修士論文

アプリケーションレベルの情報提示による ソーシャルエンジニアリング攻撃の 拡散防止手法

早稲田大学大学院基幹理工学研究科
情報理工・情報通信専攻

狩野 佑記

学籍番号　5117FE01-8
提出　2019 年 7 月 22 日
指導教授　中島達夫

A New Approach to Preventing Spread of Social Engineering Attacks by Presenting New Information at The Application Level

Yuki KANO

Thesis submitted in partial fulfillment of

the requirements for the degree of

Master in Computer Science and Communications Engineering

Student ID 5117FE01-8

Submission Date July 22, 2019

Supervisor Professor Tatsuo Nakajima

A Thesis Submitted to the Department of Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University

概要

ソーシャルエンジニアリングとは，人間の心理に付け込み特定の企業や個人の機密情報を盗み取るハッキング手法の総称である．情報技術の発展と共にインターネット上においてもソーシャルエンジニアリングが行われるようになり，インターネットを利用する全ユーザが攻撃の被害に遭う可能性があるという状況になった．そのため，インターネットにおけるソーシャルエンジニアリングの対策を論じる研究が数多く行われている．

Twitter や Facebook などのソーシャルネットワーキングサービス(SNS)の誕生により，コミュニケーションに重みを置いた新たなソーシャルエンジニアリングの脅威が生じている．例えば，被害を直接受けなくとも，ソーシャルエンジニアリング攻撃に繋がる内容の投稿を知人や信頼できる人に対して拡散してしまう危険性がある．

本研究では SNS におけるソーシャルエンジニアリングを対策するために，攻撃に繋がる投稿の拡散を防ぐことを目的とする．その提案手法として，投稿をシェアするタイミングで投稿に対するポジティブな反応とネガティブな反応をユーザに対して強制的に提示する．

提案手法が効果的であることを確かめるために，仮想 SNS を制作し 35 名のユーザを対象に評価実験を行った．その結果，ポジティブな反応はシェアの要因とならないが投稿に対する印象を変え，ネガティブな反応はシェアの要因と密接に結びつき攻撃投稿の拡散を防ぐことができるということが判明した．

提案手法が有効であった理由として，ユーザに対して思考することを誘発できたことが挙げられる．このことから，ソーシャルエンジニアリング攻撃を対策するための手法として新たな括りを定義した：「ユーザに対する思考の誘発」である．この括りを確かなものとさせるために，将来的に，本研究における提案手法とは異なる手法を用いて，ユーザに思考を誘発させることでソーシャルエンジニアリング攻撃が対策できるといった事例を作る必要がある．

Abstract

Social engineering is a method for hacking that steals confidential information from a company or individual by misusing psychology. With the development of information technology, social engineering has been increased on the Internet, so there is a risk that all users who can access the Internet will be attacked. As a result, there has been large number of studies that discuss social engineering measures on the Internet.

The emergence of Social Networking Services (SNS), such as Twitter and Facebook, has created a new social engineering threat that emphasizes communication. For example, there is a risk of spreading the post contents that can leads to social engineering attacks to acquaintances or trustworthy friends without directly receiving damage.

In this research, in order to take measures to social engineering in SNS, we take an approach to prevent the spread of posts that lead to social engineering attacks. Specifically, we proposed a method to force users to be shown positive and negative responses to social engineering posts when sharing them. In order to evaluate that our approach is effective, we created a virtual SNS and conducted experiments on 35 users. From the experimental results, we found that the positive reaction does not become a factor of the share but changes the impression on the posts, and the negative reaction is closely linked to the factor of the share and can prevent the spread of posts that lead to social engineering attacks.

The reasons why our approach was effective is that it could induce users to Unconscious thinking. From this, we defined a new bundling as an approach for take measures to social engineering attacks: Trigger unconscious thinking for the user.

In order to show that this bundling is effective as a social engineering countermeasure, in the future, it is necessary to create more case studies that can take measures to social engineering attack by using method inducing trigger unconscious thinking for the user different from our research.

目次

第 1 章　序論 1
1.1 研究背景 1
1.2 研究目的 1
1.3 論文の構成 2
第 2 章　関連研究 4
2.1 ソーシャルエンジニアリング 4
2.2 ネット社会におけるソーシャルエンジニアリング 5
2.3 SNS とソーシャルエンジニアリング 6
2.4 ソーシャルエンジニアリングの対策 7
第 3 章　提案手法 8
3.1 ソーシャルエンジニアリングの対象 8
3.2 シェアの重みづけ 8
3.3 提案手法の前提条件 10
3.3.1 攻撃投稿の拡散 10
3.3.2 反応の種類判別 10
第 4 章　アプリケーションの設計と実装 12
4.1 アプリケーション概要 12
4.2 アプリケーションの設計 12
4.2.1 サーバの設計 12
4.2.2 クライアントの設計 14
4.3 アプリケーションの実装 15
第 5 章　評価実験 16
5.1 実験の概要と目的 16
5.2 実験方法 16
5.3 事前実験 17
5.3.1 調査内容 17
5.3.2 パラメータの異なるシチュエーション 17
5.3.3 ユーザ情報の登録 19
5.3.4 各シチュエーションに対する質問の項目 20
5.3.5 事前実験終了後のアンケート項目 22
5.3.6 事前実験の結果 22
5.3.7 事前実験の結果の分析，既存研究との比較 27
5.4 評価実験 30
5.4.1 調査内容 30

5.4.2 提案手法を組み込んだシチュエーション.....30
5.4.3 実験の流れ.....32
5.4.4 各シチュエーションに対する質問の項目.....32
5.4.5 実験終了後の最終アンケート項目.....34
5.4.6 実験結果.....35
第 6 章　議論と考察.....40
6.1 提案手法の評価.....40
6.2 SNS としての質の担保.....42
6.3 ソーシャルエンジニアリングに関する知識の影響.....42
6.4 新しいソーシャルエンジニアリングの対策手法.....44
第 7 章　将来課題.....46
第 8 章　結論.....47

参考文献.....48

謝辞.....51

図目次

図 4.1　サーバの概略図.....14
図 4.2 クライアントのタイムライン画面.....15
図 5.1　攻撃投稿が表示された画面.....19
図 5.2　各シチュエーションに対して 3 段階に変換した回答.....27
図 5.3　攻撃投稿とそれに対する反応が表示された画面.....31
図 6.1　シチュエーション 1,9,10,11 に対して 3 段階に変換した回答.....41
図 6.2　シチュエーション 4,12,13,14 に対して 3 段階に変換した回答.....41
図 6.3　各グループにおけるシェアを行うかどうかの回答比率.....43

表目次

表 4.1　サーバに対するリクエストとそれに対するサーバの処理.....13
表 4.2　データベースに作成するテーブルとその内容.....13
表 4.3　作成したサーバサイドの開発環境.....15
表 5.1　8 つのシチュエーションとそれに対応するパラメータ.....18
表 5.2　質問 5.1 に対する回答.....22
表 5.3　質問 5.2 に対する回答.....23
表 5.4　質問 5.5 に対する回答.....23
表 5.5　質問 5.6 に対する，シェアを行う場合における回答.....23
表 5.6　質問 5.6 に対する，シェアを行うか決められない場合における回答.....24
表 5.7　質問 5.6 に対する，シェアを行わない場合における回答.....24
表 5.8　質問 5.7 に対する回答（一部抜粋）.....25
表 5.9　質問 5.8 に対する回答（一部抜粋）.....26
表 5.10　質問 5.9 に対する回答（一部抜粋）.....26
表 5.11　攻撃投稿内容と提案手法に対するシチュエーション番号.....31
表 5.12　質問 5.10 に対する回答.....35
表 5.13　質問 5.11 に対する，シェアを行う場合における回答.....35
表 5.14　質問 5.11 に対する，シェアを行うか決められない場合における回答.....36
表 5.15　質問 5.11 に対する，シェアを行わない場合における回答.....36
表 5.16　質問 5.12 に対する回答（一部抜粋）.....37
表 5.17　質問 5.13 に対する回答.....37
表 5.18　質問 5.14 に対する回答.....38
表 5.19　質問 5.15 に対する回答.....38
表 5.20　質問 5.16 に対する回答（一部抜粋）.....38
表 5.21　質問 5.17 に対する回答（一部抜粋）.....39

第 1 章　序論

1.1 研究背景

「ソーシャルエンジニアリング攻撃」は，人間の心理につけこみ個人情報などの機密情報を不正に手に入れるハッキング手法の 1 つである[1]．ソーシャルエンジニアリングに関する研究は日々されており，その中でも近年はネット社会における攻撃が注目されている．コンピュータサイエンス技術の発展に伴いネットセキュリティ技術は向上しており，ユーザはセキュアなネット環境を利用できつつあるが，その反面でネット社会における人間の脆弱性につけこんだにソーシャルエンジニアリング攻撃の被害が増えている[2]．さらに，ソーシャルエンジニアリング攻撃は被害がもたらされるかどうかが攻撃を受けた個人に依存するために，技術的アプローチで対策することが困難である[3]．

ネット社会において，一般ユーザ同士が繋がれるソーシャルネットワークサービス(SNS)が登場し，ソーシャルエンジニアリング攻撃はより一層脅威となっている．企業に働くユーザを特定し，そのユーザを対象に SNS 上で取引企業のクライアントを装いコンタクトを取り，そこからダイレクトメッセージなどを通して企業の機密情報を盗み取る標的型攻撃や[4]，ソーシャルエンジニアリングに繋がる SNS の投稿をシェアによって拡散するような高度なフィッシング攻撃など[5]，様々な種類の人間の脆弱性につけこんだ攻撃による被害が報告されている．前者の攻撃は 2018 年に生じた Coincheck 社における仮想通貨 NEM の流出事件の原因であるともいわれており，ネット上における人間の脆弱性をついた最大級被害を出した攻撃として被害が報告された[6]．後者の攻撃は，シェアをすることで良い事があるという内容で攻撃のきっかけとなる投稿をシェアさせ，信頼できるユーザから悪意なく無差別に攻撃投稿が拡散されるという事例を生じさせている[7]．

ネット上におけるソーシャルエンジニアリングに対して攻撃を検出しフィルタリングすることで対策する研究は多くされているが，SNS などの信頼できるサイト上で行われる攻撃は対象としていない．また，攻撃の被害を防ぐためにユーザの教育を促す研究は多くされているが，攻撃投稿の拡散そのものを防ぐことに焦点は置かれていない．そこで本研究では，SNS におけるソーシャルエンジニアリング攻撃に焦点を当て，アプリケーションレベルでユーザの行動を誘導し，攻撃投稿の拡散を対策できるような手法を論じる．

1.2 研究目的

ネット上におけるソーシャルエンジアリング攻撃は様々な対策手法が論じられているが，一方でその被害数も増加している．攻撃を対策できる手法があったとしてもそれで完全に守り切ることはできないということ，新たな攻撃手法が生み出されること，攻撃投稿を目

にする機会が増えることによって潜在的被害者になり得るユーザが増加していることが理由に挙げられる(ここでの潜在的被害者とは，経験や知識などの欠如によりソーシャルエンジアリング攻撃の被害者になる可能性のあるユーザのことを指す)[8]．これらの問題を解決するためには，既存の攻撃をより高精度で対策する手法を考案する，それとは完全に異なる新しい攻撃に対策できる手法を行うなど，それぞれの要因に対する様々な方向からの対策を実施する必要がある．そのために，攻撃が実行されるプラットフォームや(例えばWeb サイト，メールシステム, SNS など)，対策を実施させる対象(例えばサービスレベルで対策するのか， ユーザに対策をさせるのか，など)を明確にしたうえで，どのような手法がどれほど効果的であるのかを具体的に調査したデータが必要である．

そこで本研究ではソーシャルエンジニアリング攻撃を SNS に限定し，サービスレベルにおけるアプリデザインの特定機能の実装によってユーザの行動を誘導し，攻撃投稿の拡散を減らすことができないか評価検討する．本研究がコンピュータサイエンスのセキュリティ分野にもたらす貢献は, SNS におけるソーシャルエンジニアリング攻撃がアプリデザインレベルで対策できるということを示すことである．

1.3 論文の構成

本論文の構成を以下に示す．

第 2 章　関連研究

本研究のテーマであるソーシャルエンジニアリング攻撃について，アナログな手法としてのソーシャルエンジニアリング，ネット社会におけるソーシャルエンジニアリング, SNS におけるソーシャルエンジニアリングに分けて述べる．また，一般的なソーシャルエンジニアリングの対策手法についても述べる．

第 3 章　提案手法

本研究における提案手法，およびその前提条件について述べる．

第 4 章　アプリケーションの設計と実装

本研究における提案手法を実験するために用いたアプリケーションの概要，設計，および実装について述べる．

第 5 章　評価実験

評価実験の概要と目的，実験方法について述べる．また，評価実験をするにあたって行った事前実験の概要，方法，条件，結果について述べ，それを基に評価実験の方法，条件，結果について述べる．

第 6 章　議論と考察

評価実験の結果を用いて，ソーシャルエンジアリングの対策手法について議論を行う．また，ユーザが投稿をシェアする要因について考察する．

第 7 章　将来課題

本研究で生じた新たな課題について述べる．

第 8 章　結論

本研究の結論について述べる．

第 2 章　関連研究

2.1 ソーシャルエンジニアリング

本研究のテーマであるソーシャルエンジニアリングに関する研究について述べる．ソーシャルエンジニアリングという概念は 1990 年代，情報技術が発展するに伴って誕生した．その概念は，特別な技術やツールを用いずに，個人に関する機密情報を取得する手法である．Berg はソーシャルエンジニアリングの事を，システムに侵入するのではなく，必要な情報(パスワードなど)を個人から取得するためのハッカーの専門用語であると述べている[1]．また，同研究においてソーシャルエンジニアリングの実例を挙げており，例えばゴミ箱に捨てられた文書やハードウェアを漁ることで企業の機密データ(ログイン名とパスワードの印刷など)を得る Dumpster Diving，従業員として外線で電話をかけ機密情報を口頭で得るなりすましが存在することを指摘した．

同じくソーシャルエンジニアリングの被害が生じた実例を Winkler らは 挙げており，起業の機密情報を得るためにゴミをあさるだけでなく，警備員として会社に潜入すること，さらには会社に雇用される手段すら選ぶことがあると述べている．これらの手法について，人間としての脆弱性をついた攻撃を対策することは重視されていない現状があることを述べている[9]．

Tolga らはトルコ公務員のソーシャルエンジニアリングに対する意識を調査しており[10]，トルコ人の公務員である 56 人に攻撃を行ったところ電話で 38 人(68%)のユーザ名とパスワードが得られたことを報告した．このことから人間が一番の脆弱性であり，情報セキュリティの認識が最も重要であることを述べている．

Aksha らは，信頼に対する自然な傾向を悪用する方が簡単であるため，ハッカーは技術よりもむしろ人々の信頼を悪用することを好むことを主張している[3]．また，ソーシャルエンジニアリングが企業や組織のセキュリティに対して最大の脅威であること，そしてそれに対する技術的な解決手法が存在しないために最も研究不足で最も効果的なサイバー犯罪の 1 つであるということを述べた．同じく Harl は人々へのハッキングに関する研究で[11]，ハッカーにとって，クラッキングを行うよりもソーシャルエンジニアリングを行う方が簡単な場合があることを指摘しており，それに対してシステム管理者が人間の脆弱性につけこまれないためにもソーシャルエンジニアリングを防止および検出できる従業員を持つべきであると述べている．また，それが UNIX のシステムを保護することとりもはるかに少ない労力で済むことも述べている．

2.2 ネット社会におけるソーシャルエンジニアリング

ソーシャルエンジニアリングは技術的な手法ではなく，オンラインでなくオフラインで行われることが多かったが，メーリングシステムや Web サービスの発展により，ネット社会におけるソーシャルエンジニアリング攻撃が行われるようになった．Jonathan はインターネット詐欺に関して調査した研究において，ハッカーがインターネット詐欺のために用いる手法について次のように説明している：被害が生じる理由は，説得に関する社会心理学のうち「コミットメントと一貫性」が最も大きな要因である．文章を用いることでそれを生成した本人の意思に関わらず，文章を受け取る人に一貫性を訴えやすくなる[12]．インターネットがテキストベースの通信に依存しているため，コミットメントと一貫性をもった説得が無条件で達成され，インターネット詐欺の被害が生じやすい環境となる．

Mironela は，インターネットにおけるソーシャルエンジニアリングが 7 つの要素(貪欲，恐怖，緊急，好奇心，向上，敬意，信頼)で成り立っていると述べている[13]．また，その実例として Yahoo Messenger, Facebook，そして LinkedIn の脆弱性について，ソースコードから機密情報が辿れること，古いバージョンのソフトウェアを使うことで機密情報が取得できることを指摘している．

また別のインターネットにおけるソーシャルエンジニアリングとして，Jagatic らはソーシャルフィッシングという攻撃について述べた[14]．この攻撃では電子メールを通じて，大手金融機関や人気ショッピングサイトなどになりすまし，被害者から機密情報を盗み取る．この研究では実際に機密情報を提供してしまった被害者の数を調査しており，メールを受け取ったユーザのうち 19%のユーザがリンクをクリックし，3%のユーザが実際に機密情報を提供したと報告している．このフィッシング攻撃に対し Brandon らは，電子メールの要件として警告とアカウントの確認が，メール受信者の注意を引くための 2 つの主なトリガーであることを報告した[15]．

これらのインターネットにおけるソーシャルエンジニアリングは，対策技術が十分になりつつも被害が生じ続けている．その理由について Jussi-Pekka は，「被害者がセキュリティは必要ないという勘違いをしていること」，「警告やポップアウトに慣れてしまったこと」，そして「コンピュータ知識が欠如していること」を挙げており，いずれもユーザ依存であることを示した．その対策としてユーザートレーニングが効果的であること，そして Web 証明書を最新のものに更新すること挙げており，逆に警告やポップアウトを出すとユーザの作業を妨害することで逆効果になりうることを指摘している[16]．

2.3 SNS とソーシャルエンジニアリング

SNS はソーシャルネットワーキングサービス(Social Networking Service)の略称である．Danah らは SNS を「Web ベースのサービスであり，1) 閉じたシステム内における自身プロフィールを作成できること， 2) コネクションを共有している他のユーザのリストを明確にできること，3) コネクションのリストとシステム内の他のユーザによって作成されたコネクションのリストを表示，検索できること」を満たすサービスであると定義している[17]．また Nick は，SNS をユーザがコンテンツ，知識，そして経験を他の人々と共有できるオンラインプラットフォームであり，オンライン活動を促進することを述べた[18]．しかし SNS の普及に伴って莫大な数の個人情報が共有されたため，新たなソーシャルエンジニアリングが出現し，脅威となっている．

Nick らは，SNS での公開された個人情報が，ユーザの知識や同意の有無にかかわらず簡単に収集，公開，使用することができ，知らないうちにビジネス目的のために悪用される可能性があることを指摘しており，詐欺や盗難など危険に対して脆弱であることを述べた[19]．また，Jan らは同じように SNS において，公開されている個人プロフィールが機密情報を盗み取るためのソーシャルエンジニアリング攻撃に悪用される可能性を指摘しており，人々の情報セキュリティに対する意識と実際にプロフィールとして公開している情報にギャップがあるという，プライバシーパラドックスが存在することを述べた[20]．

具体的に Ralph らは Facebook を対象に，4000 人以上のユーザに対して開示する情報の量および要因を調査し，情報の開示理由が情報を開示することで得られる利益が期待以上であることが一つの要因であることを示した[21]．また，プライバシー設定に関する潜在的な脆弱性を強調した後であっても，安全なプライバシー設定に変更をしたユーザがわずかであったことも示した．

SNS におけるソーシャルエンジニアリングの他の例として，Alex は Twitter におけるスパムを報告している[22]．その特徴として，不特定多数のユーザに対しスパムサイトへ通ずる URL メッセージを送信しクリックをさせる．Twitter では 3%以上のメッセージがスパムであることを述べている．また攻撃者は bot 検出から逃れるために，短縮サイトなどを用いて異なる URL を生成し，同じ目的地へと誘導を行うことを述べている．同研究では機械学習を用いることで，Twitter におけるスパムを 89%の精度において検出に成功している．また，Sarita らは Twitter におけるスパマーの特性を分析しており，一般ユーザと比較して，アカウントを登録した日付とフォローとフォロワーの比率の違いはないが，投稿の頻度が高く，フォロー数，フォロワー数が多いことを示した[23]．

2.4 ソーシャルエンジニアリングの対策

最後に，これまでに述べたソーシャルエンジニアリングに共通する対策手法について述べる．

Fatima らはソーシャルエンジニアリングの対策手法について，「監査，ポリシー」と「教育，練習，認知」の 2 つの人間的アプローチと，「バイオメトリクス」，「センサー」，「AI」，そして「ソーシャルハニーポット」の 4 つの技術的アプローチがあることを述べている[8]．具体的な手法として，セキュリティ教育の促進，新入社員向けのセキュリティオリエンテーション，攻撃の検出ツールの利用などが挙げられている．また著者は，ソーシャルエンジニアリングを完全に対策することは困難であるために，被害を無くすアプローチではなく被害の数を軽減するアプローチが主流であることを述べている．これに関して Brandon らも，オンライン詐欺が完全に排除されることはないことを指摘しており，ソーシャルエンジニアリング攻撃を対策するための手法として重要なことが，加害者からの潜在的な脅威について公衆を教育し被害の数を減らすことであると述べた[15]．

Thomas は心理学の観点からソーシャルエンジニアリングに 3 つの重要な側面があることを指摘しており，説得の代替案，人間反応に影響する態度や信仰，説得・影響の技術に深く関わることを述べている[24]．人間心理が根本の原因であるためどれだけセキュリティ対策をしても脆弱性が生まれることを指摘したうえで，対策には教育および教育されることに対する受け入れが必須であることを述べている．しかし，Jussi-Pekka はセキュリティに対する感情に関する研究が未発達であることを述べており，それがインターネット上における人間の行動にどう影響するのか調査する必要があると述べている[16]．

第 3 章　提案手法

3.1 ソーシャルエンジニアリングの対象

提案手法について述べる前に，本研究で取り扱うソーシャルエンジアリング攻撃に関して述べる．1.1 節で述べたように，攻撃が実行されるプラットフォーム，および対策を実施させる対象を明確にする必要がある．そこで本研究では，攻撃が実行されるプラットフォームをオープンなショートメッセージ投稿型の SNS とし，対策を実施させる対象をサービス提供者と定義する．

ここでいうオープンな SNS とは，ユーザをフォローしなくともメッセージの投稿が確認できることを言う．フォローしたユーザがシェアをした投稿が自分のタイムラインにも表示され，そこでソーシャルエンジニアリング攻撃につながる投稿が表示されるというシチュエーションを考える．

サービス提供者は, SNS アプリケーションに特定の機能を付けることで対策を実施し，それを利用したユーザの傾向の変化を観測する．ここで重要なのが，サービスを利用するユーザがこの観測のために実装された特定の機能を自然に利用し行動を決めることである．対策を実施させる対象がサービスを利用するユーザと重なることで，観測されたデータの複雑性が増し,要因分析が困難となってしまうために，最善の注意を払う必要がある．

また，攻撃の内容は「懸賞詐欺を偽るフィッシング投稿」とする．これは，「ユーザがシェアをすることで当選の権利を得られる」と偽ることで，投稿を拡散しより多くの個人情報を盗むという攻撃である[7]．投稿をシェアした段階では個人情報は収集されないが，後にダイレクトメッセージにて当選の文字と共に住所，名前，クレジットカード情報などを盗み聞く．ユーザはシェアの段階で個人情報を提供することや手間をかけることが無いために，極めて簡単に悪意なく攻撃投稿のシェアを行えることが問題である．

3.2 シェアの重みづけ

3.1 節で述べたソーシャルエンジアリング攻撃を対策するための手法として，シェアの重みづけを提案する．これはつまり，アプリケーションレベルで投稿をシェアする際に追加の情報をユーザに提示する機能を実装する．既存の SNS におけるシェア機能は，投稿に対してシェアを行うための画面に遷移するシェアボタンをタップし，次にシェアを実際に行うためのボタンをタップすることで完了するような，二段階の同意を取るものが多い[25][26]．

例えば, SNS として代表的である Facebook では，シェアを許可しているユーザの投稿のみシェアするためのシェアボタンが設置され，それをタップすることでシェアを行うため

の詳細な画面へと遷移する．そこでは「投稿を自分のタイムラインにシェアする」，「投稿にコメントを添えて自分のタイムラインにシェアする」，「ダイレクトメッセージの形式で特定のユーザにのみシェアする」，「特定のグループに対してのみシェアする」，「友達のタイムラインにシェアする」のいずれかを選択する事ができ，コメントや送る対象を選ばない限りはそれを選択した時点でシェアが完了する仕組みになっている．シェアを行う対象であるユーザの範囲は，ユーザ設定から選択する事もできる．また，SNS として比較的大きな Twitter では，投稿にシェアをするためのボタンが設置されており，それをタップすることでシェアを行うための確認ポップが現れる．そこでは「投稿をそのままシェアする」，「投稿にコメントを添えてシェアするか」のいずれかを選択する事ができ，コメントを添えない限りはそれを選択した時点でシェアが完了する仕組みになっている．

本研究における提案手法では，シェアを確定させる 2 段階目の同意を取る直前にユーザに追加の情報を提示する機能をアプリケーションに実装する．追加の情報として，その投稿に対する反応の中で意味的な文章である反応を提示する．2.3 節で述べたように，意味的な文章の提示はユーザの判断を大きく変える．つまり，SNS 上における投稿をシェアするかどうかの判断も，意味的な反応を提示することで変わる可能性がある．ここで，本研究で用いる意味的な文章である反応を 2 種類用意し，次のように定義する；ポジティブな文章である反応とネガティブな文章である反応である．これらの反応を「それぞれ単体で提示した場合」，「２つを同時に提示した場合」の 3 つのシチュエーションに分けて実験を行う．

ネガティブな文章である反応が攻撃投稿と共に提示された場合，それらを見るユーザはその投稿に対して消極的な印象を持つ[27]．消極的な印象を持つことで，ユーザはその投稿をシェアしようと考えなくなると考えられる．ユーザが投稿をシェアしようと考えなくなることは，攻撃投稿の拡散を防ぐことができることと同義である．しかしネガティブな文章である反応のみが提示された場合，攻撃投稿以外の一般的な投稿に対しても消極的に印象付けられることで，SNS 全体としてアクティブ数が低下し質の低下が生じる可能性がある[28]．そこで，ポジティブな文章である反応を用いる．ポジティブな文章である反応が投稿と共に提示されることで，それを見たユーザはその投稿に対して積極的な印象をもち，シェアしようと考えること予想できる．これにより SNS 全体としてのアクティブ数の増加をもたらし質の向上を生じさせるが，ソーシャルエンジニアリング攻撃の投稿に対しても積極的に働きかけてしまう可能性が生じる．このことから，SNS の質の確保とソーシャルエンジニアリングの対策はトレードオフである．

このトレードオフを考えると，2 種類の反応を同時に提示してもユーザの判断は変わらないように思えるが，投稿に対する意味的な文章を提示するという点で，何も提示しないときと比較してユーザの判断は変わるはずである．この場合，ユーザは意味的な文章を見ることで無意識的に注意深くなり，結果として攻撃投稿のシェアをしようと考えなくなることが予想される．

提案手法に関して実施した評価実験については 4 章以降にて述べる．

3.3 提案手法の前提条件

本研究における提案手法の前提条件について述べる．

3.3.1 攻撃投稿の拡散

提案手法では，攻撃投稿にユーザからの反応が付いていることが前提条件である．要するに，ある程度の拡散が生じた状況から新たに拡散をさせないようなアプローチとなる．この手法では拡散されていない攻撃投稿に関して対策をすることはできないが，本研究の主目的は攻撃投稿の拡散を防ぐことであり，拡散されていない投稿に関してはその対象から外れている．ここで，「拡散されている」の定義を明確にする必要がある．

「拡散」の指標として考えうるのは「投稿に対するシェア数」，または「ユーザのタイムラインに表示された回数」であり，前者と後者の間には相関性がある．また，「投稿に対する反応の数」と「投稿が表示された回数」の間には相関性がある[29]．さらに，ポジティブとネガティブは極性でありパラメータ化できるため，「投稿に対する反応」は「投稿に対するポジティブな反応」と「投稿に対するネガティブな反応」の 2 つに，相対的に極性を判別することができる．以上のことから，本研究における「拡散」を「投稿に対するポジティブな反応とネガティブな反応それぞれがついた状態」と定義する．

投稿の初期状態は「拡散されていない」である．拡散されていない攻撃投稿が，シェアされるなどの要因によりポジティブな反応とネガティブの反応が付いた時点で，「拡散されている」状態へと遷移する．この定義により，「拡散されている」投稿すべてに対して本研究における提案手法を適用することができる．

3.3.2 反応の極性判別

提案手法では，投稿に対する反応の極性をすでに判別したものとして取り扱う．要するに，投稿に対する反応があったときに，それを「ポジティブな反応」か「ネガティブな反応」かに振り分ける工程は取り扱わない．あらかじめ 2 種類のうちどちらかに判別された反応のみが存在する．

本研究では反応の極性判別をブラックボックスの機能として進めるが，反応を振り分ける手法に関してはこの節にて述べておく．反応の極性判別として，例えば「文章に包含されるポジティブワード，ネガティブワードによる振り分け」が挙げられる．例えば Peter D. Turney らは単語の感情極性を判定するために，すでに極性が分かっている単語と極性を調べたい単語がインターネット上においてどの程度近接しているのかそのつながりを計算して極性判定する手法，そして潜在意味解析を用いた極性判定手法を提案し，95%を超える制度で極性の判定が行われたことを示した[30]．

また, Maite らは複数の知識源となりうる辞書を用いた単語ベースによる文章の感情分析手法を提案しており, 複数のプラッフォームにおいても感情分析の優れたパフォーマンスをもたらすことを示した[31]. これを応用することで, ある文章が「ポジティブ」な意味を持つのか「ネガティブ」な意味をもつのか容易に判別することができる.

第 4 章　アプリケーションの設計と実装

4.1 アプリケーション概要

本研究における評価実験を行うために必要なアプリケーションを製作した．アプリケーションは仮想的な SNS であり，既存の SNS の API を用いることで普段使いと同様の環境で利用できるようにする．仮想 SNS の基となる API として，Twitter のものを用いた．ソーシャルエンジニアリング攻撃に関するシェア比率を調べるために，仮想的な攻撃投稿をアプリ側で用意し，ユーザへ提示を行う．ここで言うシェア比率とは，実験を行ったユーザに対してシェアをしようと考えたユーザの比率のことであり

$$\text{シェア比率} = \frac{\textit{攻撃投稿のシェアをしようと考えたユーザ数}}{\textit{実験を行った全ユーザ数}}$$

と定義する．この時，アプリケーション側で投稿に対する反応を強制的に提示する機能も実装する．提案手法を取り入れていない従来通りのシチュエーションと，提案手法を取り入れたシチュエーションの両者を実施し，そのシェア比率を比較することで提案手法が有効であるかを調査する．

4.2 アプリケーションの設計

本アプリケーションは，サーバクライアント方式により通信を行う．この節ではサーバとクライアントに分けてアプリケーションの設計について述べる．

4.2.1 サーバの設計

サーバは Web サーバである．その概要として，クライアントがタイムラインにアクセスすると，API を経由してそのユーザがフォローしているユーザの投稿がクライアントに返される．サーバでは主に，実験に参加したユーザに関する情報とアプリケーション側で用意する攻撃投稿に関する情報を取り扱う．また，ユーザ情報などをデータとして保存するために，データベースを用い処理を行う．API を用いて Twitter に対して必要なデータを要求し，その返答を受け取る処理もサーバが行う．基本的にサーバはクライアントから要求された HTTP リクエストを受け取ると，適切な処理を行いクライアントに返答を返す． 詳細な設計として，本アプリケーションで取り扱うサーバに対するリクエストと，それに対するサーバの処理および返答を表 4.1 に示す．また，データベースに作成したテーブルの詳細を表 4.2 に示す．本アプリケーションにて設計したサーバの概略図を図 4.1 に示す．

表 4.1　サーバに対するリクエストとそれに対するサーバの処理

リクエスト	処理内容
/index	実験を行うためのメインメニュー画面を返す
/signup	クライアントに，実験に参加するために必要なユーザ登録を要求するためのフォームを返す
/auth/twitter	Twitter の API を利用するために，ユーザ認証を行う処理をクライアントに要求する
/storeUser	クライアントから受け取ったユーザ情報をデータベースに保存する
/timeline	クライアント情報に対応するタイムラインを twitter の API を用いて取得し，それをクライアントに返す．このとき，データベースに保存されている攻撃投稿情報をシチュエーションに応じて参照し，クライアントへ返すタイムラインの投稿の中に組み込む処理も行う．また同時に，その攻撃投稿をシェアするかしないか，そしてその理由を回答するためのフォームも同時に返す
/storeAnswer	攻撃投稿をシェアするかしないか，その回答をクライアントから受け取りデータベースに保存する
/questionnaire	実験を終了したクライアントに対して，本研究における最終アンケートの回答フォームを返す
/storeAnswer2	最終アンケートに対する回答をクライアントから受け取りデータベースに保存する

表 4.2　データベースに作成するテーブルとその内容

テーブル名	カラムの内容
Userlist	実験を行うユーザに関する情報： ユーザ ID，性別，年齢，実験の進行状態，その他ユーザに関する事前情報
Situations	攻撃投稿の内容に関する情報： シチュエーション ID，攻撃投稿を行う仮想ユーザの識別番号，攻撃投稿情報(投稿内容，シェア数，お気に入りの数)，提案手法を用いるか用いないかの識別番号
Spamusers	攻撃投稿を行う仮想のユーザ情報： ユーザの名前，アイコン画像，プロフィール，フォロー数，フォロワー数

Answer1	攻撃投稿をシェアするかに対する回答情報： 回答したユーザ ID，回答したシチュエーション ID，投稿をシェアしたいと思うか，その理由，その他コメント(自由記述)
Answer2	最終アンケートの回答情報： 回答したユーザ ID，アンケートに対する回答

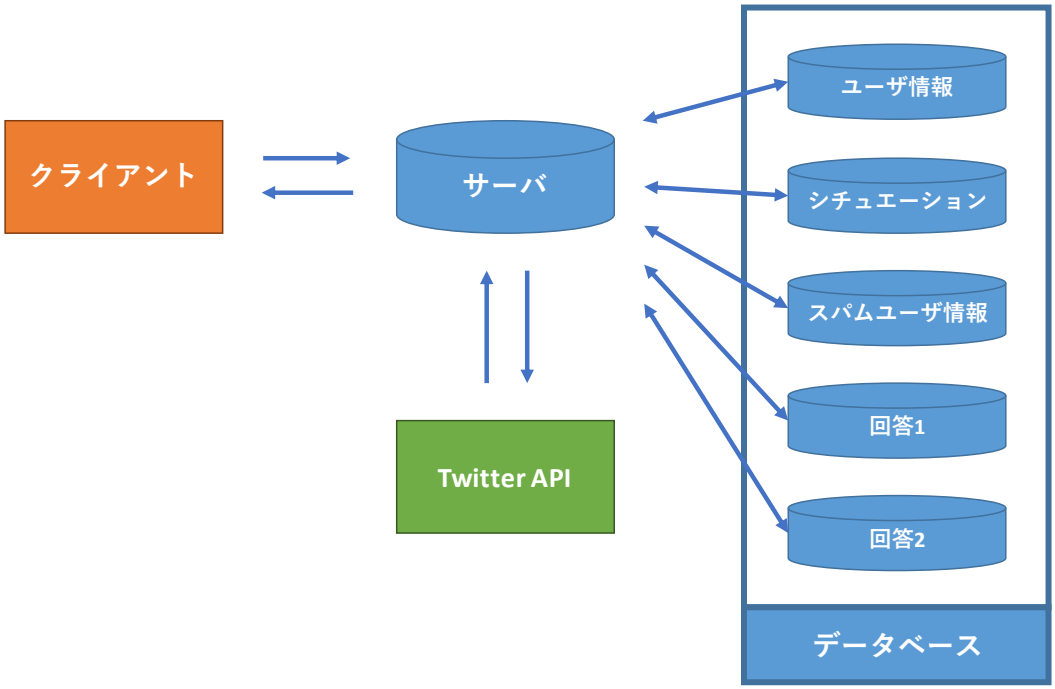

図 4.1　サーバの概略図

4.2.2 クライアントの設計

クライアントは Web ブラウザを通じて仮想 SNS を利用する．クライアントは専用のアプリケーションをダウンロードする必要はなく，Web ブラウザを搭載していればどの端末でもアクセスができる．また，ユーザから自然な回答を得るために，普段使いの SNS とサイトの UI を可能な限り再現する．そのために，デスクトップ端末とモバイル端末の 2 種類の端末用の UI を Twitter に模して設計し，本アプリケーションへの慣れを必要なく実験を行える造りとした．UI に関するテンプレート含めてサーバからクライアントに送信するため，クライアント側で行う処理はページのリクエストと回答の送信のみである．

クライアントが Web ブラウザを通じて表示されるタイムライン画面を図 4.2 に示す．

図 4.2 左）デスクトップクライアントのタイムライン画面，右）モバイル端末クライアントのタイムライン画像

4.3 アプリケーションの実装

4.2 節にて述べたアプリケーションの設計を踏まえ,サーバサイドの実装を行った．実装のために用いた開発環境について，表 4.3 にまとめた．

表 4.3　作成したサーバサイドの開発環境

サーバ	Amazon EC2
言語	Ruby v2.6.3
使用したフレームワーク	Ruby on Rails v5.2.3 jQuery v1.12.4
データベース	sqlite3 v3.28.0

第 5 章　評価実験

5.1 実験の概要と目的

シェア時のユーザに対する追加の情報提示がソーシャルエンジニアリング攻撃の対策として効果的であるかどうかを調査するために, 4 章で述べた仮想 SNS アプリケーションを用いて実験参加者からの回答を集める形式で実験を行った.

5.2 実験方法

作成したアプリケーションを実験参加者に利用してもらい, 各攻撃シチュエーションに応じる質問の回答を記録することで実験を行った.

ここで実験参加者の対象となる条件について述べる. 実験参加者は「普段から SNS(今回の場合は Twitter)を利用している」ユーザを対象として, Twitter で募り集められた. 集められたユーザ全員に対して収集するデータなどの内容について伝え, インフォームドコンセントを得た計 35 名のユーザ(10 代〜20 代を中心とする, 男性名 17, 女性 18 名)を対象に実験を行った.

ここで実験を行う前に, 実験結果により信憑性をもたせるために母体の偏りについて考慮する必要がある. ユーザ群を限定することで, 実験により判明した傾向が普遍的でなくなるが, 特定の群に対する 1 つの有益な手法を示せることに意義があるとされる[32]. 今回の場合, 「SNS を普段から利用しているユーザ」に対する実験結果の妥当性は主張できる. また, このユーザ群がもつ傾向として極めて特殊な傾向をもつ母体でないことを示すために, 既存の研究で述べられている信頼に関わる傾向の一致を確かめる必要がある. 小川らは週に 1 回以上 Facebook, mixi, ブログなどを利用して情報の発信をしている社会人を対象に, SNS においてユーザが第三者に情報を開示してしまう要因として, ユーザへの信頼, 情報開示範囲のコントロール, リスク認知, SNS でつながっている人数が挙げられることを示した[33]. また, 匿名非匿名によるものは要因ではないということも述べられている. この研究結果との傾向の一致を調査するために, SNS においてユーザが何を信頼の指標として捉えているのか, 事前実験という形式で確認調査を行う.

事前実験により既存研究とのユーザの傾向の一致が確認できれば, 実験対象となるユーザ群が特殊な傾向をもたないことを示せるため, 実験で得られるデータの信憑性が増す. 傾向が一致しなければ, その要因を分析した上でユーザ群を新たに定義し, その群中にて妥当性をもつ実験データを収集する.

実施した事前実験の詳細を 5.3 節に, 事前実験の結果を基に行う本実験の詳細を 5.4 節に述べる.

5.3 事前実験

事前実験では，本実験における参加者の群としてのユーザ傾向が，既存研究において判明したユーザ傾向と一致するかどうかを調査する．

5.3.1 調査内容

SNS における信頼性に関する既存の研究[33]では，SNS においてユーザが第三者に情報を開示してしまう要因として以下の要因を挙げた：

1. ユーザへの信頼
2. 情報開示範囲のコントロール
3. 自己顕示性
4. 情報管理に関する知識
5. SNS でつながっている人数

また，以下に挙げるものは要因でないことを明らかにした：

a. コンプライアンス意識
b. 匿名・非匿名によるもの

本実験の参加者がこれらの傾向を持つことを調査するために，要因それぞれに対するパラメータが異なるシチュエーションを用意してユーザに提示し，反応の比較を行う．また，事前実験で得られた結果を本実験でも用いて比較するために，用意するシチュエーションを次のように定義する：

SNS 上で自分のタイムラインを更新すると，最新の投稿として「この投稿をシェアすると良いことがある」という内容のものが知人よりシェアされた．これはソーシャルエンジニアリング攻撃に繋がる投稿であるが，あなたはそのことを知らない．あなたはこの投稿をシェアするか？

シェアを行うかどうかは，ユーザがその投稿を信頼するかしないかに依存するために，何の要因が信頼の基準となるのかを調査した既存研究との比較を行うことができる.このシチュエーションに沿った内容で，先に述べた情報開示の要因に関するパラメータが異なるシチュエーションをいくつか用意し，実験参加者がその投稿をシェアするかしないかの回答，およびその理由を募る．

5.3.2 パラメータの異なるシチュエーション

信頼の基準となる要因を調査するために，以下に挙げるパラメータを変えつつ異なるシチュエーションを生成する．

1. 攻撃投稿をシェアしたユーザが，信頼できるユーザか，フォローしているが関わりは

薄いユーザか(ユーザへの信頼に関する要因を調査するため)

2. *攻撃投稿を行ったユーザが，匿名であるか,非匿名であるか*
3. *攻撃投稿を行ったユーザの，フォロワーが多いか少ないか(SNS でつながっている人数に関する要因を調査するため)*
4. *攻撃投稿を行ったユーザの，プロフィールが詳細か，詳細でないか(匿名・非匿名による要因を調査するため)*
5. *攻撃投稿のお気に入り数・シェア数が，多いか，少ないか(調査内容には含まれないが，すでにシェアされているかどうかによる要因を調査するため)*
6. *攻撃投稿をシェアしたユーザ，全員が得をする内容か，確率で得をする内容か(調査内容には含まれないが，自己に降りかかる利益と現実性に関する要因を調査するため)*
7. *攻撃投稿を行うユーザが，架空の存在か，実在する企業/人物のなりすましか(匿名・非匿名による要因を調査するため)*

事前実験ではこれら 7 つのパラメータがそれぞれ異なる 8 つのシチュエーションを用意した．これら 8 つのシチュエーションと，それに対応するパラメータを表 5.1 に示す．また，実際に用いた攻撃投稿が表示された画面を図 5.1 に示す，

表 5.1　8 つのシチュエーションとそれに対応するパラメータ

パラメータ＼シチュエーション番号	1	2	3	4	5	6	7	8
シェアをしたのは信頼できるユーザか	×	○	○	○	○	○	○	○
投稿をしたのは非匿名か	×	×	○	○	○	○	○	○
投稿をしたユーザのフォロワーが多いか	×	×	×	○	○	○	○	○
投稿をしたユーザのプロフィールが詳細か	×	×	×	×	○	○	○	○
投稿のお気に入り・シェア数が多いか	×	×	×	×	×	○	○	○
シェアをした人全員が得をする内容か	×	×	×	×	×	×	○	×
実在する企業/人物のなりすましか	×	×	×	×	×	×	×	○

図 5.1　攻撃投稿が表示された画面

5.3.3 ユーザ情報の登録

事前実験を行うにあたって必要となるユーザ情報を収集するために，TwitterAPI を用いたユーザ認証，および 5.3.2 節で述べたパラメータを反映させるための実験参加者に対する質問の回答フォームを用意し回答を収集した．TwitterAPI を用いたユーザ認証と質問の回答フォームは，仮想 SNS におけるユーザ登録画面に組み込むことで，実験のためのユーザ登録を行うと同時にユーザ情報の収集が完了するようになっている．ユーザー登録時に質問した項目に関して、質問 5.1〜質問 5.4 として以下に示す．

質問 5.1

見るだけの時間も含めて，SNS を使用する頻度はどれくらいですか

- ○ 1 日当たり 2 時間以上（いつも使っている）
- ○ 1 日当たり 1 時間〜2 時間程度（だいたい使っている）
- ○ 1 日当たり 30 分〜1 時間程度（たまに使っている）
- ○ 1 日当たり 10 分〜30 分程度（1 日に数えられるくらい）
- ○ 1 日当たり 10 分未満（まれに使う）

質問 5.2

コンピュータに対する知識について教えて下さい

○ コンピュータに精通している（コンピュータ分野に関わっている）
○ そこそこ詳しい（趣味程度である）
○ 普通（平均くらいの知識である）
○ あまり詳しくない（無知ではない）
○ ほとんど知らない（コンピュータについてよく分からない）

質問 5.3

Twitter 上でフォローをしており，信頼しているユーザの名前を記入してください（自由記述）

質問 5.4

Twitter 上でフォローをしているが，関わりの薄いユーザの名前を記入してください（自由記述）

5.3.4 各シチュエーションに対する質問の項目

5.3.2 節にて述べた各攻撃シチュエーションに対して，実験参加者に投稿をシェアするかどうか，およびその理由を質問した．用いた質問の項目に関して，質問 5.5〜質問 5.7 として以下に示す．

質問 5.5

あなたはこの投稿を

○ シェアする
○ シェアしてもよい
○ 決められない
○ シェアはあまりしたくない
○ シェアしない

質問 5.6

その理由は？（複数選択可）

（シェアを行う場合）

□ 知人がシェアしたものだから

□ 信頼している人がシェアしたものだから
□ 興味がある投稿だから
□ 自分が得をするから
□ 自分が損をしないから
□ 投稿をした人のフォロー数を見て
□ 投稿をした人のフォロワー数を見て
□ 投稿をした人のプロフィールを見て
□ 投稿のシェア数を見て
□ 投稿のお気に入り数を見て
□ その他

（決められない場合）
□ 投稿をした人の詳細な情報がみたい
□ 投稿をした人の他の投稿がみたい
□ 投稿に対する他の人の反応がみたい
□ 投稿をした人について Web で調べたい
□ 迷っている
□ その他
（シェアを行わない場合）
□ 信頼できないから
□ 知らない人の投稿だから
□ 自分が得をしないから
□ 自分が損をするから
□ 投稿をした人のプロフィールを見て
□ 投稿をした人のフォロー数を見て
□ 投稿をした人のフォロワー数を見て
□ 投稿のシェア数を見て
□ 投稿のお気に入り数を見て
□ 興味がない投稿だから
□ スパムの可能性があるから
□ 投稿をした人に自分を知られたくないから
□ その他

質問 5.7
投稿に対して何かコメントがあれば記入してください（自由記述）

ここで，選択した回答の内容が他のシチュエーションを見たことに依存する可能性があるため，実験参加者には「対象となる類の投稿を初めて見た時」を想定した回答を依頼した．

5.3.5 事前実験終了後のアンケート項目

8つのシチュエーションに対して回答を終えたユーザに対してアンケートを実施した．質問の項目に関して，質問 5.8，質問 5.9 として以下に示す．

質問 5.8

あなたが何かの投稿をシェアする時の基準について，教えてください（自由記述）

質問 5.9

その他，本実験に関して何かコメントがあれば記入してください（自由記述）

5.3.6 事前実験の結果

実験参加者 35 名に対して，8 つのシチュエーションに対する質問の回答と事前実験後アンケートを完了させた．各質問に対する得られた回答を，表 5.2〜表 5.10 にまとめて以下に示す．質問 5.3，質問 5.4 に関しては，シチュエーションに応じた攻撃投稿のパラメータ変更のみに用いる回答であるため，本節では取り扱わない．

質問 5.1

見るだけの時間も含めて，SNS を使用する頻度はどれくらいですか

表 5.2　質問 5.1 に対する回答

項目	人数[人]
1 日当たり 2 時間以上（いつも使っている）	22
1 日当たり 1 時間〜2 時間程度（だいたい使っている）	9
1 日当たり 30 分〜1 時間程度（たまに使っている）	3
1 日当たり 10 分〜30 分程度（1 日に数えられるくらい）	0
1 日当たり 10 分未満（まれに使う）	1

質問 5.2

コンピュータに対する知識について教えて下さい

表 5.3　質問 5.2 に対する回答

項目	人数[人]
コンピュータに精通している（コンピュータ分野に関わっている）	6
そこそこ詳しい（趣味程度である）	8
普通（平均くらいの知識である）	16
あまり詳しくない（無知ではない))	4
ほとんど知らない（コンピュータについてよく分からない）	1

質問 5.5

あなたはこの投稿を

表 5.4　質問 5.5 に対する回答

＼シチュエーション番号	1	2	3	4	5	6	7	8
シェアする	3	4	3	3	4	3	4	5
シェアしてもよい	2	5	3	9	6	7	0	4
決められない	2	1	2	1	1	0	3	3
シェアはあまりしたくない	5	9	11	7	5	6	9	8
シェアしない	23	16	16	15	19	19	19	15

※単位は[人]である

質問 5.6

その理由は？（複数選択可）

表 5.5　質問 5.6 に対する，シェアを行う場合における回答

＼シチュエーション番号	1	2	3	4	5	6	7	8
知人がシェアしたものだから	2	2	2	3	5	5	0	2
信頼している人がシェアしたものだから	0	4	5	7	5	6	1	3
興味がある投稿だから	2	2	2	3	3	3	2	7
自分が得をするから	0	1	1	1	1	0	2	3
自分が損をしないから	2	3	1	4	3	3	2	4
投稿をした人のフォロー数を見て	0	0	0	1	0	0	0	0
投稿をした人のフォロワー数を見て	1	1	1	4	3	3	0	1
投稿をした人のプロフィールを見て	0	0	0	0	5	3	0	3
投稿のシェア数を見て	0	0	0	2	1	4	1	1
投稿のお気に入り数を見て	0	0	0	1	1	0	1	1
その他	0	1	0	0	1	0	1	5

※単位は[人]である

表 5.6　質問 5.6 に対する，シェアを行うか決められない場合における回答

＼シチュエーション番号	1	2	3	4	5	6	7	8
投稿をした人の詳細な情報がみたい	1	1	1	0	1	0	1	1
投稿をした人の他の投稿がみたい	1	1	1	1	0	0	1	2
投稿に対する他の人の反応がみたい	0	0	2	1	0	0	1	0
投稿をした人について Web で調べたい	1	0	1	0	0	0	1	1
迷っている	1	0	1	0	0	0	2	1
その他	2	0	0	0	0	0	0	1

※単位は[人]である

表 5.7　質問 5.6 に対する，シェアを行わない場合における回答

＼シチュエーション番号	1	2	3	4	5	6	7	8
信頼できないから	11	11	12	8	12	11	19	9
知らない人の投稿だから	11	10	13	10	9	9	7	2
自分が得をしないから	2	0	1	0	2	0	1	2
自分が損をするから	1	1	1	1	1	1	3	1
投稿をした人のプロフィールを見て	3	1	3	1	5	3	4	5
投稿をした人のフォロー数を見て	0	0	1	1	0	0	0	2
投稿をした人のフォロワー数を見て	0	0	1	0	0	0	0	1
投稿のシェア数を見て	3	4	3	2	2	4	0	1
投稿のお気に入り数を見て	3	3	2	2	3	0	0	1
興味がない投稿だから	16	12	12	11	10	10	9	10
スパムの可能性があるから	11	11	6	5	7	5	8	5
投稿をした人に自分を知られたくない	2	3	3	4	4	4	4	0
その他	1	2	4	2	3	2	2	1

※単位は[人]である

質問 5.7

投稿に対して何かコメントがあれば記入してください（自由記述）

表 5.8　質問 5.7 に対する回答（一部抜粋）

シチュエーション番号	回答内容
1	この人のツイート(投稿)を他に見て判断したい
1	フォローしている人の浮上数が少なく，タイムラインがほぼ動いていない場合は(シェアを)する可能性がある
1	タイムラインに自分の RT(シェア)が流れるのが嫌
4	真面目な方かどうかで決める
5	貰えるはずがないが，僅かな可能性にかけてみたいから
5	プロフィールが具体的になったことでさらに胡散臭い印象を抱いてしまう
6	知人との話題のネタになるから
6	RT(シェア)数が圧倒的に多く、自分が参加しても当たらないだろうという気持ちで拡散する意欲が下がる
7	先程の人とは違って丁寧だし、信頼できるとおもったから
7	名前の記載もあるが信頼度は微妙なので追加の情報を見て判断すると思う
7	明らかな嘘だと思うから
8	公式が出してるから
8	公式とかいてあるとシェアをしてしまうかもしれない（本当に公式か確かめはする）
8	公式の割にフォロワー数が少ない気がするので、もっとフォロワー数が多ければ RT すると思います
8	このアカウントが確実に公式アカウントと判断できる場合のみ RT します

質問 5.8

あなたが何かの投稿をシェアする時の基準について，教えてください（自由記述）

表 5.9　質問 5.8 に対する回答（一部抜粋）

回答内容
自分に不利益でないことや面白いこと，残しておきたいなと思ったらリツイート(シェア)します
面白かったり役に立つ情報等をツイートしていた時
何かおもしろい話題だったりプレゼント企画や，自分が協力できそうなツイート
自分のフォロワーさんにも見てもらいたい時，自分がまた見返したいと思う可能性がある時
面白いツイート(イラスト系)でパクツイではない場合
情報の正誤が問われるものについては公式マークがついているか，フォローしてからある程度時間が経過していて，信頼できるアカウント.それ以外については特に共感した内容のツイート(投稿)
信憑性のある情報発信系や，自分が実践してためになったこと，また意見が割れないようなもの(推しの派閥問題にならないようなもの)

質問 5.9

その他，本実験に関して何かコメントがあれば記入してください（自由記述）

表 5.10　質問 5.9 に対する回答（一部抜粋）

回答内容
自分はツイッターの RT(シェア)基準が適当だったのでこれを機会に自分の RT 基準をしっかり決めようと思いました
初めて twitter を使ったときをよりリアルに再現してくれればもっと具体的な答えが出しやすくなるのではないかと思います
リツイートした人が目に入り，その人が仲のいい人だとちょっと RT(シェア)したいと思いがちだった．関係してるんだと結構びっくりした
今になってはプレゼント企画などは詐欺が多いと聞くため「信頼出来ない」との理由で「RT しない」を選択しましたが，もしその情報が私の頭の中になければ RT してしまっていたかもしれません
プロフィールなどで判断する人は少ないと思う．普段からツイッターでツイート(投稿)をしている人はなんとなく信頼できる

5.3.7 事前実験の結果の分析，既存研究との比較

8つのシチュエーションそれぞれに対してシェアするか，しないかを示した結果が表5.4である．この設問では回答を5段階としているが，回答候補の「シェアする」と「シェアしてもよい」は「シェアを行う」に，「シェアはあまりしたくない」と「シェアしない」は「シェアを行わない」にまとめることができる．このようにして，結果を3段階に簡略化したものを図5.2に示す．

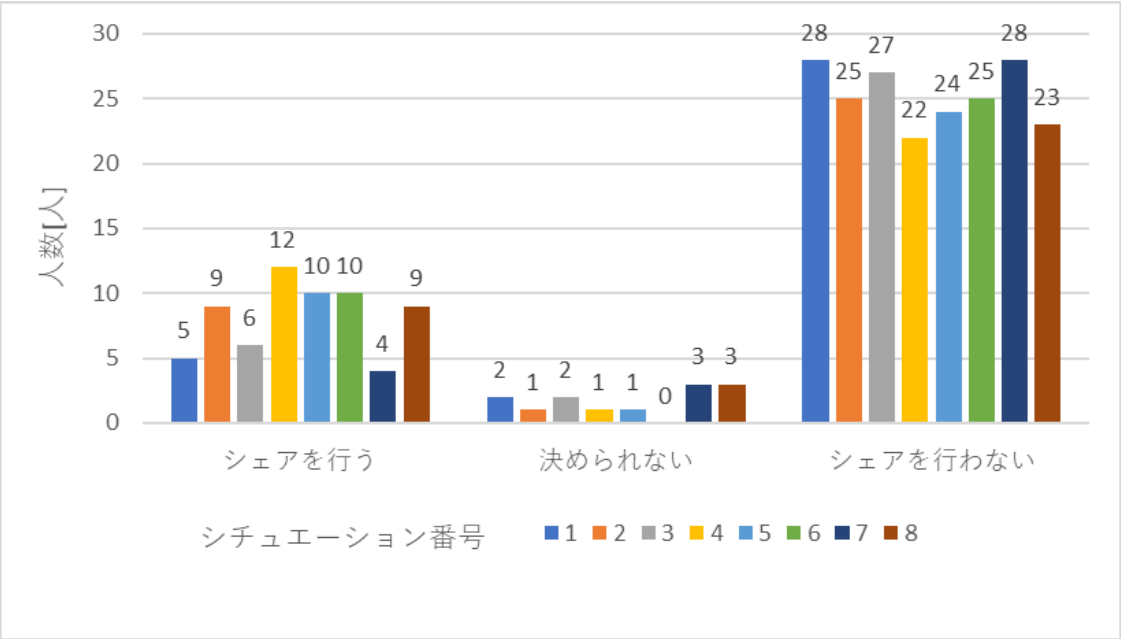

図5.2　各シチュエーションに対して3段階に変換した回答

シェアを行う要因

得られた回答よりシェアを行う要因に関して述べる．図5.2に示した結果および表5.1より，各シチュエーション同士を比較することでユーザが何の要因によりシェアを行うのか分析を行う．

シチュエーション1とシチュエーション2：

ここでの違いは「シェアをしたのは信頼できるユーザか」である．シェアをしたのが関わりの薄い人から，信頼できる人になったことで，シェアを行うと回答したユーザが5人から9人へと増加した．実際に表5.5を見ても，シェアを行う理由として「信頼している人がシェアしたものだから」と回答した人数が0人から4人へと増加した．また表5.10を見ると，実験に対するコメントとしてあるユーザは次のように述べた：

リツイートした人が目に入り，その人が仲のいい人だとちょっと RT(シェア)したいと思いがちだった. 関係してるんだと結構びっくりした
シェアをする人が信頼できる人になることで，攻撃投稿に対する信頼も生じるということが言える．以上の事から，シェアを行う要因として「ユーザへの信頼」が確認できた．

シチュエーション 2 とシチュエーション 3：
ここでの違いは「投稿をしたのは匿名か非匿名か」である．投稿したユーザが匿名から非匿名になったことで，シェアを行うと回答したユーザが 9 人から 6 人へと減少した．ここで表 5.7 を見ると，シェアを行わない理由として，非匿名に関わらず「知らない人の投稿だから」と回答したユーザが 10 人から 13 人へと増加している．以上のことから，ユーザがシェアを行う要因が「匿名・非匿名によるもの」には関係なく，「投稿を見た人がその人自身を知っているか，知っていないか」に関わることが分かる．

シチュエーション 3 とシチュエーション 4：
ここでの違いは「投稿をしたユーザのフォロワーが多いか」である．攻撃投稿を行ったユーザのフォロワーが増えたことで，シェアを行うと回答したユーザが 6 人から 12 人へと増加した．ここで表 5.7 を見ると，シェアを行わない理由として，ユーザ自体はシチュエーション 3 と同一であるのにも関わらず「信頼できないから」と回答したユーザが 12 人から 8 人に減少している．以上のことから，ユーザを信頼する要因およびシェアを行う要因として「SNS でつながっている人数」が確認できた．

シチュエーション 4 とシチュエーション 5：
ここでの違いは「投稿をしたユーザのプロフィールが詳細か」である．攻撃投稿を行ったユーザのプロフィールが明確になると，シェアを行うと回答したユーザは 12 人から 10 人へと減少したが，大きな変化は生じなかった．またその理由として表 5.8 を見ると，投稿に対するコメントとしてあるユーザは次のように述べた：
プロフィールが具体的になったことでさらに胡散臭い印象を抱いてしまう
これはプロフィールが詳細になったことで，攻撃投稿を行っているユーザに対するイメージが具体的になったことを意味しているが，それが直接信頼に繋がるわけではないと言える．以上のことから，ユーザを信頼する要因およびシェアを行う要因は「ユーザのプロフィール(匿名，非匿名に関連する)」に関わらないことが分かる．

シチュエーション 5 とシチュエーション 6：
ここでの違いは「投稿のお気に入り・シェア数が多いか」である．攻撃投稿のシェア数が多くとも少なくとも，シェアを行うと回答したユーザは 10 人で変化が見られなかった．つまり，シェアを行う要因は「攻撃投稿がすでに広まっているかどうか」に関わらないことが

分かる．

シチュエーション７：

シチュエーション 7 ではこれまでの攻撃投稿とは異なり，シェアを行ったユーザ全員に対して利益が得られるとアプローチする文章が提示された．その結果，シェアを行うと回答したユーザは 4 人に留まった．その理由として表 5.7 を見ると，シェアを行わない理由として「信頼できないから」と回答したユーザが 19 人であり，参加者のうち過半数以上を占めた．また表 5.8 を見ると，投稿に対するコメントとしてあるユーザは次のように述べた：
明らかな嘘だと思うから
文章通りの意味を取ればシェアを行う選択が最善である．しかし実際には，ユーザは投稿の内容に関する現実性をシェアの判断基準として考えるため，シチュエーション 7 における攻撃投稿はシェアされなかった．

シチュエーション８：

シチュエーション 8 でもこれまでの攻撃投稿とは異なり，攻撃投稿を行うユーザが既存の有名企業のなりすましであった．攻撃投稿の内容に関してはシチュエーション 6 と同じ条件で行ったため，シチュエーション 6 との比較を行う．攻撃投稿を行うユーザが有名企業のなりすましになると，シェアを行うと回答したユーザは 10 人から 9 人へと減少したが，大きな変化は見られなかった．また表 5.8 を見ると，投稿に対するコメントとしてあるユーザは次のように述べた：
公式とかいてあるとシェアをしてしまうかもしれない（本当に公式か確かめはする）
攻撃投稿を行ったユーザがよく知っている企業のなりすましであっても，それが本当にその企業のものでない限り，投稿のシェアがされないことが分かる．以上のことから，ユーザがシェアを行う要因が「匿名・非匿名によるもの」には関係なく，「投稿を行ったユーザが本当に自分の知っているユーザであるのか」に関わることが分かる．

既存研究との比較

以上に述べたことをまとめると，ユーザが攻撃投稿をしたユーザを信頼しシェアを行う要因は「ユーザへの信頼」と「SNS でつながっている人数」であり，「匿名・非匿名によるもの」は要因でない．これは，5.3.1 節で述べた既存研究とユーザの判断基準の傾向が一致している．ここで本実験参加者の群としての一般性が確認できたため，本研究における主張は「SNS を普段から利用している」ユーザ群に対して妥当であると言える，

5.4 評価実験

事前実験により実験参加者の群としての妥当性が確認できた．本節ではこのユーザ群に対して第 3 章で述べた提案手法が効果的であるかどうか，評価実験を行う．

5.4.1 調査内容

提案手法について改めて述べておくと，シェアを確定させる 2 段階目の同意を取る直前にユーザに追加の情報を提示する機能をアプリケーションに実装する．追加の情報として，その投稿に対する反応の中でポジティブな文章である反応とネガティブな文章である反応を用いる．これらを用いて「ポジティブな反応のみを提示する場合」,「ネガティブな反応のみを提示する場合」，そして「ポジティブな反応とネガティブな反応を同時に提示する場合」の 3 つのシチュエーションに分け実験参加者へ提示し，その投稿をシェアするかしないか質問し回答を得る．ここで，提案手法を用いらないときの回答と比較するために，事前実験で得られた結果を本実験でも利用する．事前実験と同様の形式で評価実験を実施するために，再掲ではあるが，攻撃投稿に関して用意するシチュエーションを事前実験の時と同様に次のように定義する：

SNS 上で自分のタイムラインを更新すると，最新の投稿として「この投稿をシェアすると良いことがある」という内容のものが知人よりシェアされた．これはソーシャルエンジニアリング攻撃に繋がる投稿であるが，あなたはそのことを知らない．あなたはこの投稿をシェアするか？

このシチュエーションに沿った内容で，提案手法を取り入れた形式の攻撃投稿を実験参加者へ提示し，その投稿をシェアするかしないかの回答，およびその理由を募る．得られた回答および 5.3.6 節における事前実験の結果を比較することで，提案手法が効果的であるかどうかの分析を行う．

5.4.2 提案手法を組み込んだシチュエーション

5.4.1 節で定義したシチュエーションを基に，実際に実験で取り扱う攻撃投稿の内容を生成する．提案手法が効果的であると主張するためには，「ユーザが攻撃投稿をシェアする数が減る」，もしくは「ユーザが攻撃投稿をシェアしない数が増える」事が必要条件である．また，提案手法がソーシャルエンジアリング攻撃対策に対して負の方向に効果が生じる可能性もあるため，「ユーザが攻撃投稿をシェアする数が増える」，もしくは「ユーザが攻撃投稿をシェアしない数が減る」事が起きないことを確認する必要がある．そのため，用いる攻撃投稿は事前実験で用いたシチュエーションの中で，「一番シェアすると回答した人数が多かったもの」および「一番シェアしないと回答した人数が多かったもの」を採用した．

図 5.2 を見ると，一番シェアすると回答した人数が多かったシチュエーション番号は 4 であったため，これを採用した．また，一番シェアしないと回答した人数が多かったシチュエーション番号は 1 と 7 であるが，シチュエーション 7 は他と異なる特徴をもった攻撃投稿であったために，単純比較のしやすいシチュエーション 1 を採用した．これら 2 つのシチュエーションに対してそれぞれ，「ポジティブな反応のみ」,「ネガティブな反応のみ」，そして「ポジティブな反応とネガティブな反応」を投稿と一緒にユーザに提示し，回答を得る．用いたポジティブな反応は，攻撃投稿に対して期待を込めるような一文である．また，用いたネガティブな反応は，詐欺であることを主張する一文である．

2 種類の攻撃投稿の内容と 3 種類の提案手法に対応したシチュエーションをそれぞれ 9〜14 と番号付けし，表 5.11 にまとめた．また，実際に用いた攻撃投稿とそれに対する反応が表示された画面を図 5.3 に示す，

表 5.11　攻撃投稿内容と提案手法に対するシチュエーション番号

採用した攻撃投稿内容 採用した提案手法	シチュエーション 1 （シェアしない人数が多いもの）	シチュエーション 4 （シェアする人数が多いもの）
ポジティブな反応のみ提示	シチュエーション 9	シチュエーション 12
ネガティブな反応のみ提示	シチュエーション 10	シチュエーション 13
ポジティブな反応と ネガティブな反応を提示	シチュエーション 11	シチュエーション 14

図 5.3　攻撃投稿とそれに対する反応が表示された画面

5.4.3 実験の流れ

評価実験を行うために必要な実験参加者の情報は，事前実験のユーザ登録時に回答したものを用いる．そのため，事前実験を実施したサイトと同じサイト上で，実験参加者に新たな URL を配布することで実験を実施する．サイトにアクセスすると自分のタイムラインが表示され，タイムラインの更新をするとアプリケーション側で用意した攻撃投稿が表示される．この時，投稿をシェアする画面へと遷移し，そこで攻撃投稿に対する反応が強制的に実験参加者に提示される．実験参加者はこれを見たうえでシェアを行うか行わないかの質問に回答する．その後，すべてのシチュエーションに対して回答をし終えたユーザに対してアンケートを実施する．その回答が得られた時点で実験は終了となる．

5.4.4 各シチュエーションに対する質問の項目

5.4.2 節にて述べた各攻撃シチュエーションに対して，実験参加者に投稿をシェアするかどうか，およびその理由を質問した．用いた質問の項目に関して，質問 5.10〜質問 5.12 として以下に示す．基本的に質問項目は 5.3.4 節のものと同じであり，異なる部分は**太字**で示している．

質問 5.10

あなたはこの投稿を

○ シェアする
○ シェアしてもよい
○ 決められない
○ シェアはあまりしたくない
○ シェアしない

質問 5.11

その理由は？（複数選択可）

（シェアを行う場合）

□ 投稿に対する反応を見て
□ 知人がシェアしたものだから
□ 信頼している人がシェアしたものだから
□ 興味がある投稿だから
□ 自分が得をするから
□ 自分が損をしないから
□ 投稿をした人のフォロー数を見て

□ 投稿をした人のフォロワー数を見て
□ 投稿をした人のプロフィールを見て
□ 投稿のシェア数を見て
□ 投稿のお気に入り数を見て
□ その他
（決められない場合）
□ 投稿に対する反応を見て
□ 投稿をした人の詳細な情報がみたい
□ 投稿をした人の他の投稿がみたい
□ 投稿に対する他の人の反応がみたい
□ 投稿をした人について Web で調べたい
□ 迷っている
□ その他
（シェアを行わない場合）
□ 投稿に対する反応を見て
□ 信頼できないから
□ 知らない人の投稿だから
□ 自分が得をしないから
□ 自分が損をするから
□ 投稿をした人のプロフィールを見て
□ 投稿をした人のフォロー数を見て
□ 投稿をした人のフォロワー数を見て
□ 投稿のシェア数を見て
□ 投稿のお気に入り数を見て
□ 興味がない投稿だから
□ スパムの可能性があるから
□ 投稿をした人に自分を知られたくないから
□ その他

質問 5.12
投稿に対して何かコメントがあれば記入してください（自由記述）

事前実験の時と同様に，選択した回答の内容が他のシチュエーションを見たことに依存する可能性があるため，実験参加者には「対象となる類の投稿を初めて見た時」を想定した回答を依頼した．

5.4.5 実験終了後の最終アンケート項目

すべてシチュエーションに対して回答を終えたユーザに対してアンケートを実施した．質問の項目に関して，質問 5.13〜質問 5.17 として以下に示す．

質問 5.13

投稿に対する反応が表示されることで，投稿の印象が変わりましたか？

- ○ かなり変わった
- ○ 変わった
- ○ 少し変わった
- ○ あまり変わらなかった
- ○ 変わらなかった

質問 5.14

ネガティブなコメントを見て，投稿をシェアしようと思わなくなりましたか？

- ○ かなりシェアする気がなくなった
- ○ シェアする気がなくなった
- ○ 少しシェアする気がなくなった
- ○ あまり変わらなかった
- ○ 変わらなかった
- ○ もとからシェアする気はなかった

質問 5.15

今回の実験で取り扱った投稿はスパム（詐欺）でした．実験を行う前，この事を知っていましたか？

- ○ 知っていた
- ○ 少しだけ知っていた
- ○ 聞いたことがある
- ○ 知らなかった

質問 5.16

このような詐欺投稿を拡散(シェア)させないためにはどうすればいいと思いますか？考えがあれば記入してください（自由記述）

質問 5.17

その他，本実験に関して何かコメントがあれば記入してください（自由記述）

5.4.6 実験結果

実験参加者 35 名に対して, 6 つのシチュエーションに対する質問の回答と実験終了後の最終アンケートを完了させた. 各質問に対する得られた回答を, 表 5.12〜表 5.21 にまとめて以下に示す.

質問 5.10

あなたはこの投稿を

表 5.12　質問 5.10 に対する回答

＼シチュエーション番号	9	10	11	12	13	14
シェアする	3	1	0	3	0	1
シェアしてもよい	3	0	1	7	2	2
決められない	2	1	2	2	3	3
シェアはあまりしたくない	6	3	3	6	5	4
シェアしない	21	30	29	17	25	25

※単位は[人]である

質問 5.6

その理由は？（複数選択可）

表 5.13　質問 5.11 に対する，シェアを行う場合における回答

＼シチュエーション番号	9	10	11	12	13	14
投稿に対する反応を見て	3	0	1	2	0	2
知人がシェアしたものだから	2	0	0	2	1	1
信頼している人がシェアしたものだから	0	0	0	4	2	2
興味がある投稿だから	1	0	0	0	0	0
自分が得をするから	0	0	0	0	0	0
自分が損をしないから	4	0	0	5	2	1
投稿をした人のフォロー数を見て	0	0	0	0	0	0
投稿をした人のフォロワー数を見て	1	0	0	4	0	0
投稿をした人のプロフィールを見て	0	0	0	1	0	0
投稿のシェア数を見て	0	0	0	1	0	0
投稿のお気に入り数を見て	0	0	0	0	0	0
その他	0	1	0	1	1	0

※単位は[人]である

表 5.14　質問 5.11 に対する，シェアを行うか決められない場合における回答

＼シチュエーション番号	9	10	11	12	13	14
投稿に対する反応を見て	0	1	1	1	2	2
投稿をした人の詳細な情報がみたい	1	0	0	2	2	3
投稿をした人の他の投稿がみたい	0	0	0	1	0	0
投稿に対する他の人の反応がみたい	0	0	1	1	2	1
投稿をした人について Web で調べたい	0	0	0	1	1	2
迷っている	1	1	1	1	0	1
その他	0	1	1	0	1	0

※単位は[人]である

表 5.15　質問 5.11 に対する，シェアを行わない場合における回答

＼シチュエーション番号	9	10	11	12	13	14
投稿に対する反応を見て	2	24	20	1	19	16
信頼できないから	15	21	16	14	17	17
知らない人の投稿だから	12	7	9	7	7	5
自分が得をしないから	5	3	1	2	2	4
自分が損をするから	0	1	0	0	2	2
投稿をした人のプロフィールを見て	2	1	1	1	0	1
投稿をした人のフォロー数を見て	1	1	1	0	0	0
投稿をした人のフォロワー数を見て	1	1	1	0	0	0
投稿のシェア数を見て	3	4	3	1	1	1
投稿のお気に入り数を見て	3	2	3	1	1	1
興味がない投稿だから	12	9	11	12	10	10
スパムの可能性があるから	5	8	5	4	8	5
投稿をした人に自分を知られたくない	1	1	1	1	1	2
その他	2	0	1	1	0	0

※単位は[人]である

質問 5.12

投稿に対して何かコメントがあれば記入してください（自由記述）

表 5.16　質問 5.12 に対する回答（一部抜粋）

シチュエーション番号	回答内容
9	リプライ(反応)が数個あると信じたくなる
10	返信した人のツイート(投稿)を見たい．それを見て判断する
10	シンプルに詐欺の可能性が少しでもあるものを RT(シェア)したいとは思わない．
11	どっちのリプ(反応)が多いかによって決める
11	誰か一人でも(投稿が詐欺であると)言っていたらしない
12	実際に返信している人がいるから
13	もう少し時間を置いてからもう一度見て決めたい
13	友達と繋がっているので，この様なツイート(投稿)をリツイート(シェア)しているのを見られてどう思われるのかと考えてしまう
14	詐欺の疑いがあるツイートはわざわざ RT(シェア)したいとは思えない

質問 5.13

投稿に対する反応が表示されることで，投稿の印象が変わりましたか？

表 5.17　質問 5.13 に対する回答

項目	人数[人]
かなり変わった	12
変わった	9
少し変わった	9
あまり変わらなかった	1
変わらなかった	4

質問 5.14

ネガティブなコメントを見て，投稿をシェアしようと思わなくなりましたか？

表 5.18　質問 5.14 に対する回答

項目	人数[人]
かなりシェアする気がなくなった	5
シェアする気がなくなった	9
少しシェアする気がなくなった	5
あまり変わらなかった	2
変わらなかった	0
もとからシェアする気はなかった	14

質問 5.15

今回の実験で取り扱った投稿はスパム（詐欺）でした．実験を行う前，この事を知っていましたか？

表 5.19　質問 5.15 に対する回答

項目	人数[人]
知っていた	14
少しだけ知っていた	7
聞いたことがある	4
知らなかった	10

質問 5.16

このような詐欺投稿を拡散(シェア)させないためにはどうすればいいと思いますか？考えがあれば記入してください（自由記述）

表 5.20　質問 5.16 に対する回答（一部抜粋）

回答内容
そういったことをもっと色んな人に知って貰う．そのために Twitter などで発信していく
学校で教えて，そういった知識がある人を増やす
Twitter 本社がツイートに危険ゲージを設けて，ユーザーがこのツイートは危険だと思った場合は危険ゲージボタンをタップし，タップされた回数を数値化して不特定多数に見せる事で，この様な詐欺を減らす事が出来ると思います
完全個人情報アリの SNS などがあるのであれば，このようなスパムはなくなると思う
運営が対策をする(AI による排除など)

質問 5.17

その他，本実験に関して何かコメントがあれば記入してください（自由記述）

表 5.21　質問 5.17 に対する回答（一部抜粋）

回答内容
RT(シェア)数，いいね(お気に入り)数, FF(フォロー，フォロワー)数が多ければ多いほど RT してみようかなという気持ちになりました
普段スパムツイートが流れてきても無視しているので改めて見ると勉強になりました
普段使ってる twitter で実際に見かけそうなツイート(投稿)だったのでいろんな情報が関係して RT(シェア)しているんだなと思ったし，気軽に RT できてしまうからこそ気をつけないといけないと思いました
今回の実験を通して自分の RT(シェア)の基準について見直すきっかけになりました
もしリプライ(反応)の中に「実際に貰えましたありがとうございました！」などというものがあったら RT(シェア)してしまう人も多いと思います．第三者の意見がとても影響を与えると思いました
リプライを送った人のプロフィールやフォロー，フォロワー数も見たかった
リプライに「詐欺です」などの文があるだけで RT(シェア)しようと思わなくなりました

第 6 章　議論と考察

この章では第 5 章で得られた実験結果の分析を行い，提案手法の評価を行い議論する．また，実験結果より得られた新たな知見に関して述べる．

6.1 提案手法の評価

6 つのシチュエーションそれぞれに対してシェアするか，しないかを示した結果が表 5.12 である．この結果と，表 5.4 に示した元のシチュエーションにおける結果との比較を行う．この設問では回答を 5 段階としているが，回答候補の「シェアする」と「シェアしてもよい」は「シェアを行う」に，「シェアはあまりしたくない」と「シェアしない」は「シェアを行わない」にまとめることができる．このようにして結果を 3 段階に簡略化したものを，同一の攻撃投稿内容のシチュエーションと合わせてそれぞれ図 6.1，図 6.2 に示す．

シチュエーション 1 は事前実験で取り扱ったシチュエーションの中で最もシェアしないと回答したユーザが多かったものである．表 5.4 より，提案手法を取り入れていないときにシェアすると回答したユーザは 5 人であった．図 6.1 を見ると，投稿に対してシェアすると回答した人数は，ポジティブな反応のみを提示した場合に 6 人，ネガティブな反応のみを提示した場合に 1 人，そして両方の反応を提示した場合に 1 人となった．そのシェア比率はそれぞれ, 0.14, 0.17, 0.03, 0.03 である．

シチュエーション 4 は事前実験で取り扱ったシチュエーションの中で最もシェアすると回答したユーザが多かったものである．表 5.4 より，提案手法を取り入れていないときにシェアすると回答したユーザは 12 人であった．図 6.2 を見ると，投稿に対してシェアすると回答した人数は，ポジティブな反応のみを提示した場合に 10 人，ネガティブな反応のみを提示した場合に 2 人，そして両方の反応を提示した場合に 3 人となった．そのシェア比率はそれぞれ, 0.34, 0.29, 0.06, 0.09 である．

ポジティブな反応のみを提示した場合，シェアを行うと回答したユーザは何も提示しない場合と比較して大きな変化は見られなかった．一方で，ネガティブな反応のみを提示した場合とポジティブとネガティブの反応両方を提示した場合では，シェアを行うと回答したユーザは何も提示しない場合と比較して大きな減少が見られた．また，表 14 よりネガティブな反応が提示されることでシェアする気が起きなくなったと回答したユーザは 19 人であった．以上の事から，ポジティブな反応の提示はユーザのシェアしようとする気に影響を与えず，ネガティブな反応の提示はユーザのシェアしようとする気を大きく下げることがわかる．

以上より，アプリケーションレベルによるポジティブ，ネガティブな反応の提示は，ソーシャルエンジアリング攻撃の拡散防止に有効であると主張できる．

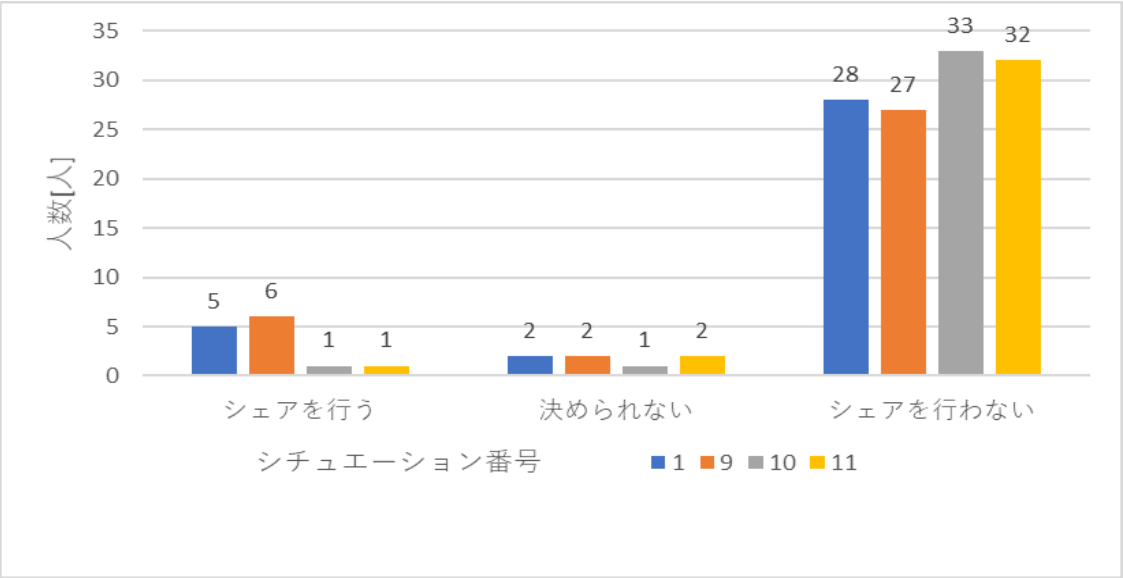

図 6.1　シチュエーション 1,9,10,11 に対して 3 段階に変換した回答

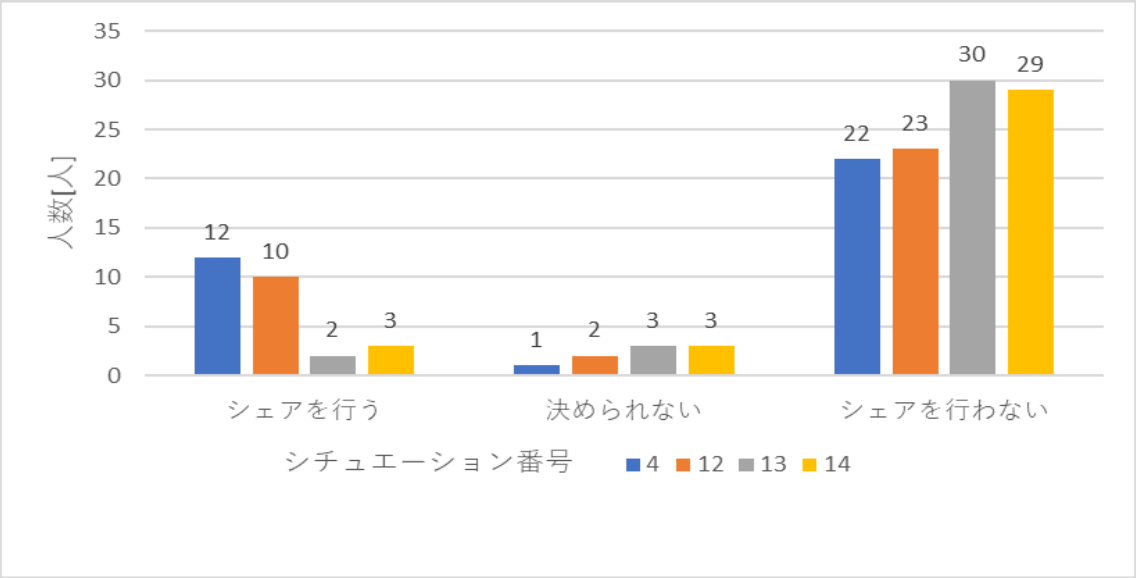

図 6.2　シチュエーション 4,12,13,14 に対して 3 段階に変換した回答

6.2 SNS としての質の担保

ネガティブな反応の提示が攻撃投稿の拡散を防止することが判明した一方で，ソーシャルエンジアリング以外の投稿に対してもシェアしようと思わなくなり,アクティブ数の低下をもたらし SNS としてのサービスの質が損なわれる可能性がある．この問題は，ポジティブな反応を同時に提示することである程度解消することができる．

表 14 よりネガティブな反応が提示されることでシェアする気が起きなくなったと回答したユーザは 19 人であったが，表 17 をみると，投稿に対する反応が提示されたことで投稿の印象が変わったと回答したユーザは 30 人であった．これは，反応の提示がシェアしようと思うかどうかに影響するだけでないことを示している．つまり，ポジティブな反応を提示することで，投稿に対してユーザが受け取る印象は変わるということである．実際に表 5.16 を見ると、ポジティブな反応のみが提示された攻撃投稿をシェアすると回答したあるユーザは，次のように述べている：
リプライ(反応)が数個あると信じたくなる
これは，ポジティブな反応が提示されたことによって投稿に対する信憑性が上がったということである．

しかしながら，ポジティブな反応を提示することによる攻撃投稿以外の投稿におけるシェアへの影響は，本実験で計測することはできない．評価実験で取り扱った投稿がすべてソーシャルエンジアリング攻撃に繋がる投稿であったためである．SNS としての質の確保が同時に成立するか調査するために，攻撃以外の投稿に対してもシェア時の情報提示を行い，シェアするかしないかに関する実験を実施する必要がある．

6.3 ソーシャルエンジニアリングに関する知識の影響

表 5.18 を見ると，ソーシャルエンジニアリングに関する攻撃投稿を「もとからシェアする気はなかった」と回答したユーザが 14 人であった．また，実験参加者のソーシャルエンジニアリングに関する事前知識も調査しており，表 5.19 を見ると今回実験で用いた投稿が悪質なものであることを「知っていた」と回答したユーザが 14 人であった．このことから，ソーシャルエンジニアリングに関して持つ知識がシェアを行うか行わないかの判断基準となることが予測できる．本節では得られたデータを基に，ソーシャルエンジニアリング攻撃に関する知識とシェアの判断基準の相関性を調査する．

質問 5.15 では，実験参加者に対してソーシャルエンジニアリングに関する知識の有無を 4 段階で質問しており，それに対する回答が表 5.19 で示されている．この回答に応じて実験参加者を 4 つのグループに分類し，回答データの比較を行う．

ここで，比較のためにシェア比率を用いる．シチュエーション 1 から 14 のすべてのシチュエーションにおいて，それぞれのグループがシェアを行うか行わないか 5 段階で回答し

た比率を調査し，グループ同士で比較したグラフを図 6.3 に示す．

この図を見ると，攻撃投稿を「シェアする」と回答した比率が最も多かったグループは攻撃投稿であることを「知っていた」グループであった．また，攻撃投稿を「知っていた」グループ以外の回答比率を見ると，攻撃投稿であることを認知していないほどシェア比率は増えている．これを説明するために，シェアを行った要因に着目する．攻撃投稿であることを「知っていた」グループ内のユーザで，シェアを行うと回答したあるユーザは次のように述べている：

貰えるはずがないが，僅かな可能性にかけてみたいから

また，同グループでシェアを行うと回答した別のユーザは次のように述べている：

知人との話題のネタになるから

以上のことから，攻撃投稿であることを認知した上でシェアを行うユーザもいるということが分かった．これは，攻撃投稿をシェアした時点においてはシェアをしたユーザが被害を直接受けるわけではないためであると言える．

一般的に，攻撃投稿であることを認知していればいるほどソーシャルエンジニアリングの被害に遭う可能性は少なくなる．しかし，攻撃投稿をシェアすることに関して言えば，攻撃投稿であることを認知しているかどうかは関係なく，各々がシェアを行う判断基準に委ねられる．

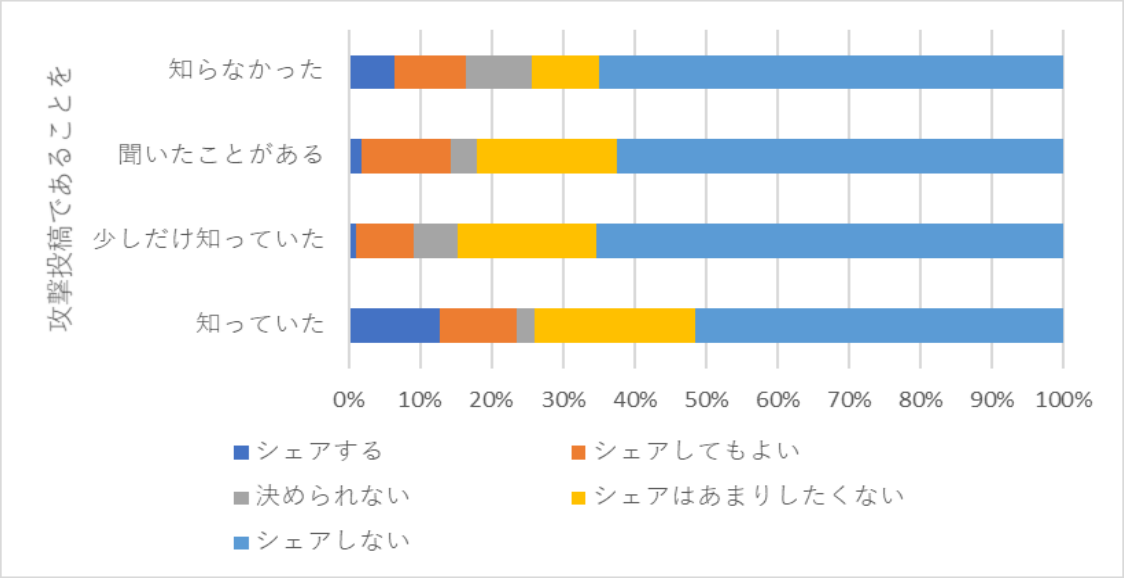

図 6.3　各グループにおけるシェアを行うかどうかの回答比率

6.4 新しいソーシャルエンジニアリングの対策手法

実験結果から導くことの出来るソーシャルエンジニアリング攻撃の対策方法に関して議論を行う．第 2 章でも述べたように，ソーシャルエンジニアリング攻撃を対策する手法として 2 つの分類に分けられることが多い．それが

1. ソーシャルエンジニアリングを検知し，それを排除する
2. ソーシャルエンジニアリングに関する知識を付ける

である [8]．実験参加者に対してソーシャルエンジニアリングを減らすための対策方法として何が考えられるか質問した回答を表 20 より見てみると，概ねこの 2 種類に当てはまることが分かる．例えば

- *学校で教えて，そういった知識がある人を増やす*
- *そういったことをもっと色んな人に知って貰う．そのために Twitter などで発信していく*

これらはソーシャルエンジニアリングの知識をつけるための方法である．また，

- *Twitter 本社がツイートに危険ゲージを設けて，ユーザーがこのツイートは危険だと思った場合は危険ゲージボタンをタップし，タップされた回数を数値化して不特定多数に見せる事で，この様な詐欺を減らす事が出来ると思います*
- *運営が対策をする (AI による排除など)*

これらはソーシャルエンジニアリングを検知し，排除する方法である．

ここで，本研究で述べた提案手法がどちらにも属さないことが分かる．そのため，3 つ目のソーシャルエンジニアリングの対策手法を次のように定義する：

3. ユーザに対して思考を誘発させる

本研究では，投稿のシェア時において投稿に対するポジティブな反応とネガティブな反応を同時に提示した．これがもたらした効果として，ユーザに新たな情報を提示したことにより思考を誘発させることができたため，攻撃投稿の拡散を防ぐことができたと考えることができる．提案手法を用いた実験において表 5.16 より，反応が提示された攻撃投稿に対するコメントとしてあるユーザは次のように述べている：

返信した人のツイート (投稿) を見たい．それを見て判断する

これは，新たな情報が提示されたことにより投稿に対する疑問が生じ，思考を誘発させることができたために得られた回答であると言える．また，別のコメントとしてネガティブな反応が一緒に提示された投稿に対してあるユーザは次のように述べている：

友達と繋がっているので，この様なツイート (投稿) をリツイート (シェア) しているのを見られてどう思われるのかと考えてしまう

これも，新たな情報が提示されたことにより思考を誘発させることができたために得られた回答である．

本実験ではポジティブの反応とネガティブの反応という新たな情報を投稿と一緒に提示したことで，思考を誘発させたことによりソーシャルエンジニアリングに繋がる攻撃投稿

の拡散を防ぐことができることを示した．これはつまり，ユーザは投稿に対してシェアを行うかどうかの基準は持っているが，そこに思考の余地が介入していないということでもある．ここでいう思考とは，自信が経験したことの無い事象に対してどう処理を行うか決定するまでのプロセスのことを指す．

ユーザは投稿に対して，すでに頭の中にある基準に従ってシェアを行うか行わないかを決定する．ソーシャルエンジニアリング攻撃では，その基準を悪用することで機密情報を奪い取るため，思考を誘発させシェアの基準を更新し続けるようなアプローチは有効である．その観点から本研究は，アプリケーションに特定の機能を実装することで思考を誘発させ，ソーシャルエンジニアリング攻撃の対策することができるという研究の一例となった．

第 7 章　将来課題

第 6 章で述べた考察を踏まえたうえで，本研究に関する将来課題について論じる．

まず，提案手法が適用できるソーシャルエンジニアリングは，投稿の拡散を促して後から直接やり取りを行い，攻撃を仕掛ける種類の攻撃のみである．実際には，SNS におけるソーシャルエンジニアリングは個人と直接メッセージのやり取りを行い，機密情報を手に入れるような標的型攻撃など多岐にわたるため，それぞれに共通して適用できるような対策について考案する必要がある．

次に，提案手法で評価した投稿はソーシャルエンジニアリング攻撃に繋がる投稿のみであった．6.2 節でも述べたように，提案手法を採用したときに SNS としての質が担保され続ける状態でなければならない．そのため，提案手法が他の一般的な投稿に対してどのように左右するのか調査する必要がある．

次に 6.3 節より，攻撃投稿のシェアを行うかどうかはソーシャルエンジニアリングに関する知識に影響しないことが分かった．要するに，攻撃投稿であることを分かったうえでシェアを行う場合がある．これに関して，ソーシャルエンジニアリングの拡散を防ぐためにも，認知したうえで拡散されていくケースに対する心理，要因を詳細に調査する必要があり，それに応じた対策についても論じる必要がある．

また本研究における提案手法は，ユーザがその手法で提示された情報に慣れた時に通用しなくなる可能性がある．そのため，情報提示の機能に慣れた場合においてシェアを行うかどうか，新たに調査する必要がある．もしくは 6.4 節でも述べたように，常に思考を誘発させ続けるようなアプローチが有効であるため，提案手法の機能に慣れた後であっても思考を誘発させ続けられるような設計を考える必要がある．

最後に，本研究は，アプリケーションの特定の機能によって使用するユーザの思考を誘発させることで，ソーシャルエンジニアリングが対策できることを示した最初の研究である．思考の誘発によりソーシャルエンジアリングが対策できるという事をより信憑性のある主張とするために，提案手法とは異なる手法でユーザの思考を誘発することで，ソーシャルエンジニアリング攻撃が対策できるといった事例を作る必要がある．

第 8 章　結論

本研究では, SNS におけるソーシャルエンジニアリング攻撃の拡散を対策するために，投稿のシェアを行う直前のタイミングでポジティブな反応とネガティブな反応を提示するような手法を提案した．そして，ポジティブな反応はシェアの要因とならないが投稿に対する印象を変え，ネガティブな反応はシェアの要因と密接に結びつき攻撃投稿の拡散を防ぐことができるということを示した．また新たな知見として，ソーシャルエンジニアリングに関する知識の有無は，攻撃投稿をシェアするかしないかの判断に影響を与えないことも示した．

また本研究ではソーシャルエンジニアリング攻撃を対策するための手法として，新たな括りを定義した：「ユーザに対する思考の誘発」 である．新たな情報の提示により，新しい体験をユーザに与え思考を誘発させることで，ソーシャルエンジニアリング攻撃を対策することができる可能性がある．これを立証するために，本研究における提案手法とは異なる手法を用いて，ユーザに思考を誘発させることでソーシャルエンジニアリング攻撃が対策できるといった事例を作る必要がある．

参考文献

[1] Berg, Al, “Cracking a social engineer: enterprising thieves use a variety of common techniques to pilfer information”, LAN Times, 1995.

[2] TREND MICRO, “2019 年 セキュリティ脅威予測”, available on: https://www.trendmicro.com/ja_jp/business.html, 2018.

[3] Akshat Jain, Harshita Tailang, Harsh Goswami, Soumiya Dutta, Mahipal Singh Sankhla, and Rajeev Kumar, “Social Engineering: Hacking a Human Being through Technology”, IOSR Journal of Computer Engineering, Vol.18, Issue 5, 2016.

[4] Markus Huber, Martin Mulazzani, Edgar Weippl, Gerhard Kitzler, and Sigrun Goluch, “Friend-in-the-Middle Attacks: Exploiting Social Networking Sites for Spam”, IEEE Internet Computing, Volume 15, Issue 3, 2011.

[5] Edwin D. Frauenstein, and Stephen V. Flowerday, “Social network phishing: Becoming habituated to clicks and ignorant to threats?”, Information Security for South Africa, 2016.

[6] 日本経済新聞, “コインチェックの仮想通貨不正流出、過去最大 580 億円”, avaiable on: https://www.nikkei.com/article/DGXMZO26231090X20C18A1MM8000/, 2018 (Retrieved July 2019).

[7] So-net, “Twitter で「当選詐欺」横行――「賞品の送料は当選者負担」とクレカ情報を要求”, available on: https://securitynews.so-net.ne.jp/news/sec_00024.html, 2019 (Retrieved July 2019).

[8] Fatima Salahdine, and Naima Kaabouch, “Social Engineering Attacks: A Survey”, MDPI, 2019.

[9] Ira S. Winkler, and Brian Dealy, “Information Security Technology?...Don’t Rely on It: A Case Study in Social Engineering”, USENIX UNIX Security Symposium, 1995.

[10] Tolga Mataracioglu, and Sevgi Ozkan, “User Awareness Measurement Through Social Engineering”, arXiv:1108.2149, 2011.

[11] Harl, “People Hacking: The Psychology of Social Engineering”, Talk at Access All Areas III Conference, 1997.

[12] Jonathan J. Rusch, “The "Social Engineering" of Internet Fraud”, INET Conference, San Jose, 1999.

[13] Mironela Pirnau, “Considerations on Preventing Social Engineering over the Internet”, Memoirs of the Scientific Sections of the Romanian Academy, 2017.

[14] Jagatic T, Johnson N, Jakobsson M, and Menczer F, “Social phishing”, Communications of the ACM, 2007.

[15] Brandon A, and Wilson Huang, “A Study of Social Engineering in Online Frauds”, Open Journal of Social Sciences, Vol.1, No.3, 2013.

[16] Jussi-Pekka Erkkila, “Why we fall for Phishing”, CHI, 2011.

[17] Danah M. Boyd, and Nicole B. Ellison, “Social Network Sites: Definition, History, and Scholarship”, Journal of Computer-Mediated Communication, 2008.

[18] M. Nick Hajli, “The role of social support on relationshipquality and social commerce”, Technological Forecasting andSocial Change, Volume 87, 2014.

[19] Nick Hajli, and Xiaolin Lin, “Exploring the Security of Information Sharing on Social Networking Sites: The Role of Perceived Control of Information”, Journal of Business Ethics, 2014.

[20] Jan Nagy, and Peter Pecho, “Social Networks Security”, Third International Conference on Emerging Security Information, Systems and Technologies, 2009.

[21] Ralph Gross, and Alessandro Acquisti, “Information Revelation and Privacy in Online Social Networks (The Facebook case)”, Proceedings of the 2005 ACM workshop on Privacy in the electronic society, 2005.

[22] Alex Hai Wang, “Don't follow me: Spam detection in Twitter”, International Conference on Security and Cryptography, 2010.

[23] Sarita Yardi, Daniel M. Romero, Grant Schoenebeck, and Danah Boyd, “Detecting spam in a twitter network”, First Monday, Volume 15, 2010.

[24] Thomas R. Peltier, “Social Engineering - Concepts and Solutions”, The EDP Audit, Control, and Security Newsletter, Volume 33, 2006.

[25] Twitter, https://twitter.com

[26] Facebook, https://www.facebook.com/

[27] Jumin Lee, Do-Hyung Park, and Ingoo Han, “The effect of negative online consumer reviews on product attitude: An information processing view”, Electronic Commerce Research and Applications, Volume 7, Issue 3, 2008.

[28] Pete Cashmore, “Should Facebook add a dislike button?”, avaiable on: http://edition.cnn.com/2010/TECH/social.media/07/22/facebook.dislike.cashmore/index.html, 2010 (Retrieved July 2019).

[29] Joshua Hardwick, “Find Out How Much Traffic a Website Gets: 3 Ways Compared”, Available on: https://ahrefs.com/blog/website-traffic/, 2018 (Retrieved July 2019).

[30] Peter D. Turney, and Michael L. Littman, “Measuring praise and criticism: Inference of semantic orientation from association”, Transactions on Information Systems, Volume 21, Issue 4, 2003.

[31] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede, “Lexicon-Based Methods for Sentiment Analysis”, Computational Linguistics, Volume 37, Issue 2, 2011.

[32] 労働政策研究・研修機構, “インターネット調査は社会調査に利用できるか”, 労働政策研究報告書, No17, 2005.

[33] 小川隆一, 安藤玲未, 島成佳, 竹村敏彦, “SNS における情報開示行動に関する要因分析”, 情報処理学会論文誌, Vol.58, No.12, 2017.

謝辞

本研究を行うにあたって，丁寧にご指導いただきました中島達夫教授に対して，この場を設けて感謝の意を表します．また，本研究の実験に参加していただいた SNS におけるフォロワーの方々，並びに研究に対する議論をしていただけた同研究室のみなさまに対して，心より感謝申し上げます．