対する教育に効果がある．
- 特定の組織や個人に対する教育としては，Psyber Social を開発し，教材を提供している．
- サイバーセキュリティ対策を行う人材に対する教育活動は，ボーンマス大学の場合，MSc Cybersecurity and Human Factors の教育過程で実施されていることが示された． [2]

質疑では，行動変容と年齢について議論された．人間の行動は短期的に変わるものだけでなく，長期的に変わっていくものがある．学術は年単位（ 1 年単位）で答えを求められることが多く，長期的な教育効果の観測は課題であり，これについて継続的な取り組みを行っている．

4　「サイバーセキュリティ」に変わる概念

2017 年 8 月 6〜8 日に行われた WIDE Project ボード合宿において,「サイバーセキュリティ」に変わる概念について探索が行われた．劇的な速さで拡大を続けるサイバー空間を「防御する」という概念は果たして今の時代に通用するのだろうか．トラスト，レジリエンス，セーフティ，様々な概念が相補的なものとして重要視されている現状を鑑みれば,「サイバーセキュリティ」のみでは十分でないことは明確である．多くのステークホルダーが長年にわたって「サイバーセキュリティ」の向上を目指す取り組みを行ってきたが，その「セキュアな状態」は絵に描いた餅となっている．

[1] Ryan Seebruck, “A Typology of Hackers”

[2] John Tayloer et al., “Teaching psychological principles to cybersecurity students”

参加者は「サイバーセキュリティ」に代わる新しい概念を探索するべく，情報工学だけではなく，心理学や意思決定，または経済，金融，経営管理といった領域の分野から興味深い論文についてその解説及び共有を行った．そして，参加者同士でグループを形成し，それぞれの概念を融合することで，新しい概念を発見するという試みに取り組んだ．

5 UnPhishMe

As discussed earlier, most anti-phishing solutions are only effective on desktop computers due to computational power issues. By considering the unsuitability of those solutions for low-powered mobile devices, we propose and develop an android application prototype, UnPhishMe, which simulates user authentication procedure through lightweight Java classes and methods. UnPhishMe intercepts a login page opened by a user and simulates the login procedure with fake credentials. Technically, an authentication attempt to a login webpage with incorrect login credentials tests the trustworthiness of that page. However, a user needs to have a prior knowledge and remembers to do so every time she encounters a suspicious page. We believe, in small size devices, this procedure is tedious when done manually. Our work addresses these issues by automating the login procedure through android application on mobile devices.

5.1 Scope and Assumptions

Our work focuses on mobile devices that use an android operating system. However, it is not limited to android-based devices only, it can be re-developed for other device operating systems such as iOS (formerly iPhone OS). Since our solution depends on RFC 2616 industrial standard response for client requests, we are limited to legitimate websites that conform to that standard. Some web servers, such as `http://www.nike.com`, that implement such server standards for an authentication failure, fit well into our solution scope. Together with other criteria, it is

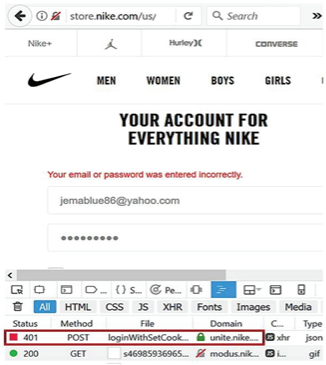

図 1: Implementation of the standard RFC 2616 for a failed authentication

easier to deduce a phishing site when it does not behave like a legitimate one. The standard information can be a certain response code equivalent to a client request such as HTTP 401 Unauthorized or 403 Forbidden. To capture this information we customize and implement several Java classes in UnPhishMe.

5.2 System Model

A big part of this work philosophy is to exploit as many authentication standard features of the HTTP protocol as possible. During authentication, if a user system does not conform to the standard, a 401 unauthorized or a 403 forbidden responses can be given. Most web servers implement web API frameworks such as AJAX and ASP.NET. Such frameworks make it easy to build HTTP services that reach a broad range of clients including browsers and mobile devices. In web API, a web server can throw a proper exception within an authentication method that returns 401 or 403 to a user who provides wrong login credentials. Figure 1 shows a 401 response as implemented by `https://store.nike.com`

Apart from exploiting HTTP protocol features for authentication, UnPhishMe also monitors the URL changes after the authentication attempt. Usually, the URL exhibited before a user logs into a website remains the same even after the authentication attempt fails on a legitimate webpage. If a user is suc-

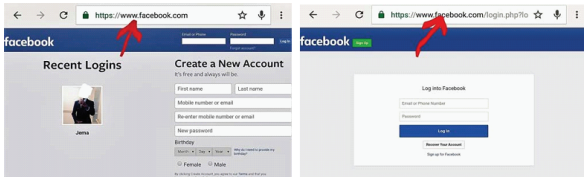

図 2: Alternative login pages on Facebook website

cessfully authenticated, it is certainly expected that the URL will change into a new one. For a malicious site, the URL always changes even when a user tries to log in with incorrect credentials. UnPhishMe computes the hashcode of the URL before and after the authentication attempt to a login page. The computed hashcodes are then compared to determine the URL or page transition. However, in our conducted experiment with 40 most popular eCommerce websites as ranked by Alexa, results show that it is not the same for all cases. In some cases, the URL of a legitimate webpage keeps changing even after an authentication fails. For example, Facebook page has at least two login page alternatives as shown in Figure 2.

To make a decision, UnPhishMe checks the URL hashcode consistency and HTTP response message. If those conditions are met as expected then the application executes. However, the URL consistency condition cannot be satisfied by a single authentication event if the URL keeps changing. Our results indicate that the change frequency that a URL exhibits does not exceed 7. The reason for having these numbers is the fact that some websites alter their URL string based on the number of authentication attempts. Therefore, in order to satisfy the condition for a page legitimacy, we implement an iteration that simulates an authentication attempt to the page for at most 10 times as the highest boundary. When the URL changing frequency is exhausted and it becomes consistent, that means one condition is satisfied, then the application checks the equivalent response code.

The detail description of this work can be shown in the article [1].

6 おわりに

SWAN Working Group は悪性ウェブサイト対策技術についての研究活動を行っており，Drive By Download やフィッシングサイトの研究技術から研究を発展させている．来年度も技術的側面にとどまらず，様々な側面からの分析を行い，対策について研究開発を行う．研究成果は引き続き WIDE 研究会及び学会発表を通じて行い，ソフトウェアなどの成果物は必要に応じた公開を検討している．

参考文献

[1] Jema David Ndibwile, Youki Kadobayashi, and Doudou Fall, *"UnPhishMe: Phishing Attack Detection by Deceptive Login Simulation through an Android Mobile App,"*, *In proceeding of the 12th Asia Joint Conference of Information Security (AsiaJCIS)*, August 2017.