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

狩野 佑記 [1,*]　中島 達夫 [1]

概要：Twitter や Facebook などのソーシャルネットワークサービス（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．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提示を行う．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ことが判明した．

キーワード：ソーシャルエンジニアリング，ソーシャルネットワーク，アプリデザイン，拡散，信頼性

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

Yuki Kano[1,*]　Tatsuo Nakajima[1]

Abstract: The emergence of Social Networking Services (SNS), such as Twitter and Facebook, has created a new social engineering threat that emphasizes user communication. 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 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 our approach, 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.

Keywords: Social Engineering, Social Network, Application design, Spreading, credibility

1. はじめに

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ソーシャルエンジニアリング攻撃の被害が増えている．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[2].

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．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など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を行う．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．

本論文の構成は以下の通りである．第 2 章ではソーシャルエンジニアリングについて行われた既存の研究について述べる．第 3 章では本研究における提案手法，およびその条件について述べ，第 4 章では提案手法を評価するための

1 早稲田大学基幹理工学研究科
Fundamental Science and Engineering, Waseda University
* konayuki@dcl.cs.waseda.ac.jp

実験を行うために用いたアプリケーションについて述べる．第 5 章では評価実験について概要，目的，方法，そして実験結果を示し，第 6 章では第 5 章で得られた結果より議論，考察を行う．最後に，第 7 章では本研究の将来課題および結論について述べる．

2. 関連研究

本論文のキーワード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．文献[1]ではソーシャルエンジニアリングを“システムに侵入するのではなく，システムの進入に必要な情報（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．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などが存在することを指摘した．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．

ソーシャルエンジニアリングは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[3]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．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に，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．”

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攻撃プラットフォームはメーリングサービスなどから SNS へと動きを広げている．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．文献[4]では 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”と定義している．SNS の普及に伴い，個人のプロフィールなどから得られる莫大な数の個人情報が公開され，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．Jan らは文献[5]において，SNS で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．SNS におけるソーシャルエンジニアリングの例として，文献[6]では Twitter におけるスパムアカウントを報告している．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．また，Twitter では 3%以上ものメッセージがスパムであることも報告している．

最後に，上述したソーシャルエンジニアリング攻撃への対策手法について論じた研究について述べる．Fatima らは文献[7]において，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．その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，そして攻撃の検出ツールの利用などが挙げられている．また同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．これに関して文献[8]で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に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．文献[9]で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」，そして「説得・影響の技術」に深く関わることを述べている．人間心理が根本の原因であるため，技術的な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要であることを述べている．

3. 提案手法

3.1 取り扱うソーシャルエンジニアリング攻撃

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ついて攻撃が実行されるプラットフォーム，そして対策を実施させる対象を明確にする．

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．ここで述べたオープンな 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目視できる SNS である．サービス提供者は SNS アプリケーションに特定の機能を実装することで攻撃の対策を実施し, それを利用するユーザの傾向の変化を観測する．

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「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，その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[10]．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で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．

3.2 提案手法：シェアの重みづけ

上述し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，本研究ではシェアの重みづけを提案する．具体的に，アプリケーションレベルで投稿をシェアする時に追加の情報をユーザに提示することで，シェアを行う事に対して重みづけを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[11][12].

本研究における提案手法では，シェアを確定させる二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を持つ文章である反応を提示する．第 2 章で述べたように，意味を持つ文章の提示はユーザの判断を大きく変えうる．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，意味を持つ文章による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を持つ文章である反応を 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る．

ネガティブな文章である反応がある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[13]. 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[14]. 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 全体のアクティブ数の増加をもたらし，SNS としての質を担保できる．しかし，ポジティブな反応の提示は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ある． 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，投稿に対する意味を持つ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，ユーザは意味を持つ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できる．提案手法が有効かどうか調査するために実施した評価実験については，第 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．提案手法では，攻撃投稿に対する反応が既に付いていることが前提条件である．要するに，攻撃投稿に対してある程度の拡散が生じた状態から新たに拡散をさせないためのアプローチとなる．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拡散の危険性が低いとみなし対象から外した．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[15]．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「拡散されている状態」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」であると定義する．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．

また提案手法では，投稿に対する反応の極性をすでに判別されているものとして取り扱う．要するに，投稿に対する反応があるときにそれを「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．反応を振り分ける手法に関しては次に述べておく．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．文献[16]では単語の感情極性を判定するために，すでに極性が判別され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．また文献[17]で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，複数のプラット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．これらの技術などを用いることで，ある文章が「ポジティブ」な意味を持つのか「ネガティブ」な意味を持つのか判別することができる．

4. 実験アプリケーションの設計と実装

4.1 実験アプリケーションの概要

提案手法に対す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アプリケーションを利用する．仮想 SNS の基となる API として，Twitter API 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．シェア比率は，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，

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

と定義する．この時アプリケーション側で，投稿に対する意味を持つ文章によ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方で実験を実施し，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 実験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．本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，

(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．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．設計したサーバの概略図を図 2　サーバの概略図に示す．

(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 にアクセスし利用する．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から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 を可能な限り近づけ再現する．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に対する慣れを必要なく実験を行える作りとした．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**エラー! 参照元が見つかりません。**に示す．

5. 評価実験

投稿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拡散の対策として効果的であるかどうかを調査するために，第 4 章で述べた仮想 SNS アプリケーションを用いて評価実験を実施した．

5.1 実験方法

実験の参加者を募り，作成した仮想 SNS アプリケーションを実験参加者に利用させ，様々な攻撃シチュエーションに対する質問の回答を記録することで実験を行った．最初に，本研究における実験参加者の対象について述べる．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 上で募り集められた．集められたユーザ全員に対し，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 17 名，女性 18 名）を対象に実験を行った．

ここで実験を行う前に，実験参加者グループとしての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[18]．今回の場合「SNS を普段から利用しているユーザ群」に対する実験結果の妥当性は主張できる．

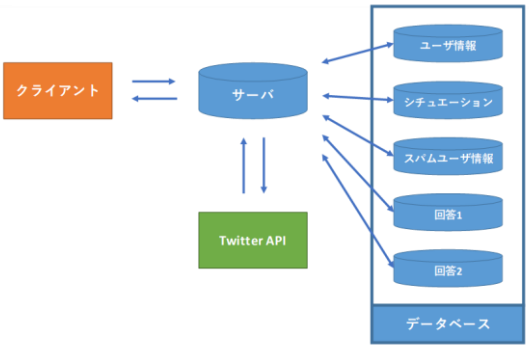

図 2　サーバの概略図
Figure 2　Server Overview.

図 1　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面
Figure 1　Screenshots of Client Timeline (Desktop and Mobile).

また本実験参加者のグループの傾向として，極めて特殊な傾向を持たないことを示すために（述べた条件のユーザ群に対して普遍的であることを示すために），既存の研究で示さ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．本研究における実験参加者がこの傾向と一致するか確かめるために，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．実施した事前実験について 5.2 節に，事前実験の結果を基に行う本実験について 5.3 節に述べる．

5.2　事前実験

5.2.1 調査内容

SNS における信頼性に関する既存の研究[19]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ている：

1. ユーザへの信頼
2. 情報開示範囲のコントロール
3. リスク認知
4. 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. 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．実験で取り扱うシチュエーションは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．この投稿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収集する．

5.2.2 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，信頼に関わる様々なパラメータを変えつつ異なる攻撃の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 1 に示す．

表 1　8 つのシチュエーションとそれぞれに対応するパラメータ

Table 1　8 situations and corresponding parameters.

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	×	×	×	×	×	×	×	○

5.2.3 事前実験の結果および考察

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を得た．回答内容は，シチュエーション毎に異なる攻撃投稿をシェアしようと思うか・思わないかである．得られた実験結果を図 3 に示す．

シチュエーション 1 と 2 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．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 3 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．ユーザがシェアを行う要因が「匿名・非匿名によるもの」には関係ないことが確認できた．シチュエーション 3 と 4 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．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．

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 5.2.1 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．つまり本実験参加者のグループの妥当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．

5.3　評価実験

5.3.1 調査内容

提案手法について改めて述べると，シェアを確定させる

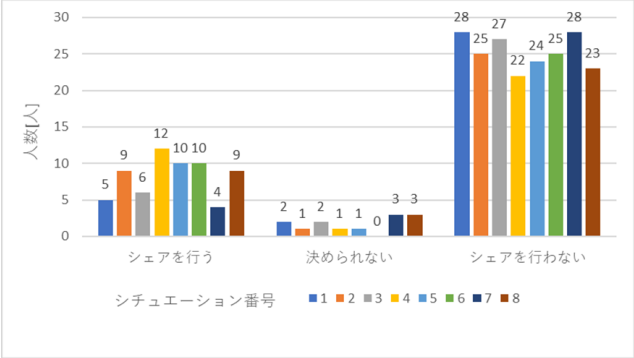

図 3　各シチュエーションに対して得た回答

Figure 3　Answers obtained for each situation.

二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．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．攻撃投稿に関して用意するシチュエーションは事前実験のものと同様である．得られた回答と 5.2.3 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．

5.3.2 提案手法を組み込んだシチュエーション

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

表 2　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
Table 2　Situation Number according to proposed method.

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	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	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両方	シチュエーション 11	シチュエーション 14

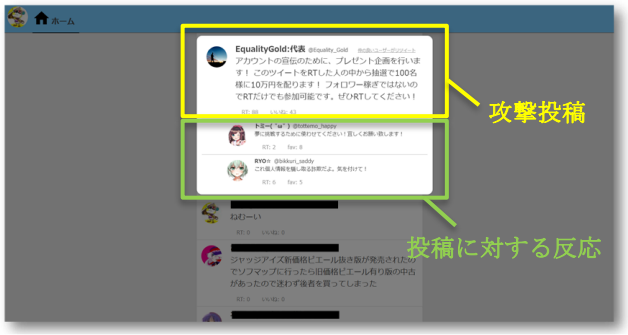

図 4　攻撃投稿とそれに対する反応が表示された画面
Figure 4　Screenshot showing attack post and their responses.

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もある．そのため，用いる攻撃シチュエーション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．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．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 2 に示した．また，攻撃投稿とそれに対する反応が表示された画面を図 4 に示す．

5.3.3 実験結果

実験参加者 35 名に対して，新たな 6 つのシチュエーションに対する質問の回答を得た．回答内容は事前実験と同様，シチュエーション毎に異なる攻撃投稿をシェアしようと思うか・思わないかである． 得られた実験結果および同一攻撃シチュエーションにおける回答の比較を表したものを図 5 に示す．

6.　議論と考察

6.1　提案手法の評価

得られた実験結果より提案手法の評価を行う．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．図 5（左）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．一方で，提案手法を取り入れた時に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 人であった．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. 図 5(右)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．一方で，提案手法を取り入れた時に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 人であった. 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. これはつまり，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．

6.2　ポジティブなリアクションによる投稿の印象変化

実験を終えたユーザに対してアンケートを実施し，ネガティブな反応が提示されることでシェアをする気が起きなくなったかどうか調査した結果，起きなくなったと回答したユーザは 19 人であった．また同時に，投稿に対する反応が提示されたことによって投稿に対する印象が変わったかどうかを調査した結果，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．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

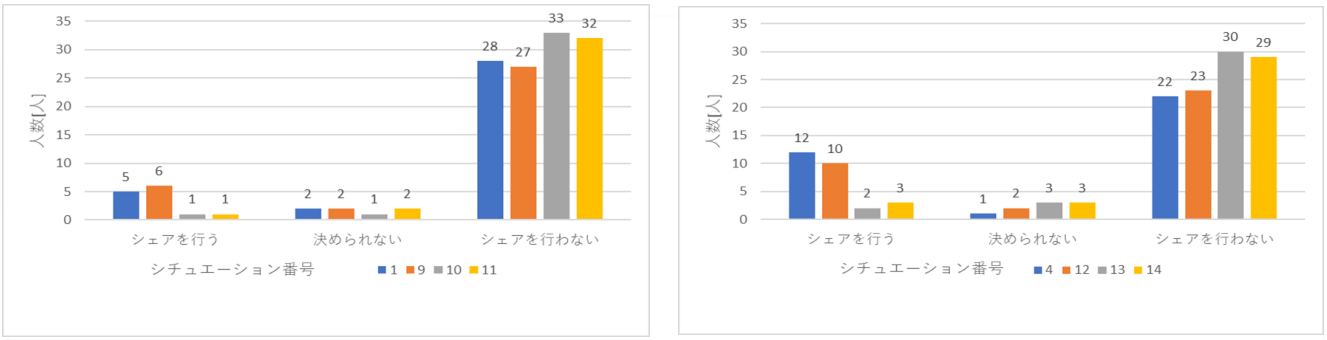

図 5　同一の攻撃シチュエーションにおけるシェアを行うかの回答数およびその比較

Figure 5　Number of responses and comparison of whether to share in the same attack situation.

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．そのため，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評価実験を実施する必要がある．

7. 結論と将来課題

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アプリ側から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ことを示した．

将来課題として，本研究における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ため，他の一般的な投稿に対してどのように影響を与えるのか調査する必要がある．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．

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] 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.

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

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

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

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

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

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

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

[10] “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”. https://securitynews.so-net.ne.jp/news/sec_00024.html, (参照 2019-07-20).

[11] “Twitter”. https://twitter.com, (参照 2019-07-20).

[12] “Facebook”. https://www.facebook.com/, (参照 2019-07-20).

[13] 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.

[14] “Should Facebook add a dislike button?”. http://edition.cnn.com/2010/TECH/social.media/07/22/facebook.dislike.cashmore/index.html, (参照 2019-07-20).

[15] “Find Out How Much Traffic a Website Gets: 3 Ways Compared”. https://ahrefs.com/blog/website-traffic/, (参照 2019-07-20).

[16] 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.

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

[18] 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.

[19] 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. Chang, C. L. and Lee, R. C. T.. Symbolic Logic and Mechanical Theorem Proving. Academic Press, 1973, 331p.