個人に合わせた巧妙な標的型メールの分析とその対策手法の研究

博士学位論文

個人に合わせた巧妙な標的型メールの分析と その対策手法の研究

２０２０年１２月

静 岡 大 学

大学院自然科学系教育部

情報科学専攻

西川弘毅

論文要旨

特定の企業・組織の機微情報や設備破壊による稼働防止，経済損失の発生を目的とする標的型攻撃はより深刻となっている．近年でも，日本国内の重要政府機関やインフラ企業が標的型攻撃の対象となり，機密情報や個人情報の漏洩など深刻な被害を受けており，対策が重要である．標的型攻撃では，標的に特化したメール（標的型メール）を利用して標的組織をマルウェアに感染させたのち，機密情報の窃取や，データの不正な暗号化によるシステム停止，といった攻撃者の目的を達成する活動を行う．その背景には，情報通信技術（ICT）の発展がある．マルウェアや不正な通信の検知技術，サーバの堅牢化など，サイバー攻撃防御技術の開発は進んでおり，攻撃者は，標的組織での目的を達成することが以前と比べて難しくなっている．そのため攻撃者は，セキュリティ上で最も弱い点(Weakest Link)である「人」を対象に攻撃を実施する．そして ICT の発展が，攻撃者が人に対する攻撃をより巧妙，かつ，より容易に行うことを可能としている．

攻撃は 1 点を突破すれば成功するのに対し，防御はすべての攻撃を防がなければならず，攻防はそもそも攻撃者有利の関係にある．その中でも「人」の価値観や振る舞いは非常に多様であり，その結果，人を対象とした攻撃も多岐に渡ることになる．このため，防御側がそのすべてに対応することは格段に難しく，標的型攻撃は攻高防低が特に顕著となる事案である．この課題に対し，現在の防御策は全ての人に対して一律の対策を実施するに留まっており，巧妙な攻撃を防ぎ切ることができていないという現状にある．本研究の最終目的は，今後，より深刻化していくことが予想される，人に対する脅威の可能性について明らかにし，人ごとに適した対策を取り入れていくことで，より効果的なセキュリティ対策を実現することである．その第一歩として，本論文では，攻撃者が標的を直接揺さぶることができるメールに着目して研究を行った．今後予想される標的型メールの脅威と対策を見据え，攻撃側が擬態精度あるいは心理操作効力の高いメールをどの程度作成することができるのかに関する検討と，その結果を防御側がどのように対策に活用していくことができるのかに関する検討を行った．

まず，擬態精度を高めたメールの脅威と対策を示した．具体的には，攻撃者は標的者の名前のみを手掛かりとして，インターネットに公開されている情報源から標的者に関する様々な情報を Open Source Intelligence（OSINT）ツールによって次々と取得し，これらの情報をメールに組み入れることによって，擬態精度が高いメールを標的者ごとに作成することが可能であることを示した．このような擬態精度を高める攻撃への対策として，攻撃者による OSINT の進行を状態遷移図によって定式化し，個々の攻撃段階（どこまでの情報が攻撃者に漏洩しているか）ごとに作成され得る標的型メール文面を整理した．各組織は，自らも OSINT を実行することによって攻撃者が自組織の情報をどこまで入手可能であるか確認し，本状態遷移図に照らし合わせることによって，自組織に届く可能性があ

る標的型メールの文面をあらかじめ把握することが可能となるため，それを踏まえたプロアクティブな対策選定に資することができる．

次に，心理操作効力を高めたメールの脅威と対策を示した．攻撃者は，AI ツールによって推定した標的者の特性に基づいて，その標的者に有効なチャルディーニの法則を推定することで，心理操作効力の高い標的型メールを標的者ごとに作成することができる．本研究では，100 人規模のアンケート調査を実施し，個人の性格因子とチャルディーニの法則の相関関係が，行動特性によって異なることを示した．これまで，種々の既存研究において説得のされやすさ（チャルディーニの法則に対する感受性）と性格因子との相関関係，あるいは，説得のされやすさ（詐欺師の説得を受け入れてしまいやすさ）と行動特性との相関関係がそれぞれ個別に調査されてきたのに対し，説得のされやすさ（チャルディーニの法則に対する感受性）を性格因子と行動特性のコンビネーションによって分析したことが，本研究の第一の貢献である．このような心理操作効力を高める攻撃への対策として，標的者が標的型メールを受けた際には，メールに含まれている説得のフレーズを通知することで標的者に注意を促すことが有効であると考えられる．本研究では，400 人規模のアンケート調査を実施し，効果的な注意喚起の方法（メール内のどの説得フレーズに対してアラートを提示すると，標的者は標的型メールへの不信感を高めるか）も，標的者の性格因子と行動特性の両者によって異なることを示した．

これまでの既存研究においては，標的型メールの脅威分析（攻撃側が擬態精度あるいは心理操作効力の高いメールをどの程度作成することができるのか）に主眼が置かれていた．これに対し本研究では，標的型メールへの対策（防御側が脅威分析の結果をどのように対策に活用していくことができるのか）にまで足を踏み込んだ．攻撃者による脅威をあらかじめ把握することで，標的者に応じた有効な対策を配備することができることを示したことが，本研究の第二の貢献である．本研究は，著者が知る限り，個々人の特性に応じてセキュリティシステムの防御方法を変える考えを具体化した初めての研究である．本研究が，Weakest Link である人に着目したサイバーセキュリティ防御研究の今後の足掛かりとなることを期待する．

目次

図一覧

表一覧

第1章　　序論

1.1. 本研究の背景と目的

特定の企業・組織の機微情報や設備破壊による稼働防止や経済損失の発生を目的とする標的型攻撃はより深刻となっている[1][2]. 警察庁によれば,「新型コロナウイルス感染症の発生に乗じたものを含め，サイバー攻撃やサイバー犯罪が国内外において発生している状況にあり，サイバー空間における脅威は，引き続き深刻な情勢.」と言及しており，引き続き対策が求められている. 近年でも，日本国内の政府機関や企業が標的型攻撃の対象となり被害を受けており，対策が求められている[3][4][5]. IPA が発足させている機器製造業者向けのサイバー攻撃の情報共有と早期対応の場であるサイバー情報共有イニシアティブ（J-CSIP：Initiative for Cyber Security Information sharing Partnership of Japan ）に提供された情報において[6]，標的型メールとみなした件数が 2019 年度では 401 件，2020 年度の上期では 96 件が報告されており，依然として標的型メール攻撃が行われていることが分かる. また，IPA が発足させたサイバーレスキュー隊（J-CRAT : Cyber Rescue and Advice Team against targeted attack of Japan）では，相談を受けた組織の被害の低減と攻撃の連鎖の遮断を支援する活動を実施している[7]. J-CRAT では,「標的型サイバー攻撃の被害の発生が予見され，その対策の対応遅延が社会や産業に重大な影響を及ぼすと判断される組織や，標的型サイバー攻撃の連鎖の元（ルート）となっていると推測される組織などに対しては，レスキュー活動にエスカレーションして支援」を実施していると記載されている. J-CRAT における 2019 年度での支援件数は 139 件，2020 年度の上期では 45 件が支援の活動実績として報告されており，サイバー攻撃の脅威が身近であり，対策が急務であることを物語っている. また，標的型攻撃のうち標的型メールが起点となっているものは 76%以上であると報告されており[8], 近年でも標的型メールを起点とした被害が報じられている[5].

標的型攻撃の対策を考える上では，標的型攻撃がどのように行われるか，その手口を理解することが重要である. 標的型攻撃の手口は様々なセキュリティ企業や団体によって整理されている[5][9][10][11]. 攻撃者はターゲットの情報を収集したのち，標的組織に標的型メールなどを手段でマルウェアを感染させた後，感染させたマルウェアを活用して標的組織で情報窃取や破壊活動を実施する[5].

マルウェアや不正な通信の検知技術，サーバの堅牢化など，サイバー攻撃防御技術の開発は進んでおり[12]，攻撃者は，標的組織での目的を達成することが以前と比べて難しくなっている. そのため攻撃者は，技術的対策が困難であり，セキュリティ上で最も弱い(Weakest Link)「人」を対象に攻撃を実施する. 攻撃は 1 点を突破すれば成功するのに対し，防御はすべての攻撃を防がなければならず，攻防はそもそも攻撃者有利の関係にある. その中でも「人」の価値観や振る舞いは非常に多様であり，その結果，人を対象とした攻撃も多岐に渡

ることになる．このため，防御側がそのすべてに対応することは格段に難しく，標的型攻撃は攻高防低が特に顕著となる事案である．この課題に対し，現在の防御策は全ての人に対して一律の対策を実施するに留まっており，巧妙な攻撃を防ぎ切ることができていないという現状にある．本研究の最終目的は，今後，より深刻化していくことが予想される，人に対する脅威の可能性について明らかにし，人ごとに適した対策を取り入れていくことで，より効果的なセキュリティ対策を実現することである．その第一歩として，本論文では，攻撃者が標的を直接揺さぶることができるメールに着目して研究を行った．今後予想される標的型メールの脅威と対策を見据え，攻撃側が擬態精度あるいは心理操作効力の高いメールをどの程度作成することができるのかに関する検討と，その結果を防御側がどのように対策に活用していくことができるのかに関する検討を行った．

IPA では，巧妙な標的型攻撃を新しい攻撃として，「ソフトウェアの脆弱性を悪用し，複数の既存攻撃を組み合わせ，ソーシャル・エンジニアリングにより特定企業や個人を狙った攻撃の総称.」と定義している[13]．文献[14]では，ソーシャルエンジニアリングは，「人を操って行動を起こさせる行為．ただし，その行動が当人の最大の利益に適合しているか否かを問わないこともある.」と定義されている．また文献[15]では，各種の説得研究から説得を「送り手が，おもに言語コミュニケーションを用いて非強制的なコンテキストの中で，納得させながら受け手の態度や行動を意図する方向に変化させようとする社会的影響行為あるいは社会的影響過程」と定義している．説得の定義を踏まえると，ソーシャルエンジニアリングは説得の一態様であり，ソーシャルエンジニアリングは説得対象に行動を起こさせる点と説得を受けた当人への利益を問わない点に特に重きを置いていることが分かる．

まず，擬態精度を高めたメールの脅威と対策を示した．今日では，企業や個人が自身の情報をインターネット上で自ら発信することが当たり前となってきた．Web 上には企業や個人に関する情報が氾濫しており，公開されている情報を組み合わせることで，個人の詳細な情報を得ることが可能であることも報告されている[18][19]．このような，公開されている情報源からの情報収集は Open Source Intelligence（OSINT）と呼ばれ，OSINT を半自動的に実行する OSINT ツールも種々出回っている．攻撃者は OSINT ツールを利用し，標的者の所属組織・上司・友人の名前・メールアドレス・関心事などを取得し，これらの情報をメールに組み入れることによって，標的者が正規のメールと区別を付けることが難しい標的型メールを，標的者ごとに作成することが可能である[20]．このような擬態精度を高める攻撃への対策として，攻撃者による OSINT の進行を状態遷移図によって定式化し，個々の攻撃段階（どこまでの情報が攻撃者に漏洩しているか）ごとに作成され得る標的型メール文面を整理した．各組織は，自らも OSINT を実行することによって攻撃者が自組織の情報をどこまで入手可能であるか確認し，本状態遷移図に照らし合わせることによって，自組織に届く可能性がある標的型メールの文面をあらかじめ把握することが可能となるため，それを踏まえたプロアクティブな対策選定に資することができる．

次に，心理操作効力を高めたメールの脅威と対策を示した．近年の研究から，インターネ

ットに公開されている情報を基に，個人の性格因子や行動特性を推定することが可能となってきている．性格因子に関しては，機械学習によって個人のツイートやブログ記事から性格因子（ビッグファイブ）を表す指標を推定する AI ツールが実運用されている[21][22][23]．かねてより，人間の行動に対しては，ある程度の心理操作（チャルディーニの法則）が可能であることが知られており[24]，チャルディーニの法則はフィッシングメールにも影響を与えることが判明している[25]．さらには，個人の性格因子に応じて有効なチャルディーニの法則が異なることも報告されている[26]．行動特性に関しては，人間が詐欺に引っ掛かる傾向（詐欺による説得の受けやすさ）にあるか否かを推定する心理テストが開発されている[27]．そのため，上述の性格検査ツールと同様に，個人のツイートやブログ記事から行動特性を推定する AI ツールを作成することも可能であると考えられる．さらに，個人のツイートやブログ記事は，OSINT ツールによって取得することが可能な情報である．このため，「攻撃者が，OSINT ツールにより得られる標的者の情報を基に，AI ツールにより標的者の性格因子や行動特性を推定し，それらに応じた心理操作テクニックを悪用することで，心理操作効力の高い標的型メールを，標的者ごとに作成する」，という新たな攻撃が考えられる．ここで，既存研究では，性格因子，あるいは行動特性のどちらかに着目して，セキュリティ上の影響を評価していたが，本研究で検討する新たな攻撃は，人間に対して有効な心理操作効力は，人間の性格因子，行動特性の両面から観察する必要があるという仮説に基づく．

技術が発展し，攻撃者が人に対してカスタマイズした巧妙な攻撃を，OSINT ツールや AI ツールにより容易に行うことができるようになってきている．一方で，現状では全ての人に対して一律の対策を実施しているため，前述したようなカスタマイズされた巧妙な攻撃を防ぎ切ることができない．本研究の位置付けは，今後，より深刻化していくことが予想される，人に対する脅威の可能性について明らかにし，人ごとに適した対策を取り入れていくことで，より効果的なセキュリティ対策を実現することである．その第一歩として，攻撃者が標的を直接揺さぶることができるメールに着目して研究を行った．今後予想されるメールの脅威として，擬態精度の高いメールと，心理操作効力の高いメールの実現性と，その対策を検討した．

ここまでの，OSINT を起点とした擬態精度と心理操作効力を高める攻撃の全体像を図 1 に示す．本論文では，擬態精度と心理操作効力のそれぞれに対し，その脅威の実現性と対策を示す．

このような擬態精度を高める攻撃への対策として，どのような情報が漏洩しているかによって作成される標的型メール文面を状態遷移図によって整理した．本状態遷移図とそれに対応するメール文面により，標的となる組織に対して届く可能性があるメール文面をあらかじめ把握することで，注意を促すことができる．

このような心理操作効力を高める攻撃への対策として，標的者が心理操作効力の高いメールを受けた際に，利用されている説得手法を通知することで標的者が説得を受けようとしていることを示し，攻撃者の望む行動を取らないように説得することが有効であることを

示す．具体的には，メールにおいて，チャルディーニの法則が検知された際には該当箇所に対して，標的者を説得する文言が含まれていることをアラートの形で通知する．この時，アラートの提示でも同様に性格因子と行動特性が，提示の有効性に寄与すると考えて，性格因子と行動特性によってアラートの効果が異なることを検証した．400 人規模のアンケート調査を実施し，効果的な注意喚起の方法（メール内のどの説得フレーズに対してアラートを提示すると，標的者は標的型メールへの不信感を高めるか）も，標的者の性格因子と行動特性の両者によって異なることを示した．

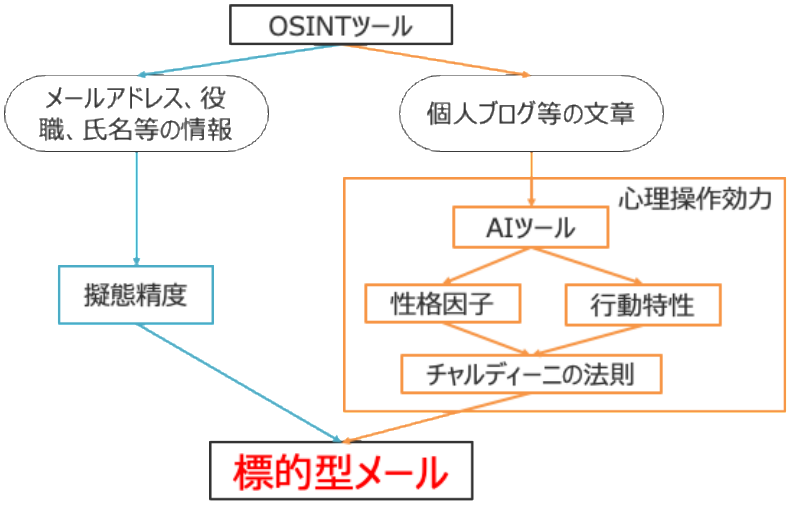

図 1　OSINT ツールと AI ツールを用いた標的型メール攻撃の全体像

1.2. 本論文の構成

1 章では本研究の背景と全体の概要を述べた．2 章では，不審メール（標的型メールや，特定個人を狙わずに同一の内容のメールを多数の送信先にばら撒くフィッシングメール）に対する対策と，説得心理学とソーシャルエンジニアリングの関係，人を対象とするセキュリティの既存研究，OSINT ツールによる擬態精度・心理操作効力が高いメール作成につながる既存研究を示す.

3 章から 5 章では，全体のリサーチクエスチョンである「組織・企業でのサイバーセキュリティにおいてもっとも弱い箇所である「人」に着目し，標的型攻撃のリスクを低減することを目的に，人に着目した新たな攻撃の可能性を示した上で，攻撃の対策としても人に着目することで有用な対策を提示できるか」を示すために，攻撃側と防御側の両面から擬態精度と心理操作効力の脅威の実現性とその対策方法を示す（図 2）.

3 章では，擬態精度を高めたメールの脅威と対策を示す．具体的には，インターネットに公開されている情報源から標的者に関する様々な情報を OSINT ツールによって次々と取得し，これらの情報をメールに組み入れることによって，擬態精度が高いメールを標的者ごとに作成することが可能であることを示す．さらにこのような擬態精度を高める攻撃への対策として，攻撃者による OSINT の進行を状態遷移図によって定式化し，個々の攻撃段階（どこまでの情報が攻撃者に漏洩しているか）ごとに作成され得る標的型メール文面を整理した.

4 章では，心理操作効力を高めたメールの脅威を示す．攻撃者は，AI ツールによって推定した標的者の特性に基づいて，その標的者に有効なチャルディーニの法則を推定することで，心理操作効力の高い標的型メールを標的者ごとに作成することができる．本研究では，100 人規模のアンケート調査を実施し，個人の性格因子とチャルディーニの法則の相関関係が，行動特性によって異なることを示す．これまで，種々の既存研究において説得のされやすさ（チャルディーニの法則に対する感受性）と性格因子との相関関係，あるいは，説得のされやすさ（詐欺師の説得を受け入れてしまいやすさ）と行動特性との相関関係がそれぞれ個別に調査されてきたのに対し，説得のされやすさ（チャルディーニの法則に対する感受性）を性格因子と行動特性のコンビネーションによって分析したことが，本研究の第一の貢献である.

5 章では，このような心理操作効力を高める攻撃への対策として,メール文面でチャルディーニの法則が使われている箇所を通知することで，標的者に注意を促すことが有効であると考えられる．本研究では，400 人規模のアンケート調査を実施し，効果的な注意喚起の方法（メール内のどの説得フレーズに対してアラートを提示すると，標的者は標的型メールへの不信感を高めるか）も，標的者の性格因子と行動特性の両者によって異なることを示した.

6 章では，本研究のまとめと今後の展望を示す．これまでの既存研究においては，標的型メールの脅威分析（攻撃側が擬態精度あるいは心理操作効力の高いメールをどの程度作成することができるのか）に主眼が置かれていた．これに対し本研究では，標的型メールへの対策（防御側が脅威分析の結果をどのように対策に活用していくことができるのか）にまで足を踏み込んだ．攻撃者による脅威をあらかじめ把握することで，標的者に応じた有効な対策を配備することができることを示したことが，本研究の第二の貢献である．本研究は，著者が知る限り，個々人の特性に応じてセキュリティシステムの防御方法を変える考えを具体化した初めての研究である．

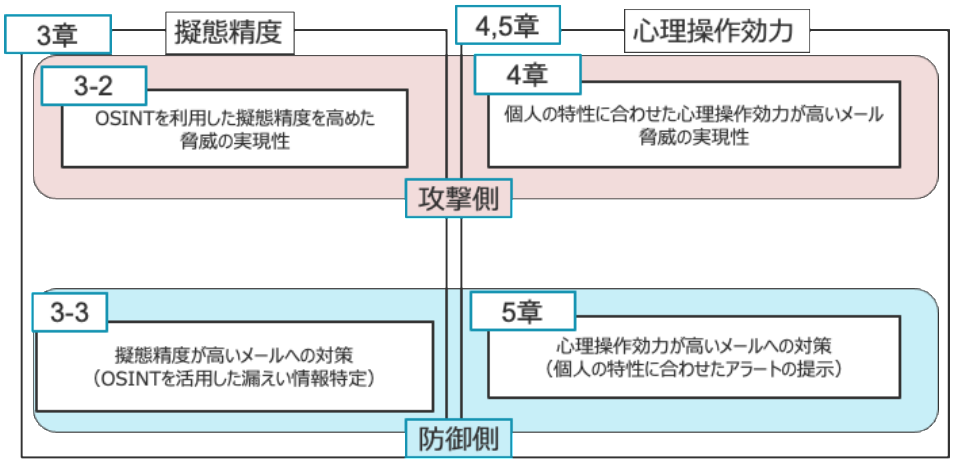

図 2　論文の構成

第2章　　メールや人に対するセキュリティの既存研究

本章では，不審メール（標的型メールや，特定個人を狙わずに同一の内容のメールを多数の送信先にばら撒くフィッシングメール）に対する対策と，説得に関わる心理学の研究，人を対象としたセキュリティの既存研究を示す.

2.1. 不審メール対策の既存研究

CipherCraft/Mail[28]は，受信メールを，送信ドメイン認証結果や送信経路といった挙動と，名称やアイコン偽装といった添付ファイルに関する不審点をもとに検査し，自動隔離・注意喚起する技術である．信頼のおける人物に感染した後に，その人物のメールアドレスを利用してメールを送る攻撃では，挙動に関する不審点は検知できず，高度な攻撃者による添付ファイルが作成される場合，サンドボックスによる検知を通過するため，本技術では検知できない.

Disarm[29]は，添付ファイルのドキュメントが悪性である可能性があるコード（マクロ等）を含む場合，該当コードを除去し，ドキュメントを再構成することで，悪性マクロの実行を予防する．しかし，マクロ等を活用している組織である場合，Disarm を無効にすることが公式で推奨されているため，そのような組織では有効に働かない.

Sevtap D らの手法[30]は，個人ごとに，メール文面に特徴が存在することを利用し，不審なメールを検知する手法を提案している．本手法では，まず不審であるかの識別対象である個人ごとにメールを収集し，個人ごとの特徴量をサポートベクターマシン(SVM)で学習する．学習した分類器により，受信したメールが，予め学習した人物からのものであるかを判定することで，届いたメールが，正しく本人からの文章であるかを判断し，不審な成りすましメールを検知することができる．しかし，認識精度は 67%～100%とまばらであり，確度を持って本人からメールであると言うには信頼性が低いことと，本人識別を通過するように，本人の特徴を学習する巧妙な攻撃には無力である，という課題がある.

Gascon らの研究[31]では，70 万通以上のメールを利用して，特にヘッダ情報を正規のメールに見えるように改竄するような標的型メールの特徴を分析・抽出し，添付ファイルの種別や，In-Reply-To ヘッダ情報といった特徴量を 13 個選別し，その特徴量をもとに標的型メールを検知する手法を提案している．本手法は，対象のドメイン情報などを知らない攻撃者モデル（Blind Spoofing）の場合には高い効果を示す．一方で，送信者を偽らないような攻撃者は対象ではないため，これまで関係がないような，製品の問合せを行うような外部の攻撃者を検知することは困難である.

Ho G らの研究[32]では，企業のメールデータセットを利用した大規模な学習データをも

とに，不審メールを検知し，その傾向を分析している．本研究における検知手法は，セキュリティ企業やユーザが不審であると判断することで不審であるとラベル付けされたメールを学習データとして利用することで識別器を作成し，不審メールを検知している．本手法では，セキュリティ企業が前もって不審なメールであると気づくか，社内にて不審であると誰かが気付くことによって初めて機能するため，標的組織に気づかれないように巧妙に侵攻する攻撃の場合には攻撃を検知することが困難である．

これらの技術が有する課題は，巧妙な攻撃者による標的型攻撃メールを検知することができない点である．例えば，企業の問い合わせ窓口に対して商品の問い合わせを行い，正常なやり取りを行う人物であると信頼を獲得した後，攻撃者の目的であるマルウェア感染等を引き起こすメールを送るような攻撃を考える．このような攻撃の場合，攻撃者が外部のアドレスからメールを送信することはありうるため，送信者ドメイン認証や送信経路から不審であると判断することはできない．さらに，本人性を検証しても，初めから攻撃者本人からのメールであるため，本人であるかどうかでは攻撃を判断することができない．

2.2. 説得心理学とソーシャルエンジニアリングとの関係

本節では，説得についての研究である説得心理学と，ソーシャルエンジニアリングの関係について示す．その上で，説得の一種であるチャルディーニの法則を説明する．

2.2.1. ソーシャルエンジニアリングと説得心理学

文献[14]では，ソーシャルエンジニアリングは，「人を操って行動を起こさせる行為．ただし，その行動が当人の最大の利益に適合しているか否かを問わないこともある.」と定義されている．また文献[15]では，各種の説得研究から説得を「送り手が，おもに言語コミュニケーションを用いて非強制的なコンテキストの中で，納得させながら受け手の態度や行動を意図する方向に変化させようとする社会的影響行為あるいは社会的影響過程」と定義している．説得の定義を踏まえると，ソーシャルエンジニアリングは説得の一態様であり，ソーシャルエンジニアリングは説得対象に行動を起こさせる点と説得を受けた当人への利益を問わない点に特に重きを置いていることが分かる．さらに，ソーシャルエンジニアリングは対面や，電話，メールなど，人とコミュニケーションを取るもの全般において行なわれる活動である．

ソーシャルエンジニアリングを成功させるために重要な点は，標的から攻撃者であると疑われないように，攻撃ではない正規の問い合わせのように見せる（擬態精度）ことと，説得対象に行動を起こさせるために心理的に働きかけること（心理操作効力）の両面となっている．

2.2.2. ソーシャルエンジニアリングにおける擬態精度

ソーシャルエンジニアリングを成功させるためには，標的に説得されているという違和感を感じさせないために，ソーシャルエンジニアリングであることを悟られないようにする必要がある．ソーシャルエンジニアリング中のやりとりが害のないようなやりとりと思わせるために，役作りが必要となる（擬態精度を高める）[14]．以下に，文献[14]に記載されている役作りの根幹を一部示す．

- 調査を多くすればするほど，成功の可能性が高まる．
- 役作りは単純であればあるほど，成功の可能性が高まる．
- 役作りは自然に見えなければならない．

2.2.3. チャルディーニの法則

チャルディーニの法則は，社会心理学者であるチャルディーニによって提唱された，相手を自分の思いどおりに誘導させるための心理法則である[24]．チャルディーニの法則には，希少性，返報性，権威，一貫性，好意，社会的証明の 6 つの法則が存在する．以下に，それぞれの概要を示す．

- 希少性（Scarcity）
  「限られたものほど，価値があると感じてしまう」という心理法則．
- 返報性（Reciprocation）
  「人から受けた恩は，返したくなる（返さなければならないと考える）」という心理法則．
- 権威（Authority）
  「肩書や経験などの“権威”を持つ者に対して，信頼を置いてしまう」という心理法則．
- 一貫性（Commitment and Consistency）
  「自分の行動に一貫性を持たせようとする（持たせたいと考える）」という心理法則．
- 好意（Liking）
  「好意を持っている人からの要請を受けると，積極的に応えようとする」という心理法則．
- 社会的証明（Social Proof）
  「周囲の動きに同調したくなる」という心理法則．

2.3. 人を対象とするセキュリティの既存研究

本節では，人を対象とするセキュリティの既存研究を示す．まず，人の特性の一つである性格因子として，ビッグファイブを示したのちに，心理特性とセキュリティ意識の既存研究と，メールと心理特性，説得手法に関する既存研究を示す．

2.3.1. ビッグファイブ

ビッグファイブは，ゴールドバーグによって提唱された，主要 5 因子性格モデルと呼ばれる性格評価尺度の 1 つである[21]. パーソナリティを理解するうえでの包括的かつ明瞭なモデルであるといわれ，医療や消費嗜好調査，仕事のパフォーマンス分析等多くの領域で用いられている[22]．ビッグファイブは以下の 5 つの要素に基づき，性格の各要素の強さを定量的に評価する．

- 情緒不安定性(Neuroticism)
  怒りや心配事といったネガティブな心理的反応に過剰に反応する性質を持つ．スコアが高い場合，人が不合理な考えを持っている傾向があり，衝動を制御する能力が低く，ストレスへの対処が不十分であることを示す．スコアが低い場合は，感情が安定していることを示しています．これらの人々は通常，落ち着いていて，気性があり，リラックスしていて，動揺することなくストレスの多い状況に直面することができる傾向がある．
- 外向性(Extroversion)
  社会性を持ち，自己主張を行い，饒舌な性質を持つ．スコアが高い場合，エネルギッシュであり，楽観的である．スコアが低い場合は，ひかえめな性格であり，独立性を重んじる傾向がある．
- 開放性(Openness)
  感情や芸術に対して理解が深く，好奇心が強い，独立心が高い，といった性質を持つ．スコアが高い人は，型にはまらない傾向がある．スコアが低い人は，行動が慣習的で，見通しが保守的である傾向がある．
- 調和性(Agreeableness)
  利他的であり，他者へ共感し，だれかを助けると同時に他人も自身を助けると信じる性質を持つ．スコアが低い人は，自己中心的で，他人の意図に懐疑的で，協力的というよりは競争的である傾向がある．
- 誠実性(Conscientiousness)

組織的な行動を好み，計画性を持ち，効率的に行動する性質を持つ．スコアが高い人は，目的があり，意志が強く，秩序を重んじる傾向があるため，気難しい側面もある．スコアが低い人は，高い人ほどは規則を厳密に適用しない傾向がある．

2.3.2. メールと擬態精度に関する既存研究

Rajivan らは，攻撃者の行動を推定するための実験[16]．どのような攻撃が有効であるかを，攻撃者側として実験してもらうことで調査した．調査の結果，メールをリマインドすることや，権威がある人物を装うこと，親しみを感じさせること，などが有効であることが分かった．

Goel らは，メールの内容が当事者に取って関係するかどうかによってフィッシングの効

果が変わるかを，実際にメールを送信して調査した[17]．学生に対してメールを送付する実験を実施した結果，ギフトカードや iPad が当たるという内容と比べ，授業の履修登録期限が迫っていることを伝えるメールの方が開封されやすく，メールに記載されている URL をクリックする可能性も高いことが分かり，対象が興味を惹くような内容を送ることがフィッシングメールにおいて重要であることを示した．

2.3.3. 心理特性とセキュリティ意識の既存研究

小川らは，不用意に SNS などで情報を開示する人間の心理的特性が，ユーザへの信頼，情報開示範囲のコントロール，リスク認知といった要因であることを，アンケートによって明らかにした[33].

寺田らは，ウイルス感染や不正利用，プライバシ漏洩といった IT 被害に遭いやすい人間を特徴づける因子があることを，アンケートにより明らかにした[34]．さらに，片山らは，文献[34]の結果を基に，被害リスク判定ツールを試作し，評価した[35][36].

Modic らは，詐欺における説得のされやすさ（詐欺師の説得を受け入れてしまいやすさ）が，個々人の行動特性によって異なることを報告している[27]．そのうえで，社会心理学や行動経済学で行われてきた研究を基に，詐欺のされやすさを評価する指標(Susceptibility to Persuasion-II: StP-II) を，54 の行動特性で示した．詐欺に引っかかりやすいかを確認するための質問に対する回答と，行動特性との間で因子分析を行い，主要な行動特性を抽出することで StP-II は構成されている．

八藤後らは，対話を想定した攻撃の場面で標的者がとる行動の関係を分析した[82]．本研究では，人間の脆弱性を狙った攻撃に対するチャルディーニの法則(ハドナジーがあげる人間の脆弱性)の影響を，シナリオベースのアンケートによって調査している．

個人の性格因子に応じて有効なチャルディーニの法則が異なることも報告されている一方で，Egelman らは，ビッグファイブは個人のプライバシに対する趣向を説明する因子としては弱く，意思決定の傾向と個人のプライバシに対する趣向との間に，強い相関が存在していることを示している[58]．この結果は，標的型メールに対する反応度も，個々人の性格因子以上に，行動特性によって左右される可能性があることを示唆している．

これらの分析は，「人」に着目しているが，人の特性として，どのような攻撃が有効であるか，人がどのようなミスをするかを示すことで教育に応用できると論ずるにとどまっており，具体的な対策については言及されていない．

2.3.4. メールと心理特性，説得手法に関する既存研究

Butavicius らは，チャルディーニの法則のうち，希少性，権威，社会的証明がフィッシングメールや標的型メールに有効であるかを実験により調査した[37]．実験で利用したメールは，大学に実際に届いたメールを基にしている．実験は，研究室に実験協力者を呼び，研究室に配置されたコンピュータ上で 12 件のメールを提示し，各メールに対して URL をク

リックするかどうかを 5 件法で尋ねることで実施した．実験の結果，標的型メールではチャルディーニの法則が有効であるが，フィッシングメールではむしろ効果が下がることが分かった．

Williams らは，従業員に対して模擬的な標的型メールを送信し，緊急性や権威がメールに記載されている URL をクリックする確率を上げるかを評価した[38]．実験の結果，権威が有効であることが分かった．

文献[37][38]では，一部のチャルディーニの法則が標的型メールの文面に与える影響を評価しているが，全てのチャルディーニの法則において効果があるかを評価していない．

文献[37]では，一貫性，好意，返報性の法則は，相手と何通かのやりとりがなければ文面に入れ込むことはできないという考えのもとで，希少性，権威，社会的証明の三つに絞り実験を行っていた．これは，実験として一通のメールを提示した際の反応を評価しているためである．一方で，文献[25]では，文献[37]と同様の実験ではあるが，全てのチャルディーニの法則に対して文面を作成し，評価している．本論文では文献[25]の考えを採用し，全てのチャルディーニの法則の効果を検証する．

2.4. OSINT ツールによる擬態精度・心理操作効力が高いメール作成につながる既存研究

攻撃者は OSINT ツールを利用し，標的者の所属組織・上司・友人の名前・メールアドレス・関心事などを取得し，これらの情報をメールに組み入れることによって，標的者が正規のメールと区別を付けることが難しい標的型メール（「擬態精度」が高い標的型メール）を，標的者ごとに作成することが可能である[20]．

一方，近年の研究から，インターネットに公開されている情報を基に，個人の性格因子や行動特性を推定することが可能となってきている．性格因子に関しては，機械学習によって個人のツイートやブログ記事から性格因子（ビッグファイブ）を表す指標を推定する AI ツールが実運用されている[21][22][23]．

行動特性に関しては，人間が詐欺に引っかかる傾向（詐欺による説得の受けやすさ）にあるか否かを推定する心理テストが開発されている[27]．個人のツイートやブログ記事から性格因子を推定するアルゴリズムは，ツイートやブログ記事と性格因子とを対応づけることでその出力が得られるように AI ツールを学習させている．それと同様に，ツイートやブログ記事と行動特性とを対応づけて学習させることで，ツイートやブログ記事から行動特性を推定する AI ツールを作成することが可能となる．そのため，上述の性格検査ツールと同様に，個人のツイートやブログ記事から行動特性を推定する AI ツールを作成することも可能であると考えられる．

アリストテレスの弁論術など，かねてより，人間の行動に対しては，ある程度の心理操作が可能であることが知られている．心理操作の有名な手法の１つとして，相手を自分の思い

どおりに誘導させるための心理法則であるチャルディーニの法則も知られている[24]．チャルディーニの法則はソーシャルエンジニアリングに応用できる可能性が示されており，具体的には，不特定多数に対する詐欺メールであるフィッシングメールに影響を与えることが判明している[25]．さらには，個人の性格因子に応じて有効なチャルディーニの法則が異なることも報告されている[26]．また，個人の行動特性によって詐欺に引っかかる傾向が判別できることから[26]，個人の行動特性に応じても有効なチャルディーニの法則は異なることが推察される．また，チャルディーニの法則は，特定人物に対する詐欺メールである標的型メールの成功率を高める（「心理操作効力」が高い標的型メール）ことが報告されている[37][38]．

2.5. 2 章のまとめ

本章では，不審メール（標的型メールや，特定個人を狙わずに同一の内容のメールを多数の送信先にばら撒くフィッシングメール）に対する対策と，ソーシャルエンジニアリングや説得心理学との関わり，人を対象としたセキュリティの既存研究として，心理特性とセキュリティ意識の関係性の示した研究や，メールにおける心理特性や説得手法の研究を示した．

機械的な対策である不審メール対策技術では，巧妙な攻撃者による標的型攻撃メールを検知することができないことを課題としてあげた．このようなメールへの対策を効率的に実施するために，個々人に合わせた対策が必要となる．

人を対象としたセキュリティの研究はいくつかあるが，それらは「どのように人を攻撃することで攻撃者が目的を達成することができるか」を示すにとどまる．

さらに，性格因子のみと説得手法との関係性や，行動特性とセキュリティ行動との関係性についての言及はあるが，性格因子と行動特性の両方を同時に見ることで，より個人に対して有効となる説得手法を調査した研究は著者が知る限りない．

第3章　　擬態精度を利用した攻撃と対策

本章では，擬態精度を高める攻撃として，OSINT を利用する巧妙な攻撃者による脅威を分析し，その対策として，状態遷移モデルによる脅威分析を示す.

3.1.　OSINT による擬態精度を高めたメール作成

Open Source Intelligence(OSINT は，)公開されている情報の中から必要な情報を収集する諜報活動をいう．公開されている情報（以下, OSINT データと呼ぶ）とは，新聞の報道, 政府の公報から電話帳に書かれている情報まで幅広い範囲が含まれる. 近年は, SNS や Web サイト上で，多くの個人や企業が自身に関わる情報を発信しており，個人や企業に関わる膨大な OSINT データが入手可能な現状となっている.

OSINTデータは，経済予測や流行分析などといった目的で有用な情報である. しかし，これら OSINT データは，標的型メール攻撃，より正確には，ソーシャルエンジニアリングをしかける攻撃者にとっても有益な情報となる．OSINT データを組み合わせることによって個人情報やプライバシ情報を得ることが報告されており[19]，また, OSINT 活動をサポートするツールも出回っている．今や攻撃者は, HUMINT や SIGINT 等の「手間のかかる諜報活動」を行わずとも，OSINT データから攻撃対象者に関する情報を収集することが可能な状況にある．そして，その情報を利用することによって，攻撃者は信憑度の高い（対象者が騙されやすい）標的型メールの作成が可能である.

以下に，これら脅威について具体例をあげて説明する．下線部が OSINT データに対応する部分である.

【具体例】

攻撃者が，大学教授である A 氏に標的型メールを送信する場面を想定する.

1. 攻撃者は A 氏の名前を用いて，メールアドレスを Web 検索し，ドメインが「*.ac.jp」であるメールアドレスを得る.
2. 攻撃者は名前やメールアドレスを用いて関連する Web ページを検索し，研究室の HP から, A 氏がどのような研究を行い，業績を挙げているか調べる.
3. 攻撃者は続けて関連する Web ページを調査し, A 氏が以前, X 社で講演したという情報を得る.
4. 攻撃者は X 社について Web 検索して関連する Web ページを調べ，現在注力している事業や社員のメールアドレスの書式といった情報を得る.
5. 攻撃者はメールのヘッダを偽装し X 者の社員になりすまし，以前の講演が素晴らしかったこと，現在注力している事業が A 氏の研究と関連していること, A 氏に再び講演して欲しいことをメールに記載し，詳細は添付ファイルを参照して欲しいと述べ，エクスプロイトコードを含む PDF ファイルを添付し A 氏にメールを送信する.

本シナリオにおける攻撃者は，当初，A 氏の名前と職業に関する情報しか有してなかったが，Web 上から多くの OSINT データを収集することができた．その結果，攻撃者は，手順 5 に示したような信憑度の高い標的型メールを作成することに成功している．

3.1.1. ソーシャルエンジニアリングにおける OSINT の有効性

ソーシャルエンジニアリングにおける OSINT の有効性について，文献[20][39][40][41]の研究が行われている．

Ball らは，世界中で蓄積されるデータ量が指数関数的に増加していることから，OSINT がソーシャルエンジニアリングに活用され，犯罪行為に利用されてしまう可能性について言及し，OSINT を使用して組織の従業員に標的型メール攻撃を仕掛ける方法について議論している[20]．その中で，攻撃者が情報収集のために OSINT ツールを使用することや，標的型メール攻撃を行うために専用のツールを使用する方法について紹介している．

Edwards らは OSINT データを，Bootstrap（攻撃のきっかけとして利用されるデータ）と Accentuator（攻撃の有効性を高めるために補助的に使用されるデータ）の 2 つに分類し，ソーシャルエンジニアリングにおいて，それらのデータがどのように用いられるかを示した[40]．また，水道・ガス・電気などの公共事業を担う企業について，従業員の名前や電話番号，メールアドレスといった情報が，OSINT からどの程度集めることができるのかを調査し，それらがもたらすソーシャルエンジニアリングの脅威について明らかにしている．

Silic らは，フォーチュン 500 の企業を対象として，SNS を活用したソーシャルエンジニアリングの有効性について検証を行っている[41]．具体的には，OSINT を活用して，攻撃対象の企業の従業員になりすました「偽の SNS アカウント」を作成し，正規の従業員により構成される SNS プライベートグループのメンバーになることで，企業に関する情報を入手することが可能であるかの実験を行った．その結果，従業員は容易に欺かれ，ソーシャルエンジニアリングの被害を受けやすいことや，組織は現状，SNS メディアからのセキュリティ脅威を制御する術がないことが明らかになっている．

本稿は，複数の OSINT ツールを調査して攻撃者の情報収集プロセスを「状態遷移図」として体系化し，標的型メール攻撃に焦点を当てて各状態における情報の具体的な活用方法（標的型メールの作成）を分析しており，上記の既存研究を補完あるいは補強する関係にある．

機械学習とソーシャルエンジニアリングについて文献[42]のような研究が行われている．

Singh らは，標的型攻撃の対象を企業の CFO（最高財務責任者）として，その人物が「標的となるか（騙されやすいか）」を機械学習によって予測する学習モデルを提案している[42]．学習データには，CFO の年齢や性別，Twitter のフォロワーやツイート数，Twitter や LinkedIn 経由のフィッシング成功の有無などを用いている．学習によって，80%の精度で標的になる（騙される）人物の特定が可能であるという結果が出ており，機械学習を用いたソーシャルエンジニアリングの進化について明らかにしている．

標的型メールの自動生成に関しては, 岩田らが標的型メール攻撃対策訓練における, 訓練メールの自動生成のための受信メールの分析手法の提案を行っている[43]. ユーザの受信 BOX メールを分析して, どのようなメールを信頼して開封しているか確認した上で, 「普段のメールと似ているが, 標的型攻撃メールだと気づくことができる不自然さ」をメールに盛り込むことによって, 訓練メールが作成される.

Singh らや岩田らの研究は, 攻撃者がビッグデータ技術や AI 技術の活用することによって, 洗練されたソーシャルエンジニアリングを実行可能であることを示している. 今後は, より高度なソーシャルエンジニアリングや標的型メールに対して備えることが重要であると考えられる.

3.1.2. OSINT ツールの例

OSINT ツールとは, Web 上にある膨大な OSINT データの中から, 検索対象に関わる OSINT データだけを効率的に入手する為に用いられるツールである. Maltego, Creepy など, 既に数多くの OSINT ツールが公開されている. 以下に, OSINT ツールの代表例である Maltego と Tinfoleak を利用して, OSINT ツールの詳細な動作を説明する.

Maltego は Paterva 社によって開発されたデータ収集・可視化ツールである. 名前, ドメイン, URL のいずれか, または, その組合せを入力すると, それに紐づくありとあらゆる情報を Web 上から自動的に収集する. 例えば, Maltego に対して名前「minamigaki」を入力した場合, 図 1 の形式で電話番号, メールアドレス, 関連 Web サイトのリストが得られる. 本例では SNS アカウントが見つからなかったが, Twitter や Facebook 上で一致するアカウント名が存在する場合は, 該当アカウントの情報も取得することが可能である.

また, ここで得られた情報を更に入力することによって, 連鎖的に情報を収集していくこともできる. 例えば, 名前を入力することによって得たドメイン名「minamigaki.cs.inf.shizuoka.ac.jp」（図 3）を Maltego に入力した場合, PDF ファイル, 関連 Web サイト, 関連ドメイン, Web サーバソフトウェア情報, メールアドレス, IP アドレス, IP アドレスから求まる位置情報が得られる（図 4）.

このように, 前節のシナリオにおける手順 1～4 の検索作業の多くが, Maltego によって半自動的に実施可能であることが分かる.

Tinfoleak は, TwitterID を入力情報と以下の情報を得ることができる. Maltego が Web 上全体を検索対象としているのに対し, Tinfoleak は特定サービス（Twitter）上に限定した情報収集を行う.

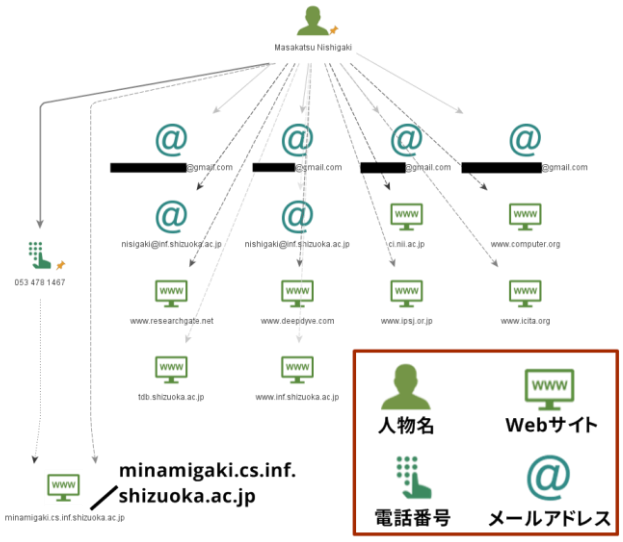

図 3　Maltego による結果 1（入力：名前）

図 4　Maltego による結果 2（入力：ドメイン名）

- ユーザの基本情報（名前，画像，所在，フォロワー等）
- ユーザが使用するデバイスと OS
- ユーザが使用するアプリケーションと連携 SNS
- ツイートに付与された位置情報（座標）
- ユーザが投稿した写真
- ユーザが使用したハッシュタグと使用された時刻
- ユーザと親密な関係にある他ユーザの特定
- ユーザが興味を示すトピック（趣味等）

これらの情報は，本来，対象ユーザの Twitter におけるプロフィールやツイート，他ユーザとのやりとりをくまなくチェックして分析を行ったり，位置情報が付与されているツイ

ートを手作業で探し出したりすることで得られる情報である．Tinfoleak は，これらの情報を自動的かつ瞬時に提供する．

これら OSINT ツールは，正規ユーザが，自身のパーソナル情報がどの範囲まで拡散しているか確認したり，自身が利用しているサーバの脆弱性を診断する（デーモンやアプリケーションのバージョンを確認する），といった場面で非常に有用である．しかし，場合によっては，標的型メール攻撃（より正確には，ソーシャルエンジニアリング）を支援するツールとして攻撃者に悪用されてしまうという二面性を有する．

3.2. OSINT フローチャートによる擬態精度を高めた標的型メールの脅威

本節では，OSINT の手法を，既存のフローチャートを基に，著者らの知見を合わせたフローチャートを提供する．本フローチャートにより，擬態精度を高めた標的型メールの脅威の実現性を明らかにし，その脅威を認識する．

さらに，OSINT のフローチャートを作成することによって，次の利点がある．

① セキュリティに関する高度な知識を持っていない人でも，簡単に「攻撃の起点となる情報」をオープンソースから収集することができる
攻撃を実際に行うことでシステムに弱い箇所がないかを調べるペンテスターは単価が高いが，フローチャートを利用して，高度な知識を持たない人でも作業ができるようになることで，コストを抑えることができる．また，専門的なペンテスターが，より高度な業務に集中することができる．

② OSINT 情報を人手によってデータ化（把握）することができる

人海戦術的に，OSINT で取れる情報を（可能であればカテゴリ別に）データ化することにより，公開されている（漏れている）情報を把握することができる．「推測」等のヒューリスティックな作業も含まれるので，機械での全自動化がまだ難しい．公開を意図していなかった「想定外の情報」も，多くの人間を投入することができるため，見つけることが期待できる．

本節では，まず既存のフローチャートを紹介した後，著者らの知見を合わせて作成したフローチャートを説明する．

3.2.1. 既存の OSINT マニュアル

Buscader というディストリビューションと共に，OSINT のフローチャートが公開されている[44]. 図 5 に, 公開されている OSINT のフローチャートを引用したものを示す[46]. 本図では，”Email Address”を起点として，ノードの背景が灰色になっている”Email Address”, “User Name”, “Employer”, “Real Name”, “Websites/Blogs”, “User Names”, “Social Networks”，といった情報を収集する際の作業がフローチャートで示されている．

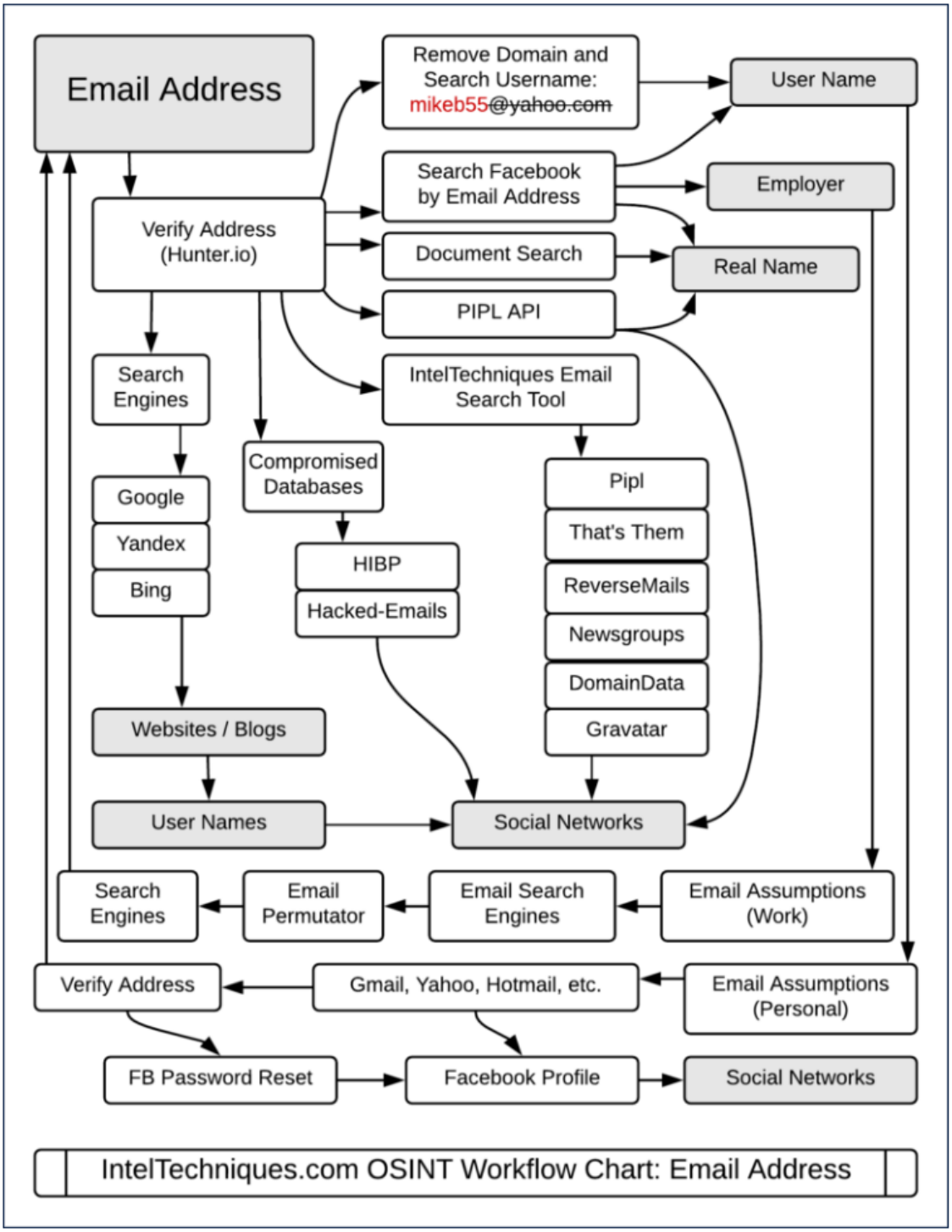

図 5　OSINT のフローチャート[46]

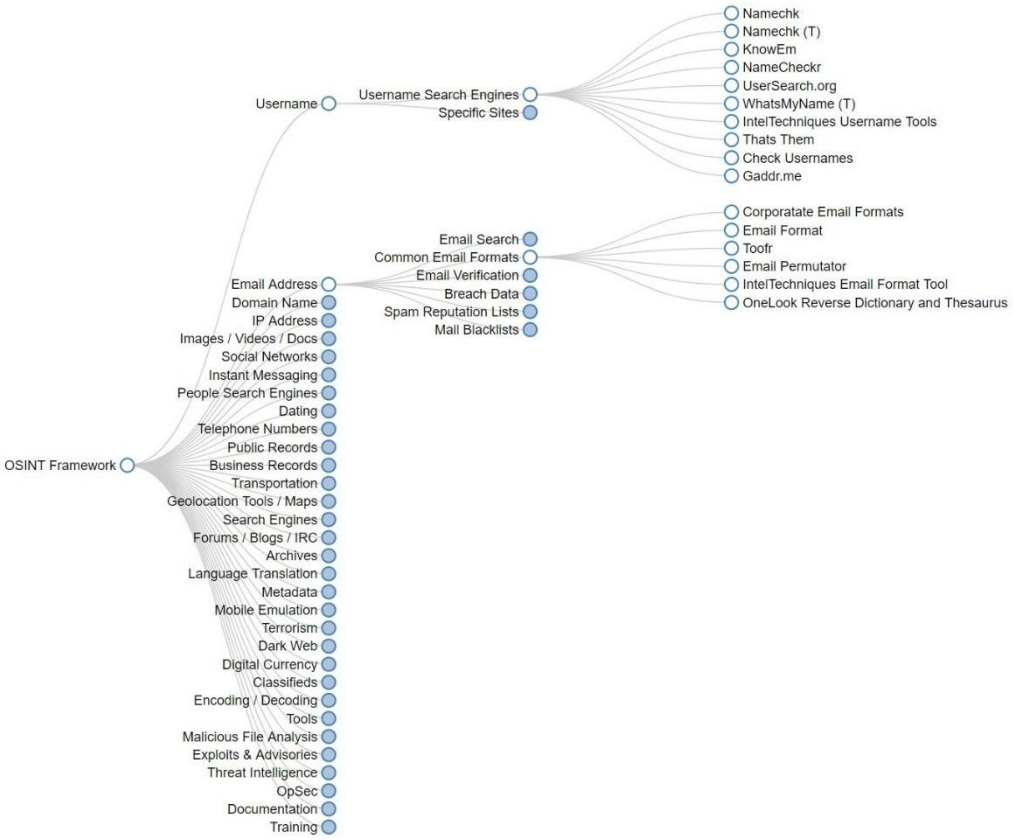

図 6　OSINT Framework[47]

3.2.2. 著者らが作成した OSINT マニュアル

3.2.1 項で紹介したフローチャートは，英語圏で提案されているものであるため，英語の情報を軸に検討されている．また，作成された時期が古いものがあるため，現在は使えないツールの情報も一部含まれている．そのため著者らは，これらの既存の OSINT フローチャートの知見と著者らの経験を踏まえて，日本向けの OSINT フローチャートを作成した[48][49][50][51][52][53][54][55][56]. 著者らが作成した OSINT フローチャートを図 7 に，図中の各要素の詳細を表 1 に示す.

図 7 のフローチャートの見方を説明する．丸四角で表現されるノードはデータであり，四角で表現されるノードは，処理を指す．操作・アプリケーションのノード近くに記載してある丸で囲われた数字は，表 1 の”No”列の値に対応している.

まず，赤い丸四角で表示されている”Target Domain”が調査の起点となる．OSINT をする人間は，Target Domain ノードと接続しているノード 6 つのノードである，(1)Whois Lookup, (3)Email Hunter, (4)Email Harvester, (6)Maltego, (8)”site” Search, (18)Metagoofil，という処理によって情報を収集する．ここで，括弧内の数字は，図中の丸で覆われた数字に対応している．各処理ノードにおける処理区分（位置情報の収集やメールアドレスの調査といった情報），属性（機械による処理なのか人間による処理なのか），処理への入力とその出力等を整理したものを表 1 に示す.

図 7 と表 1 を用いて，OSINT の例を示す．まず，Target Domain を起点に(6)Maltego を用いることで，攻撃者は「Email Address, Employee Name, Phone Number」を取得することができる．ここで得られた Email Address から，法則を導き出すことで，Email Address Format を推定する（例：Yamada.Taro@xxcompany.co.jp であれば，ユーザ名が“(名字)．(名前)”，と推定することができる）．この推定結果を用いて，攻撃者が標的としたい人物の氏名によって，対象人物のメールアドレスを推定することが可能となる.

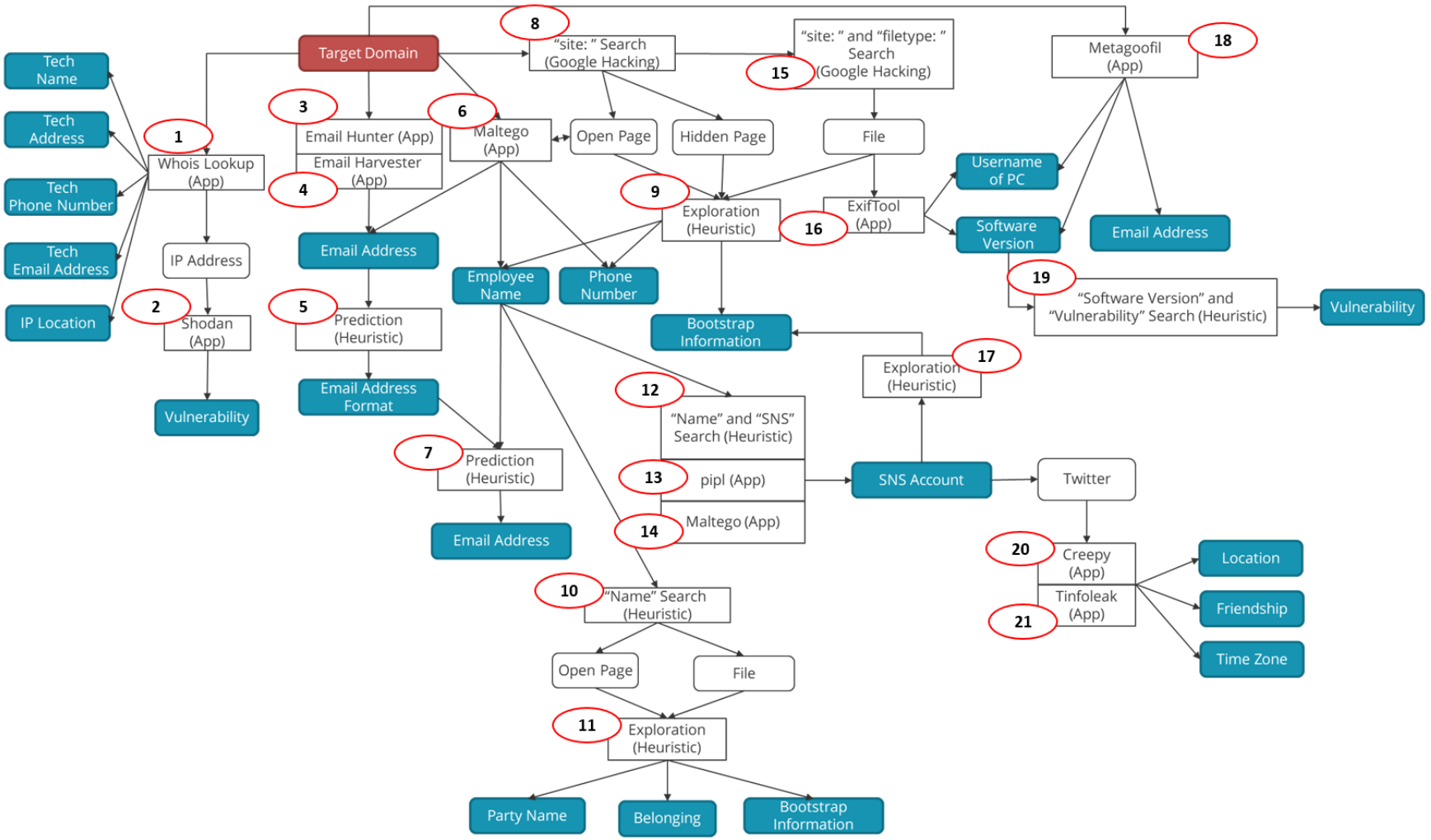

図 7　著者らが作成した OSINT フローチャート

表 1　OSINT 攻撃の整理

No	処理名（図記載名称）	処理区分	属性	入力	出力	対策（機械が情報を抽出することを難しくするもの）	対策（攻撃者が，抽出された情報が正しいかを判断することを難しくするもの）	補足
1	Whois Lookup	位置情報収集	App（機械）	Target Domain	Tech Name, Tech Address, Tech Phone Number, Tech Email Address, IP Location	なし	なし	
2	Shodan	脆弱性調査	App（機械）	IP Address	Vulnerability	パッチ適用	わざと脆弱性のある端末を残してハニーポットにする	
3	Email Hunter	メアド調査	App（機械）	Target Domain	Email Address	・メアドを画像に置き換え ・アットマークを(at)などに置き換え	・嘘の Hidden Page に，嘘のメアドを用意	指定した企業のドメインを指定することで，指定ドメインに存在する社員のアドレスをリストアップする
4	Email Harvester	メアド調査	App（機械）	Target Domain	Email Address	・メアドを画像に置き換え ・アットマークを(at)などに置き換え	・嘘の Hidden Page に，嘘のメアドを用意	恐らく，Web ページを get して，そこから正規表現でメールアドレスを抽出している． 恐らく，myparser クラス

								の email メソッドで処理している
5	Prediction	メアドフォーマット推定	Heuristic（人間）	Email Address	Email Address Format	N/A	・ドメイン名にランダムな文字列を含める ・メールアドレス名で，従業員名を使わない ・とにかく規則性をもたせない	ドメイン名を見ると，組織で通用するメールアドレスのフォーマットが推定できる
6	Maltego	組織情報収集	App（機械）	Target Domain	Email Address, Employee Name, Phone Number	・メアドを画像に置き換え ・アットマークを(at)などに置き換え	・嘘の Hidden Page に，嘘のメアドを用意	次のデータを収集し利用する：DNS records, whois records, search engines, social networks, various online APIs and extracting meta data
7	Prediction	メアド推定	Heuristic（人間）	Employee Name, Email Address Format	Email Address	N/A	・ドメイン名にランダムな文字列を含める ・メールアドレス名で，従業員名を使わない ・とにかく規則性をもたせない	メールフォーマットを基に，他の情報（従業員名等）を基にして，標的者のメールアドレスを推定する

8	"site:"Search	Google Hacking	Heuristic（人間）	Target Domain	Open Page, Hidden Page	N/A	・嘘の Hidden Page を用意し，ハニーポットにする	the site: operator narrows a search to a particular site, domain or subdomain.
9	Exploration	組織情報収集	Heuristic（人間）	Open Page, Hidden Page, File	Employee Name, Phone Number, Bootstrap Information	N/A	Hidden Page や File に対し，偽情報を混ぜ込む	公開されているページ等から情報を収集する
10	"Name" Search	個人情報収集	Heuristic（人間）	Employee Name	Open Page, File	N/A	正規の従業員に紐づく，嘘の情報が入ったファイルを登録する	従業員名を基に，情報収集を行う
11	Exploration	個人情報推定	Heuristic（人間）	Open Page, File	Party Name, Belonging, Bootstrap Information	N/A	ファイル名に，嘘の所属を記載	
12	"Name" and "SNS" Search	SNS アカウント推定	Heuristic（人間）	Employee Name	SNS Account	N/A	同姓同名のアカウントを作成し，嘘の情報を登録．そちらの情報を用いてきたら，攻撃と判断．（SNS によっては，ポリシー違反になる可能性があるため注意が必要）	

13	pipl	SNS アカウント推定	App（機械）	Employee Name	SNS Account	（アルゴリズムの詳細が不明のため，対策も困難）	（アルゴリズムの詳細が不明のため，対策も困難）	アルゴリズムの詳細は不明だが，ダークウェブも検索対象としているようだ.
14	Maltego	SNS アカウント推定	App（機械）	Employee Name	SNS Account	・SNS の使い方を制限		
15	"site:" and "filetype:" Search	Google Hacking	Heuristic（人間）	Target Domain	File	N/A	通常は利用していない，偽のファイルをアップロードしておく	
16	ExifTool	ファイル情報利用	App（機械）	File	Username of PC, Software Version	ファイルから情報を削除する	・ファイルに嘘の情報を混ぜ込む ・嘘のファイルを置いておく	
17	Exploration	SNS 関連情報収集	Heuristic（人間）	SNS Account	Bootstrap Information	N/A	・SNS の使い方を制限	
18	Metagoofil	メアド，ソフトウェア情報収集	App（機械）	Target Domain	Username of PC, Software Version, Email Address	ファイルから情報を削除する	ファイル名に，嘘の所属を記載	gooogle 上で公開されているファイルのメタ情報を抽出する

19	"Software Version" and "Vulnerability" Search	脆弱性調査	Heuristic（人間）	Software Version	Vulnerability	N/A	ファイル名に，嘘の所属を記載	
20	Creepy	Twitter 解析	App（機械）	Twitter Account	Location, Friendship, Time Zone	画像からメタ情報を削除する	画像のメタ情報に嘘の情報を入れる．例えば，写真アプリで，メタ情報には偽データを入れるなど.	twitter や instagram から，ユーザの位置情報を取得する python ツール
21	Tinfoleak	Twitter 解析	App（機械）	Twitter Account	Location, Friendship, Time Zone		デバイス情報や写真のメタ情報に，嘘の情報を入れる	

3.3. 状態遷移モデルによる脅威分析による擬態精度への対策

OSINT ツールを利用した標的型メール攻撃の検討を行うにあたり，本稿では「OSINT ツールと標的型メール攻撃の相乗効果」のモデル化を試みる．まず，攻撃者が OSINT ツールを用いて攻撃対象の情報を収集していく過程を状態遷移モデルとして体系化する．そして，その各状態において攻撃者が生成可能な標的型メールを類型化する．

本研究の第一歩となる本稿においては，標的型メール攻撃の攻撃対象が「特定の個人」である場合に焦点を当ててモデル化を行う．「特定の部署に所属する任意の構成員」を攻撃対象とする標的型メール攻撃も同手順でモデル化を行うことが可能であると考えているが，詳細については今後の検討項目とする．

3.3.1. 攻撃者が保有する情報の遷移

OSINT ツールを利用した標的型メール攻撃では，攻撃者が

① その時点で攻撃者が有する攻撃対象者に関する情報を OSINT ツールに入力して，新たな情報を収集する．

② ①で入手した情報を更に OSINT ツールに入力して，連鎖的に情報を収集する．

という手順を経ることに鑑みて，攻撃者が OSINT ツールを用いて攻撃対象の情報を収集していく過程を状態遷移モデルとして体系化する．

初期状態において，攻撃者が保有する「攻撃対象者に関わる情報」を{X0}と記す．攻撃者が OSINT ツールに{X0}を入力することによって，攻撃対象者に関わる新たな情報 X1 が入手できた場合，状態は（{X0}から）{X0, X1}に遷移する．引き続き，攻撃者が OSINT ツールに{X0, X1}を入力することによって，攻撃対象者に関わる更に新たな情報 X2 が入手できた場合，状態は（{X0, X1}から）{X0, X1, X2}に遷移する．以降，これが繰り返される．

例えば，攻撃者が攻撃対象の｛名前，電話番号｝という情報を保有している状態において，OSINT ツールに「名前」，「電話番号」，あるいはその両方を入力することによって，攻撃対象の「住所」が入手できた場合，攻撃者の状態は｛名前，電話番号，住所｝という情報を保有している状態に変化する．

攻撃者は，任意の OSINT ツールを自由に使用して，上述の手順 1～2 の OSINT 活動を繰り返していく．今回は，攻撃者が使用する OSINT ツールとして，OSINT 収集用 Linux 仮想マシンである Buscador[44]に標準で用意されている OSINT ツール群（Recon-NG, Maltego, Creepy, Metagoofil, Tinfoleak, EmailHarvester, theHarvester, SpiderFoot, ExifTool の 9 つの OSINT ツール）を想定した．

3.3.2. 状態遷移モデルの構築

著者らが攻撃者の立場になり，前節に示した 9 種の OSINT ツールを実際に利用して OSINT 活動を試行することによって，「攻撃者が，OSINT ツールを用いて，攻撃対象者に

関する情報を次々と取得していく過程」を状態遷移モデルとして体系化した（図 8）.

なお，今回構築する状態遷移モデルにおいては，モデルの簡素化のために，以下の前提を設けている.

- 取り扱う OSINT データの種類を，名前，電話番号，メールアドレス，SNS アカウント，趣味，位置情報，友人関係，所属組織名，組織における役職者の名前，組織の公開情報（事業，関連企業等）に限定する.
- 攻撃者が攻撃対象の｛名前｝のみを保有している状態を初期状態と仮定して状態遷移モデルを構築する.
- 各保有情報から推測されうる情報は考慮しない.（例えば，攻撃対象のメールアドレスのドメインが ac.jp であることが分かった時点で，対象者が教育機関に属することが容易に推測できるが，今回は，そのような演繹は一切行わないこととする.）
- OSINT ツールに SNS アカウントを入力すると，そのアカウントの人物の趣味，位置情報，友人関係に関する情報については必ず取得できるとする（図 8 中の※1）.（実際には，例えば Twitter でツイートに位置情報を含める者がいる一方，全く含めない者もいるが，今回は簡単化のため，SNS から概ねこれら 3 つの情報が分かるという前提を置く.）
- 現在多くの組織が Web ページにおいて IR 情報等を公開していることから，OSINT ツールに所属組織を入力すると，その組織の役職者の名前，事業内容，グループ企業に関する情報については必ず取得できるとする（図 8 中の※2）.

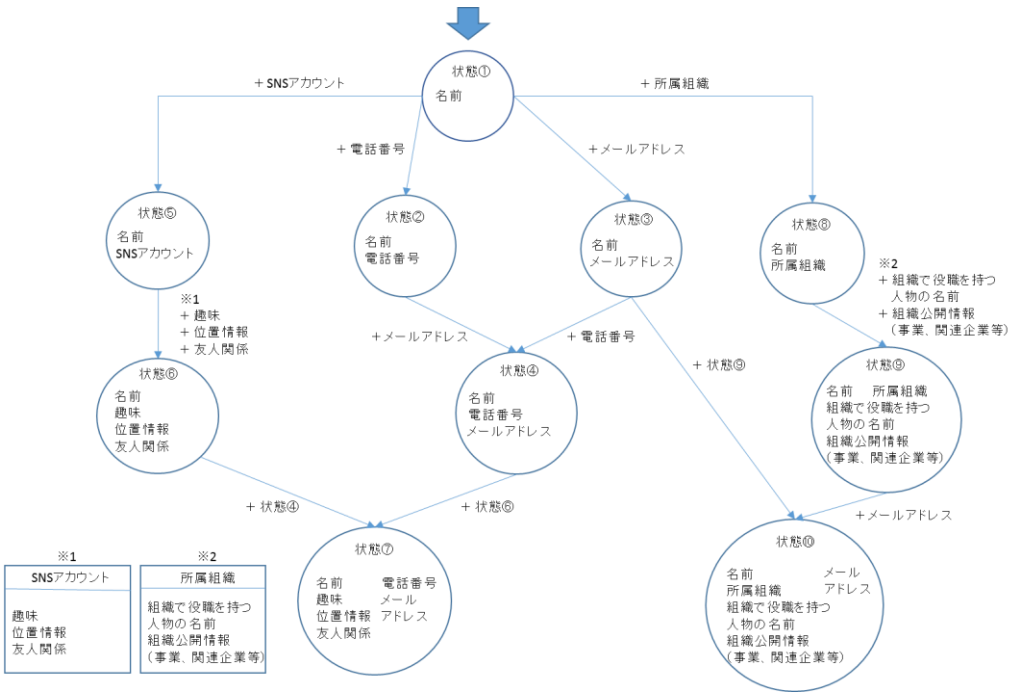

図 8　攻撃者が情報を収集していく家庭の状態遷移モデル

- 今回は標的型メール攻撃（電子メールを介した標的型攻撃）を分析対象とするため，郵便に関する情報である「住所」については，取り扱う OSINT データから除外する*1.

3.3.3. 各状態において攻撃者が生成可能な標的型メール

図 8 の状態遷移モデルの中で，メールの送信が可能な 4 つの状態（状態③，④，⑦，⑩）において，攻撃者が作成可能な標的型メールの典型例をそれぞれ図 9～図 13 に示す.

- 状態③（図 9）:

状態③では，攻撃者が保有している攻撃対象に関する情報が「メールアドレス」と「名前」だけなので，メールの文面は，攻撃対象者に依らない内容になる．メールの文面に名前が記されているため標的型メールと言えるが，内容的には一般的なフィッシングメールに近い.

- 状態④（図 10）:

状態④のでは，攻撃者はメールの文面の中に，攻撃対象者の「電話番号」を含めることできる．身に覚えがないメールであっても，メールの本文に記載されている情報が確かに自分のものであるので，攻撃対象者が騙される（メールに記されているリンク先にアクセスする）可能性は高まると考えられる．また，メールの送信者に自分の電話番号が知られているという空恐ろしさは，攻撃対象者の不安を煽り，冷静な判断を失わせる可能性もある.

なお，今回は，攻撃対象者の「住所」については検討の対象外としたが，もし攻撃者が攻撃対象者の（電話番号の代わりに）住所を取得できた場合には,図 11 のような標的型メールが作成可能である.

- 状態⑦（図 12）:

状態⑦では，攻撃者は，攻撃対象者の SNS アカウント名から，本人の趣味，位置情報，友人関係に関する情報を取得できているので，攻撃対象者の友人になりすました標的型メールを作成することが可能である．この状態では，攻撃対象者のプライベートな情報をメールに盛り込むことが可能であり，信憑度の高い標的型メールを作成できる．また，親密な他者を偽って送信されるメールには，受信者を盲目的にする効果があることが知られているため，脅威の度合いは高いと考えられる.

- 状態⑩（図 13）:

状態⑩では，攻撃者は，2015 年 5 月に日本年金機構が起こした大規模な情報流出事件のきかっけとなった標的型メールを作成できる．この事件の際には，メールの件名や本文に，日本年金機構が Web ページで公開していた事業内容や役職者名を記載することによって，信憑度の高い標的型メールが作成され，使用された.

*1 今回，著者らが実際に OSINT 活動を試行したところでは, OSINT ツールによって住所が取得できたケースは稀（設定ミスによって顧客管理データベースが公開状態にある等の場合のみ）であった．このため，今回の分析において住所を除外することは，住所が「OSINT ツールによって容易に取得できる情報ではない」という意味からも妥当であると言えるかもしれない.

差出人	Amazon_customer_support@gmail.com
件名	Amazon.co.jpのアカウントの修正
宛先	*******@yahoo.co.jp メールアドレス
本文	○○○○様 名前 Amazon.co.jp をご利用いただき、ありがとうございます。お客様のリクエストに沿って、パスワードを再設定いたしましたのでお知らせします。 設定を変更したお心当たりがない場合は、こちら（悪性URL）までお問い合わせ下さい。 Amazon.co.jpのまたのご利用をお待ちしております。 このEメールアドレスは配信専用です。このメッセージに返信しないようお願いいたします。

図 9　状態③の標的型メール

差出人	*******@example.com
件名	電話番号登録確認のご連絡
宛先	*******@yahoo.co.jp メールアドレス
本文	○○○○様 弊社サービスをご利用頂きありがとうございます。 お客様の電話番号のご登録が確認されましたのでお知らせします。 お客様氏名：○○○○　様 お客様電話番号：090-1234-5678 電話番号 なお、詳細を確認する場合は、こちら（悪性URL）をご覧下さい。 このEメールアドレスは配信専用です。このメッセージに返信しないようお願いいたします。

名前

図 10　状態④の標的型メール

差出人	*******@example.com
件名	発送準備完了のお知らせ
宛先	*******@yahoo.co.jp メールアドレス
本文	○○○○様 ××ショップをご利用いただき、まことにありがとうございます。 ご注文いただきました商品の発送準備が整いましたのでご確認ください。 【送り主】 住所 [送り主住所] 4328011　静岡県浜松市中区（省略） [送り主氏名] ○○　○○　さま 【お届け先】（省略） お心当たりがない場合は、こちら（悪性URL）をご確認下さい。

名前

図 11　攻撃者が｛名前，メールアドレス，住所｝を取得した状態の標的型メール

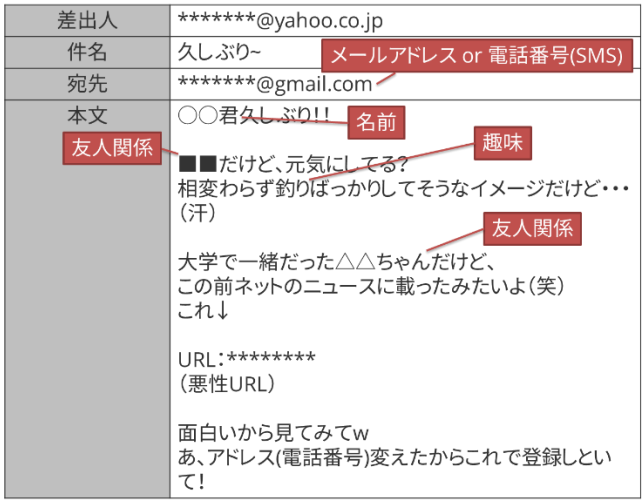

図 12　状態⑦の標的型メール

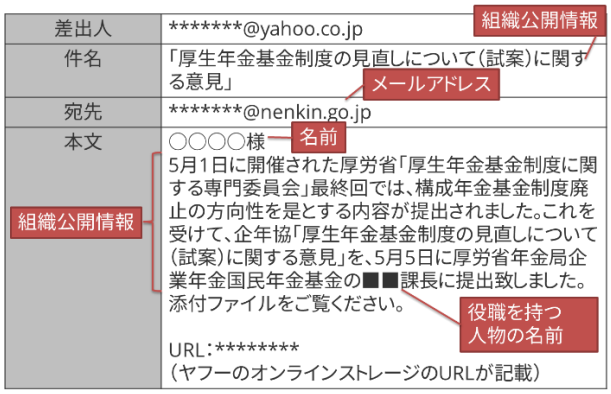

図 13　状態⑩の標的型メール

3.3.4. 標的型メール防御に向けての OSINT 状態遷移モデルの活用

3.3.4.1. 各状態における標的型メールのタイプ

3.3.1 で述べたように，本稿では標的型メール攻撃の攻撃対象を「特定の個人」に絞ったが，ここでは，「特定の個人」を更に 3 つのカテゴリに分類する（表 2）．まず，攻撃対象が「特定の組織に所属する個人」であるか否かによって，「特定の個人」を 2 つに分ける．そして，攻撃対象が一意に特定されるか否か（攻撃対象となり得る人物の数 k が 1 であるか 2 以上か）によって，両者を更に 2 つに分ける．表 2 に示したように，この内の 3 つのタイプを「特定個人型」，「絨毯爆撃型」，「特定構成員型」と呼称することとする．

例えば，状態①の攻撃者が，OSINT 活動によって「攻撃対象者の名前に関連する Gmail アドレス」を取得することができたとしても，それが同性同名の別の人物のメールアドレスである可能性も存在する．つまり，状態③の状態では，攻撃対象者を一意に特定できるとは言えない（図 14）．このような状況は，攻撃者の目的が「複数の同姓同名の人物の内のいずれかを欺くこと」である場合などに生じ得る．すなわち，状態③における標的型メールは，

複数の同姓同名の攻撃対象者に同一文面で送られる（絨毯爆撃型）.

これに対し，電話番号や所属組織は特定の個人を識別するためのユニークな情報であると言える（図 15）[*2]．したがって，状態④や状態⑧における標的型メールは，攻撃者が「特定の人物を欺くこと」あるいは「特定の組織に所属する特定の人物を欺くこと」を目的として，その攻撃対象者のみを狙って送信される形となる（特定個人型，特定構成員型）.

各状態における標的型メールのタイプの傾向を図 16 に示す.

表 2　標的型メールのタイプ

攻撃対象	攻撃対象者の数 k	
	k = 1	k >= 2
特定の個人	特定個人型	絨毯爆撃型
組織に属する個人	特定構成員型	

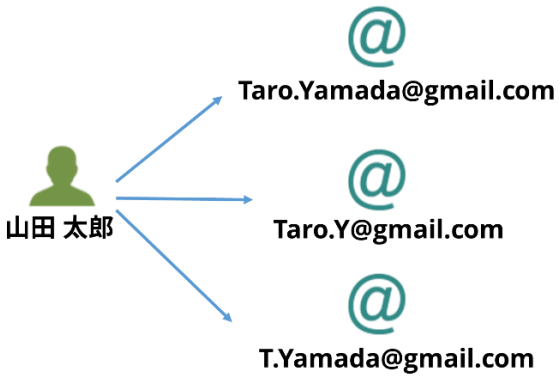

図 14　個人を一意に特定できない状態

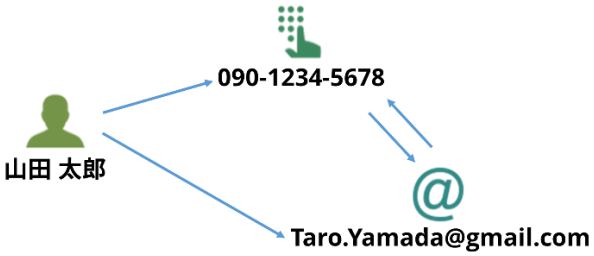

図 15　個人を一意に特定できる状態

*2 構成人員の多い大規模な組織の場合は同姓同名の人物が存在する場合もあるが，ここでは適度な規模の組織を考えることとする.

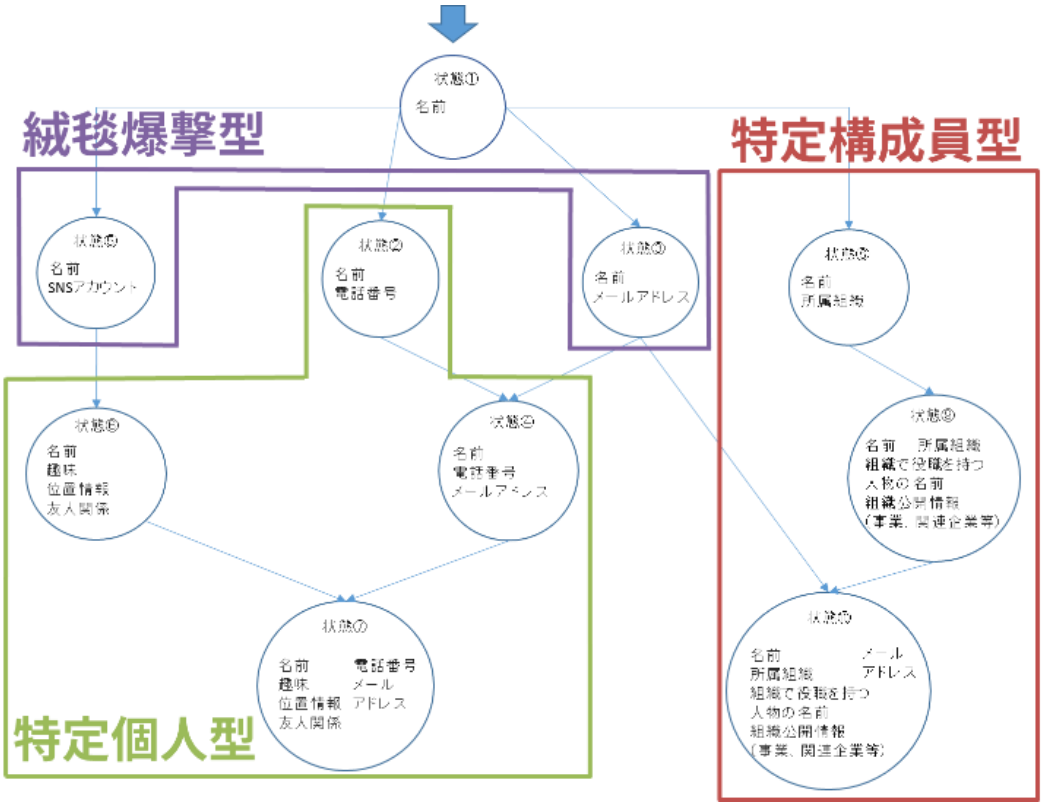

図 16　各状態における標的型メールのタイプ

3.3.4.2. 本モデル及びテンプレートを活用した防御

攻撃者が攻撃対象者の情報を収集していく過程の状態遷移と，各状態で作成される標的型メールの類型化は，以下に述べる 3 つの知見として防御に活用できると期待される．

1 つ目は，標的型メール攻撃の深度の把握である．自分の手元に標的型メールが送られてきた際に，攻撃者が自分に対する情報をどの程度収集できているのか目算できる．また，今後，攻撃者の情報収集が進んだ際にどのような文面の標的型メールが送られてくる可能性があるのか予測できる．これにより，予想される標的型メールに対する注意喚起や，別の攻撃者が似たような標的型メールを送信してくる可能性に気を配ることができる．すなわち，リスク分析や事前対策に役立つと考えられる．

2 つ目は，自分が公開している「自身に関する情報」の中で，どの情報の公開を止めれば，信憑度の高い標的型メールを攻撃者に作成されてしまうことを回避できるかの把握である．図 3 の状態遷移モデルからは，例えば，攻撃者を状態⑦に至らせないためには，自分の名前（実名）と SNS アカウントが共起する形での情報公開を控えることが効果的であることが見て取れる．

3 つ目は，個人や組織の情報公開の度合いに応じた標的型メール対策の導入が可能となることである．OSINT によって得られる情報が多いほど，信憑度の高い標的型メールが送ら

れてくる可能性を考慮し，標的型メール攻撃に対する対策を高める必要がある．

3.4. 3 章のまとめ

本章では，OSINT によって擬態精度を高める攻撃方法の脅威を確かめるために，OSINT ツールを利用することで目的とする情報が収集可能なフローチャートを作成した．

また，攻撃者が OSINT ツールを用いて攻撃対象の情報を収集していく過程を状態遷移モデルとして体系化し，その各状態において攻撃者が生成可能な標的型メールの類型化を行った．

得られた知見は，攻撃の深度の把握，情報を公開することでどのような標的型メールを受ける可能性があるかの想定，必要なセキュリティ対策の選定といったリスク分析や事前・事後対策に活用できると期待される．

今回は主に，「特定の個人を標的とした標的型メール攻撃」を中心に分析を行ったが，今後「組織を標的とした標的型メール攻撃」についても調査していく．また，今回は OSINT ツールで機械的に収集できる情報しか考慮に入れていなかったが，それらの情報から推測され得る情報や，ユーザが意図せず公開してしまっている情報（公開ファイルのメタデータなど）についての議論も必要であると考えられる．今後は，それらの情報も考慮に入れ，状態遷移モデルを改良していく．

第4章　心理操作テクニックと性格特性および行動特性との関係性分析

本章では，将来起こりうる脅威である「攻撃者が，OSINT ツールと AI ツールを駆使し，標的者の性格因子や行動特性を収集して，標的に合った標的型メール文面を作成する」という脅威について，その脅威が現実的であることをアンケート実験により示す．

4.1. OSINT ツールを用いた心理操作効力の高い標的型メール作成

本節では，攻撃者が，OSINT ツールと AI ツールを駆使し，標的者の性格因子や行動特性を収集して，標的型メールの効果を高めるために応用しうる可能性について，関連研究の紹介を交えながら説明する．まず，人の性格因子と心理操作に関して，ビッグファイブとチャルディーニの法則の説明を行う．さらに，ソーシャルエンジニアリングにおける心理操作に関して，チャルディーニの法則とフィッシングメールとの関係を説明する．続いて，ビッグファイブとチャルディーニの法則との関係と，チャルディーニの法則とフィッシングメールとの関係とを基に，性格に基づいて効果的なチャルディーニの法則を標的型メールの文面に反映することができる可能性について説明する．次に，行動特性とチャルディーニの法則（説得されやすさ）との関係性と，セキュリティ（プライバシ）意識の個人差は，性格因子よりも行動特性によって引き起こされることから，標的型メールへの引っかかりやすさは，むしろ行動特性に左右される可能性が示唆されることを説明する．最後に，OSINT ツールによって収集した情報を基に，AI ツールによって性格因子と行動特性を推定することで，攻撃者が標的型メールの効果をより高められることを説明する．

不特定多数に対する詐欺メールであるフィッシングメールに対しては，文献[25][57]などで，チャルディーニの法則との関連について研究が行われている．また，文献[37][38]では，一部のチャルディーニの法則が標的型メールの文面に与える影響を評価している．

Alkış らは，チャルディーニの法則は万人に通用すると述べたうえで，その“影響の受けやすさの度合い（反応度）”は性格因子に左右されるとし，検証実験を行っている[26]．したがって攻撃者は，標的者の性格因子（ビッグファイブ）を取得できれば，その標的者に対して効果的なチャルディーニの法則を知り，その法則に応じた文面をメールに組み入れることによって，心理操作効力を高めた標的型メールを作成することが可能である．

Modic らは，詐欺における説得のされやすさ（詐欺師の説得を受け入れてしまいやすさ）が，個々人の行動特性によって異なることを報告している[27]．したがって攻撃者は，標的者の行動特性を取得できれば，その標的者に対して効果的なチャルディーニの法則を知り，その法則に応じた文面をメールに組み入れることによって，心理操作効力を高めた標的型

メールを作成することが可能である．

一般的に，ある人物の SNS やブログから，正しいビッグファイブを“人間が”推定することは難しいことが知られている[45]．これは，読み手である人間のバイアスがビッグファイブの推定に影響を与えることに起因する．すなわち，読み手によって，その人物の印象や評価が様々であり，ビッグファイブの正確な推定ができないのである．したがって，従前の攻撃者であれば，たとえ標的者の SNS を特定できたとしても，効果的なチャルディーニの法則を選択することが困難であったと言える．

しかし，近年の AI の発展により，SNS やブログからその人物のビッグファイブを至って正確に推定できることが明らかになっている．すなわち，“機械が”推定できるようになってきている．代表的なものに，IBM Watson の Personality Insights が挙げられる[23]．Personality Insights は IBM Watson Developer Cloud により API が公開されており，誰でも利用可能な状況にある．当該 API を用いて著者らが作成したプログラムのシーケンス図とその実行結果を，それぞれ図 17 と図 18 に示す．なお，本実験で実施した OSINT は，サーバに負荷をかけないよう十分に配慮して実行されたことに注意されたい．また，図 18 において，入力に使用した TwitterID は，プライバシ保護の観点から黒塗りにしている．

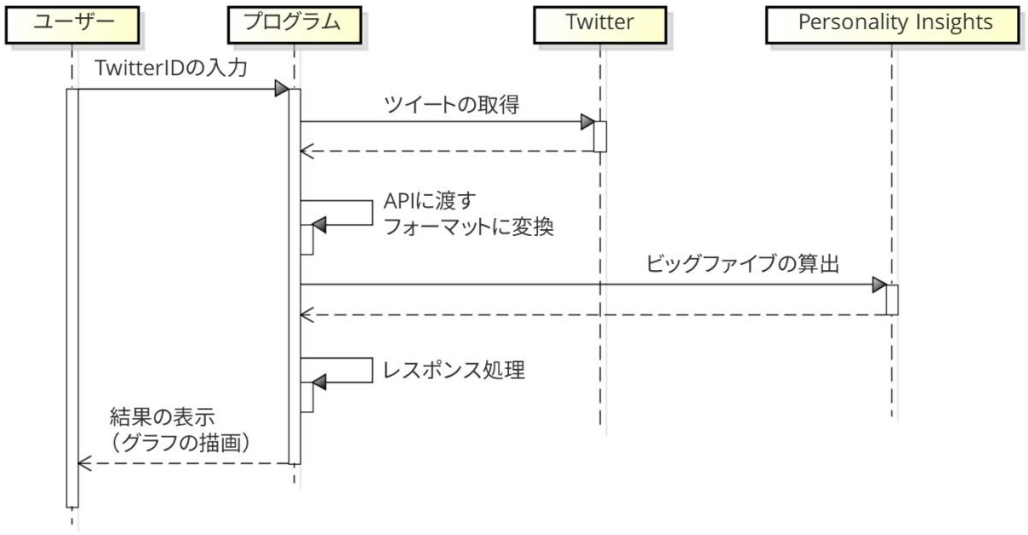

図 17　ビッグファイブ算出プログラムのシーケンス図

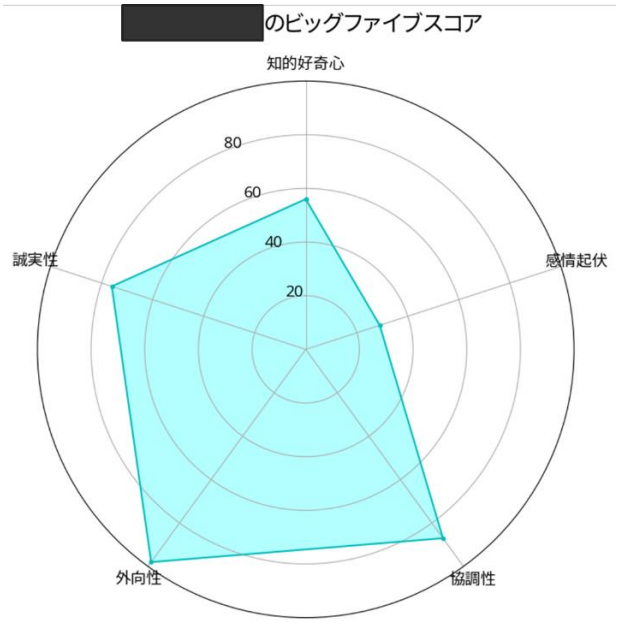

図 18　プログラムの実行結果[23]

個人の性格因子に応じて有効なチャルディーニの法則が異なることも報告されている一方で, Egelman らは, ビッグファイブは個人のプライバシに対する趣向を説明する因子としては弱く, 意思決定の傾向と個人のプライバシに対する趣向との間に, 強い相関が存在していることを示している[58]. この結果は, 標的型メールに対する反応度も, 個々人の性格因子以上に, 行動特性によって左右される可能性があることを示唆している.

ここで, Personality Insights と同様のアルゴリズムで, Modic らにより開発された上述の StP-II 検査[27]の結果と, 各個人のツイートやブログ記事を機械学習することで, 当該人物の「詐欺に対する説得のされやすさ」に関する行動特性についても OSINT により推定することが可能になると考えられる. すなわち, 攻撃者は, OSINT ツールにより収集した情報を, 説得のされやすさを推定する AI ツールに入力することで, 詐欺に対する説得のされやすさについても推定することが可能となる. このように, 攻撃者は, 標的者に StP-II 検査を受診させることなく, OSINT ツールで得られた標的者のツイートやブログ記事を基に, 標的者の行動特性を推定可能であることが示唆される.

本節の内容をふまえると, 「攻撃者が, OSINT ツールと AI ツールを駆使し, 標的者の性格因子や行動特性を収集して, 標的型メールの効果を高めるために応用するという新たな攻撃」が現実の脅威となりうることが分かる.

本論文では, チャルディーニの法則を組み込んだ標的型メールを開くかどうかという行動に関与する行動特性として, 「そもそも, 通常のメール（チャルディーニの法則が用いられていない標的型メール）に対しても, それを開いてしまう傾向を持つ標的者であるのか否か」という観点の行動特性に対して特に焦点を当てる.

4.2. リサーチクエスチョン

攻撃者は OSINT ツールと AI ツールを活用することで, インターネットに公開されている情報をもとに, 標的者の性格因子と行動特性を推定することが可能となる. 今後, インターネットと AI の進歩により, OSINT ツールで収集可能な情報が増殖することで, ソーシャルエンジニアリングの脅威がますます深刻化することを暗示している.

そこで本章では, 個々人の性格因子や行動特性が, OSINT ツールと AI ツールにより収集することが可能となることを前提として,「標的型メールの文面に用いられたチャルディーニの法則」と「個人の性格因子および行動特性」との間の関係性を, ユーザ実験により調査する.

ここで,フィッシングメールとチャルディーニの法則には関係性があることは示されており, また標的型メールにおいても, その一部の法則については関係性が示されている. しかし, 標的型メールにおいて, 6 つのチャルディーニの法則全てに対する関係性は検証されていない. そのうえで,チャルディーニの法則に対する反応度は, 個々人の性格因子と行動特性の両者に影響されることが予想される. すなわち, 標的型メールの開封率も個々人の性格因子と行動特性の両者に影響されると考えられる.

以上より, 本論文のリサーチクエスチョンとして, 以下を掲げる.

RQ1：チャルディーニの法則を標的型メールの文面作成に応用することで, 標的型メールが開封されやすくなるのか.

RQ2：個々人で有効なチャルディーニの法則は, 性格因子と行動特性の両者によって異なる傾向を示すのか.

4.3. ユーザ実験

本章では, 4.2 で掲げた 2 つのリサーチクエスチョンに答えるために, クラウドソーシングを通じて募集した実験協力者 100 名に対する性格検査と疑似標的型メールに対する反応度調査を用いたユーザ実験を行い, その結果を分析する. 本節では, ユーザ実験の手順を説明する.

4.3.1. 実験・分析の流れ

本節では, 今回実施するユーザ実験（A～D）および分析（E, F）の流れを説明する.

A) 名字の入力：

実験協力者に自分の名字を入力してもらう. 標的型メールでは, 標的者の名前を文面に用いることで, 標的者に正規のメールであると信じさせる手口が用いられている. そのため, 実験で利用する疑似標的型メールの文面を実際の攻撃に似せるため, 名字をメールの宛名に埋め込んだ. 名字は, 疑似標的型メールを作成するためだけに利用

し，記録しない．

B) 属性情報の調査：
実験協力者に，性別，年齢層，職種，業務，IT 利用形態などの情報を入力してもらう．

C) 性格検査による性格因子（ビッグファイブ）の調査：
実験協力者に対し，4.3.3 に示す性格検査を実施する．

D) 疑似標的型メールを用いたチャルディーニの効果に対する反応度調査：
実験協力者に 4.3.5 に示す反応度調査を実施する．

E) RQ1 に対する分析：
全実験協力者に実施した(A)から(D)の実験で得られたデータに対し，ウィルコクソン符号付き順位和検定を用いて 4.4 節に示す分析を実施する．本分析では，実験協力者の集合全体に対して分析を行う．

F) RQ2 に対する分析：
全実験協力者に実施した(A)から(D)の実験で得られたデータに対し，ウィルコクソンの順位和検定とスピアマンの順位相関係数を用いて 4.5 節に示す分析を実施する．本分析の一部では，実験協力者の集合を二つの群に分けて分析を行う．

4.3.2. 実験協力者

本ユーザ実験では，クラウドソーシングのランサーズ[63]を利用して 100 名の実験協力者を集め，実験を行った．実験協力者への支払いは，1 名あたり税込み 324 円とした．
実験協力者数を決定した手順は以下のとおりである．実験の実施に先立ち，検出力分析を行った．検出力，有意水準，効果量の 3 つを次のように定めることで，十分なサンプルサイズを決定した．有意水準と検出力は，文献[59]などで適正値として示唆されている 5%, 0.8 とした．効果量を大として，文献[59]をもとに各検定に対する効果量の値（ウィルコクソン検定：0.8，相関分析：0.5）を決定した．その理由は,4.1 で述べた既存研究より，チャルディーニの法則が標的型メールの開封率を上げる効果や，標的型メールの開封率における性格因子および行動特性とのチャルディーニの相関は大きいと考えたためである．その結果，ウィルコクソン符号付き順位和検定（片側検定）のサンプルサイズは 12 名，ウィルコクソンの順位和検定（両側検定）では 27 名，スピアマンの順位相関係数（両側検定）では 29 名となる．(F)の実験では，2 つの実験協力者群が必要となるため，29 名の 2 倍である 58 名がサンプルサイズとして必要な人数となる．外れ値の除外によりサンプルサイズの実数が減少することに備え，実験協力者の人数を 100 名に設定した．

実験協力者は，あらかじめ以下の(ア)～(オ)に関する説明を受け，同意をした上で，ユーザ実験に参加している．

(ア) 本ユーザ実験で収集した情報は，個人が特定できない状態に加工したうえ，学術目的で利用すること．

(イ) 名字の入力を求めるが，名字は実験中の質問時にのみ利用し，記録しないこと．

(ウ) ユーザ実験の質問に最後まで回答していない場合には報酬を受け取ることができないこと.

(エ) 「問題文をしっかり読んでない回答」と実験後に実験実施者により判断された場合には報酬を受け取ることができないこと.

(オ) 会社員を対象とした実験のため, 会社員以外は回答しないこと.

(オ)の条件を追加した理由は, 標的型メールは特定の企業など組織をターゲットとして送信されるため, 会社員の経験があるほうが, 今回の実験趣旨にあった回答が得られると考えたためである. ただし, Web アンケートの性質上, 参加者の正確な職業を知ることはできない. そのため本ユーザ実験の結果には, 会社員以外の参加者の回答も含まれている可能性がある.

4.3.3. 性格検査

和田は, 性格因子（ビッグファイブ）を測定するための日本語版の性格検査を開発している[60]. 60 個の日本語の形容詞に対し, 自身があてはまるかどうかを回答者に 5 件法（「1：あてはまらない」, 「2：あまりあてはまらない」, 「3：どちらともいえない」, 「4：ややあてはまる」, 「5：あてはまる」）で答えさせることで, 個人の性格因子を測定する.

本実験では, 実験協力者の負担を考慮し, 並川らの研究を参考に, 質問項目数を 29 個に絞った短縮版の性格検査を利用して, 各実験協力者の性格因子を測定する[61]. 29 個ある性格検査の質問は, 情緒不安定性 5 問, 外向性 5 問, 開放性 6 問, 調和性 6 問, 誠実性 7 問で構成されている. それぞれ, 正の相関がある質問と, 負の相関がある質問が存在している. 正の相関がある質問に対しては, 「1：あてはまらない」, 「2：あまりあてはまらない」, 「3：どちらともいえない」, 「4：ややあてはまる」, 「5：あてはまる」の回答を, それぞれ 1 点, 2 点, 3 点, 4 点, 5 点と点数化する. 負の相関がある質問に対しては, それぞれの回答を 5 点, 4 点, 3 点, 2 点, 1 点と点数化する.

4.3.4. 行動特性

本研究では, チャルディーニの法則を組み込んだ標的型メールを開くかどうかという行動に関与する行動特性として, 「そもそもチャルディーニの法則を用いていない標的型メールを開いてしまうかどうか」という行動特性に対して特に焦点を当てる. すなわち, 4.5.2 項の実験手順に記載している「プレーンメール（チャルディーニの法則を組み込んでいない標的型メール）への反応度」が高いか低いかによって, 各実験協力者の行動特性が分類される.

4.3.5. チャルディーニの法則の反応度調査

4.3.5.1. 擬似標的型メール作成手順

疑似標的型メールのデータセットを作成する手順を以下に示す.

① 日本サイバー犯罪対策センター(JC3)[62]で公開されている, 実際の標的型攻撃で利

用された標的型メールの文面から無作為に 21 種類のメール（オリジナルメールと呼ぶ）を選択する．

② 21 種類のオリジナルメールをランダムに 3 つの組に分割する（それぞれをオリジナルメールデータセットと呼ぶ）．

③ 3 組のオリジナルメールデータセットから 1 組選択する．

④ オリジナルメールデータセットのオリジナルメールからチャルディーニの法則が利用されていると考えられる部分を削除し*3，チャルディーニの法則が使われていない 7 種類の疑似標的型メール（プレーンメールと呼ぶ）を作成する．

⑤ チャルディーニの各法則（希少性，返報性，権威，一貫性，好意，社会的証明）を誘引するフレーズをメール本文に組み込むことで，プレーンメール 1 つにつき異なる 6 種類のチャルディーニの法則が組み込まれた疑似標的型メール（チャルディーニメールと呼ぶ）を作成する*4．結果的に，プレーンメールを含め，計 49 種類の疑似標的型メールからなる標的型メールデータセットが得られる．

⑥ 残りのオリジナルメールデータセットから 1 つ選び，④，⑤を実施する．

⑦ 3 組全てのオリジナルメールデータセットから標的型メールデータセットを作成した時点で終了する．

ここで，3 組の標的型メールデータセットを i 番目（1≦i≦3）のデータセットと呼ぶことにする．7 種類のプレーンメールを j 番目（1≦j≦7）のプレーンメールと呼ぶことにする．チャルディーニの希少性，返報性，権威，一貫性，好意，社会的証明の各法則に番号を振り，k 番目の法則（1≦k≦6）と呼ぶことにする．また，チャルディーニの法則を組み込まない場合を k=0 で表すことにする．i 番目（1≦i≦3）のデータセットにおける j 番目（1≦j≦7）のプレーンメールに対し，k 番目の法則（0≦k≦6）を組み込むことによって作成した疑似標的メールをe_{jk}^{i}と記載する．3 組の標的型メールデータセットを，e_{jk}^{i}を用いて定義すると，以下のようになる．

$$E^{i} = \begin{pmatrix} e_{10}^{i} & \cdots & e_{16}^{i} \\ \vdots & \ddots & \vdots \\ e_{70}^{i} & \cdots & e_{76}^{i} \end{pmatrix} (i = \{1,2,3\})$$

表 3 に，④で作成したチャルディーニメールの例を示す．紙面の都合上全てを掲載することはできないので，ここではチャルディーニメールの一部を示している．下線部が，④の操作において組み込まれたチャルディーニの各法則を誘引するフレーズである*5．この下線は，チャルディーニの各法則のフレーズを読者に示すためのものであり，実際に実験協力

*3 チャルディーニの法則に該当すると考えられる箇所を実験実施者らで議論し，削除した．削除後のメール文面に実験実施者の 1 人でも違和感を持った場合には，実験実施者全員で再度議論し，メール文面の修正を行っている．削除・修正にあたっては，原文をなるべく尊重するよう配慮した．

*4 疑似標的型メールの文面を実験実施者らで議論し，実験実施者の 1 人でも違和感を持ったメールがあった場合には，実験実施者全員で議論し，文面を修正した．その際，オリジナルメールの文意を変更しないことを前提に，違和感の無い日本語となるよう文面を修正するようにした．

*5 希少性の法則の文面例 1 では，24 時間が経過したからといって添付ファイルが消えるわけではないが，「自由の減少に対する（中略）心理的リアクタンス理論」が希少性の法則の趣旨であるため[24]，文面例 1 も希少性の法則に属する．

者に提示したメールでは表示していない．下線部を削除したものがプレーンメールである．

表 3　チャルディーニメールの例

用いたチャルディーニの法則	文面 1	文面 2
権威の法則	田中様 お世話になります． 解約に関する書類です． 部長より，対応の要請がありましたので， 捺印後に，返信して頂ければ幸いです． よろしくお願いいたします．	田中様 お世話になっております． 注文書を添付します． 貴社の課長より，依頼があった分ですので，添付ファイルをご確認下さい．
希少性の法則	田中様 XLS 版にて送付致します． 24 時間以内に添付ファイルのご確認，宜しくお願いいたします．	田中様 いつもお世話になっております． 添付データの通り，発注をいたしますのでよろしくお願いいたします． ※ファイルは会社規定により本日のみ閲覧可能です． 有限会社全行団
社会的証明の法則	田中様 お世話になっております． 先ほどは，FAX にて申込書類が不鮮明だったためメールにて再度送らせていただきます． 他の皆様と同じように，田中様も， ご確認のほど，どうぞ宜しくお願い致します． 宜しくお願いします．	田中様 お世話になっております． イメージをお送りします． 関係者の内，貴方だけまだ確認していないようでしたので，ご連絡させて頂きました． ご要望含め，ご確認いただけると幸いです． よろしくお願い致します．

4.3.5.2. 実験手順

4.3.5 で作成した標的型メールデータセットを用い，次の手順で反応度調査を実施する．

① 3 組のデータセットE^1, E^2, E^3からランダムに 1 つ選ぶ．

② ①で選んだデータセットから，e^i_{jk}を 7 個選択する．このとき，j を 1 から 7 までの範囲で 1 ずつ増加させながら，また，k は 0 から 6 までの値をランダムかつ重複がないようにして選択する．

③ 7 つのメールすべてをランダムな順序でスクロール可能な 1 ページの画面に表示する．実験協力者は，「メールに書かれている指示に従ってしまう度合い」を「1：絶対に従わない」，「2：従わない」，「3：どちらともいえない」，「4：従う」，「5：絶対に従う」の 5 件法[*6]で，メールごとに回答する．7 つのメール全てに対して回答を行う．

④ 残りのデータセットに対しても①～③を実施する．3 組全てのデータセットに対して回答を終えた時点で反応度調査を終了する．

反応度調査において，同じ文面，同じ法則のメールが 2 度以上出てくると，順序効果（1 回目の文面を見たことが 2 回目の回答に影響を与える）が発生する．これを防ぐために，②で，49 個のe^i_{jk}の中から，j と k に対して排他的に 7 個の疑似標的型メールを選んでいる．

4.4. RQ1 に対する分析

本節では，4.3 の項目 D の実験結果を基に，RQ1 に答えるために実施した分析結果を示す．すなわち，標的型メールに対する反応度調査の結果より，標的型メールにおけるチャルディーニの法則の有効性を検証する．

4.4.1. 回答時間による外れ値除去

実験結果のうち，実験協力者が 4.3 の項目 A～D への回答に要した総回答時間の分布において，極端に大きな値および小さい値を外れ値として除外した．外れ値を求める際には回答時間の四分位数を用いた．第 1 四分位数より四分位範囲の 1.5 倍以上小さい値，あるいは第 3 四分位数より四分位範囲の 1.5 倍以上大きい値を外れ値とした．実験協力者の回答時間の平均は 7 分 40 秒であり，最も短い回答時間は 2 分 52 秒，最も長い回答時間は 22 分 34 秒であった．外れ値を除外した結果，有効な回答数は 94 名となった．なお本ユーザ実験は，Web アンケート形式であり，実験協力者が実験実施者の想定しているとおりに実験を実施しているかどうかを把握することはできないことに注意されたい．

*6 本実験では，チャルディーニの法則によって，標的型メールに対する反応がどのように変化するかを観測しようとしている．その変化の度合いを適度な粒度で分析するために本実験では 5 件法を用いた．

4.4.2. 反応度と相対反応度の定義

4.3.5.2 で示したように，標的型メールへの反応度調査においては，3 つの疑似標的型メールデータセットそれぞれに対し，メールの指示に従うかどうかの回答を実験協力者に質問した．E^1から排他的に選ばれた 7 つの中には，k=0 のメール（チャルディーニの法則を組み込んでいないプレーンメール）が 1 つある．このメールに対する各実験協力者の回答（「1：絶対に従わない」，「2：従わない」，「3：どちらとも言えない」，「4：従う」，「5：絶対に従う」）の値をR_0^1とする．同様に，E^2, E^3に対しても，k=0 のメールに対する回答の値をR_0^2, R_0^3とする．これらR_0^1, R_0^2, R_0^3の算術平均を，当該実験協力者の「プレーンメールに対する反応度S_0」と定義する．同様に，k=1 のメール（プレーンメールにチャルディーニの希少性の法則を組み込んだメール）に対しても，各実験協力者の回答R_1^1, R_1^2, R_1^3の算術平均を「希少性に対する反応度S_1」と定義する．このように，それぞれの k に対して，チャルディーニの各法則に対する反応度S_kを定義する*7.

それぞれの実験協力者ごとに，チャルディーニの各法則の反応度S_kから，プレーンメールの反応度S_0を引いた値を，チャルディーニの各法則に対する相対反応度$S'_k = S_k - S_0$ $(k = \{1,2,3,4,5,6\})$と定義する．4.2 節に示した 2 つの RQ の目的は，チャルディーニの法則を標的型メールの文面作成に応用することで，標的型メールが開封されやすくなるか否かを確認することにある．よって 4.3 項の項目 D に対しては，相対反応度が「チャルディーニの法則を標的型メールの文面作成に応用することで，標的型メールが開封されやすくなるか否かを確認する」にあたっての尺度となる．

4.4.3. 相対反応度による外れ値の除去

実験協力者の相対反応度の分布において極端に大きな値および小さい値を，外れ値として除外した．外れ値を求める際には相対反応度の四分位数を用いた．ある実験協力者の相対反応度が，全実験協力者の相対反応度の分布において，第 1 四分位数より四分位範囲の 1.5 倍以上小さい場合，あるいは第 3 四分位数より四分位範囲の 1.5 倍以上大きい場合を外れ値とした．この時，実験協力者の相対反応度$S'_k(k = \{1,2,3,4,5,6\})$のうち，一つでも外れ値となる場合には，当該実験協力者のデータは全て外れ値であるとした．外れ値を除外した結果，有効なデータは 79 名となった．得られた全データの相対反応度及び属性情報の統計値を表 4，表 5 に示す．以後，79 名のデータセットに対して処理や分析を実施した結果を示す．

*7 リッカート尺度による回答（「1：絶対に従わない」，「2：従わない」，「3：どちらとも言えない」，「4：従う」，「5：絶対に従う」）は間隔尺度ではないが，個々の被験者に同じ趣旨の質問を複数提示し，被験者ごとに回答の算術平均をとる方法であれば，その回答を間隔尺度とみなして分析しても妥当な結果が得られる[64]．そこで今回の実験では，チャルディーニの各法則のメールをそれぞれ 3 通用意し，個々の被験者ごとにその回答R_k^1, R_k^2, R_k^3の算術平均をとることによって，反応度S_kを間隔尺度とみなして以降の分析を行っている．

表 4　全実験協力者の相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	79	1.33	-1.33	0.14	0.52
返報性	79	1.00	-0.67	0.09	0.43
権威	79	1.00	-0.67	0.17	0.47
一貫性	79	1.00	-1.00	-0.02	0.43
好意	79	1.33	-0.67	0.17	0.46
社会的証明	79	1.33	-1.33	0.01	0.50

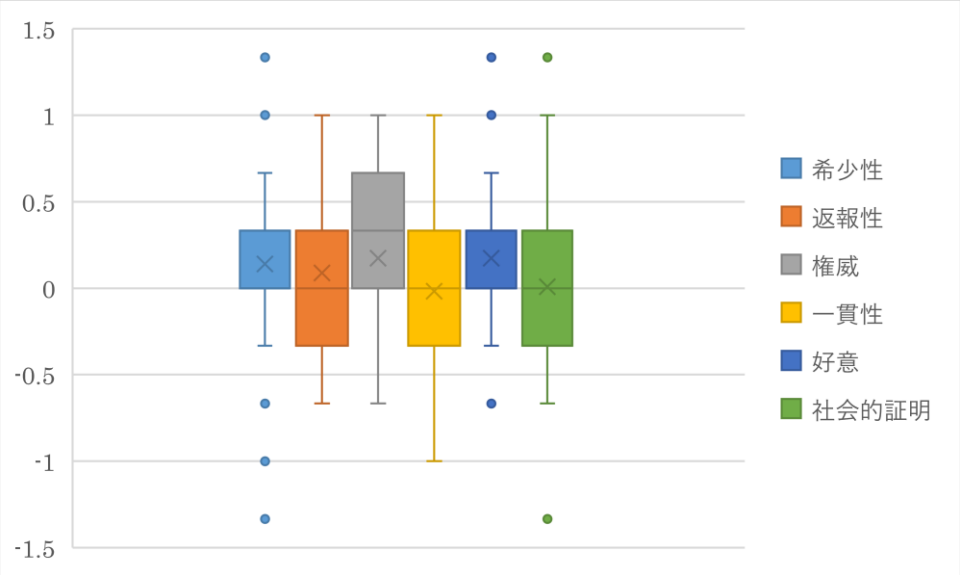

図 19　全実験協力者の相対反応度の箱髭図

表 5　全実験協力者の属性情報

属性種別	属性	人数（人）
性別	男性	49
	女性	30
年齢層	10 代	0
	20 代	14
	30 代	23
	40 代	29
	50 代	11
	60 代	1
	70 代	1
	80 代	0
職種	購買・仕入れ	2
	調査・広告・宣伝	1
	製造・生産・品質管理	2
	経営・事務計画	4
	物流・配送	2
	技術開発・設計	6
	情報処理（システム）	8
	広報・編集	1
	営業・販売	15
	個人事業主・店主	3
	人事・総務・経理	16
	その他	19

4.4.4. 分析結果

RQ1「チャルディーニの法則を標的型メールの文面作成に応用することで，標的型メールが開封されやすくなるのか」に答えるために，チャルディーニの法則が用いられているかどうかで，フィッシングメールと同様に標的型メールでも開きやすくなる傾向があるかを分析した[76][77][78]．分析は，プレーンメールの反応度[*8]とチャルディーニメールの反応度に有意差があるかを，次の帰無仮説と対立仮説により片側検定を実施した．今回，有意水準は5%とした．

帰無仮説：チャルディーニの法則の付与によって標的型メールに対する反応度は変わらない

対立仮説：チャルディーニの法則の付与によって標的型メールに対する反応度は増加する

各実験協力者にプレーンメールと各法則のチャルディーニメールの反応度を聞いているので，二つの反応度は対応したデータ（対標本から得られたデータ）である．このため，ウィルコクソン符号付き順位和検定により検証した．

検定の結果を表 6 に示す．一貫性の法則，社会的証明の法則，返報性の法則に対しては，有意差を確認することができず，帰無仮説を棄却できなかった．希少性の法則，権威の法則，好意の法則では有意差（$p<0.05$）をそれぞれ確認することができ，帰無仮説を棄却することができた．

ただし，希少性の法則，権威の法則，好意の法則における効果量 d は 0.30～0.42 であるため，プレーンメールとチャルディーニメールとの間で反応度の差は小～中[59]であることが分かる．希少性の法則，権威の法則，好意の法則における検出力はいずれも 0.8 以上である．以上より，標的型メールにおいては，一部のチャルディーニの法則が標的型メールを開かせることに対し，有効に働く可能性があることが判明し，RQ1 が部分的に成り立つことを示すことができた．

*8 3.2 節で述べたように，2 つの RQ の目的は，チャルディーニの法則を標的型メールの文面作成に応用することで，標的型メールが開封されやすくなるか否かを確認することにあり，相対反応度を用いてこれを評価することとなる．ウィルコクソン符号付き順位和検定では，反応度を入力として検定を行えば，検定処理の内部で相対反応度が計算される．

表 6　プレーンメールとチャルディーニメール間での反応度の検定結果

	p 値(α)	効果量 d
希少性	0.016*	0.30
返報性	0.101	0.03
権威	0.003*	0.34
一貫性	0.614	0.09
好意	0.001*	0.42
社会的証明	0.425	0.04

*:p<0.05　†：p<0.1

4.5. RQ2 に対する分析

本節では, 4.3 の項目 C と D の実験結果を基に, RQ2 に答えるために実施した分析結果を示す[76][77][78]. すなわち, 性格検査と標的型メールに対する反応度調査の結果より, 性格因子と行動特性がチャルディーニの各法則の効果（チャルディーニの各法則が用いられることによって, 標的型メールを開きやすくなるか）が異なるかを検証する. 検証は, (RQ2-1)性格因子によってチャルディーニの各法則の効果が異なるか, (RQ2-2)行動特性によってチャルディーニの各法則の効果が異なるか, (RQ2-3)性格因子と行動特性の 2 軸によってチャルディーニの各法則の効果が異なるか, の 3 つの観点から分析を実施する.

4.5.1. 性格因子スコアの算出

4.3 の項目 C に対しては, 並川らの性格検査によって性格因子スコアを算出した[61]. 実験協力者ごとに各性格因子に対応する複数の質問に対して, それらへの回答の点数をそれぞれ算術平均し, これを各実験協力者の 5 つの性格因子のスコアとする. 得られた全データの統計値を表 7 に示す.

表 7　性格因子の統計値

	N	最大値	最小値	平均	標準偏差
情緒不安定性	79	5.00	1.80	3.45	0.78
外向性	79	4.40	1.40	2.81	0.77
開放性	79	4.33	1.33	2.94	0.67
調和性	79	4.33	1.17	3.23	0.69
誠実性	79	4.57	2.00	3.23	0.60

4.5.2. 性格因子に関する分析結果

本項では, RQ2-1「性格因子によってチャルディーニの各法則の効果が異なるか」に答えるために, 性格因子スコアとチャルディーニの各法則の相対反応度との間で, スピアマンの順位相関係数を算出した. 実験協力者全体（79 名のデータセット）に対して分析を行った. 有意水準は 5%とした.

算出した相関係数を表 8 に示す. 外向的スコアが高い人間に対しては権威の法則を用いた標的型メールが有意（$p<0.05$）に効果的であり, 誠実性スコアが高い人間に対しては一貫性の法則を用いた標的型メールが有意（$p<0.05$）に効果的であることが分かる（ただし, 検出力はそれぞれ 0.54, 0.75 であるので, 特に前者の外向的スコアと権威の法則の間の相関には相応の第二種の過誤が含まれうる）.

表 8　全データにおける性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	0.11	0.15	-0.09	0.15	0.04
返報性	0.12	0.06	-0.05	0.06	-0.06
権威	0.02	0.23*	0.11	0.06	0.08
一貫性	0.06	-0.03	0.06	-0.05	0.29*
好意	0.06	0.06	-0.05	0.17	0.03
社会的証明	0.15	0.08	-0.02	0.02	0.06

**: p<0.01, *: p<0.05, †: p<0.1

これは, 外向的な人間は, 社会性を有する傾向がある[22]ため, チャルディーニの権威の法則が組み込まれたメールのように, 組織の上下関係を利用したメールに対してはこれを開きやすくなるのだと考えられる. 一方, 誠実性が高い人間は, 計画性を持った行動を好む傾向がある[22]ため, チャルディーニの一貫性の法則が組み込まれたメールのように, 以前までの動作の継続を依頼するメールに対してはこれを開きやすくなるのだと考えられる.

今回の分析では,既存研究[26]で示された性格因子とチャルディーニの各法則との相関関係とは, 異なる結果が得られた. 誠実性スコアが高い人間は一貫性の法則が有意に効果的である点は共通している. 一方, 今回は外向性スコアが高い人間は権威の法則が効果的であるという結果が得られたのに対し, 既存研究では一貫性の法則が効果的であるという結果が得られている. さらに, 今回は有意差が確認できなかったが, 既存研究では調和性スコアが高い人間は, 返報性の法則, 権威の法則, 好意の法則, 一貫性の法則が効果的である結果が得られている. このような差が生じた理由として, 既存研究で実施した実験と今回の実験で, 質問の形態が異なることが考えられる. 既存研究では, チャルディーニの各法則への反応度を直接的に尋ねているのに対し, 今回の実験では, チャルディーニの各法則を組み込んだ

標的型メールの文面に対する反応度を尋ねている. この点については, 今後さらに検討していく必要があるだろう.

以上の結果より, 標的型メールを開かせることに対してチャルディーニのどの法則が有効であるかが, 性格因子に応じて異なることが判明し, RQ2-1 が成り立つことが示された.

4.5.3. 行動特性に基づく実験協力者の分割

RQ2-2, RQ2-3 に答えるための分析を行うにあたり, その準備として, 「標的型メールを開きやすいかどうか」という行動特性（4.3.4 項）に着目して, 実験協力者を 2 つの群に分割した. 今回は, プレーンメール（チャルディーニの法則を組み込んでいない標的型メール）への反応度が 3 以下の実験協力者を「標的型メールを開きにくい群」とし, 3 より大きい実験協力者を「標的型メールを開きやすい群」として分類した. 結果, 開きにくい群の回答数は 26 名, 開きやすい群は 53 名となった. 開きにくい群と開きやすい群の, それぞれの相対反応度及び属性情報の統計値を表 9〜表 12 に示す.

表 9　開きにくい群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	26	1.33	-1.33	0.15	0.64
返報性	26	1.00	-0.33	0.26	0.47
権威	26	1.00	-0.67	0.21	0.45
一貫性	26	1.00	-0.67	0.03	0.43
好意	26	1.33	-0.67	0.23	0.53
社会的証明	26	1.33	-0.67	0.08	0.55

表 10　開きにくい群における属性情報

属性種別	属性	人数（人）
性別	男性	16
	女性	10
年齢層	10 代	0
	20 代	5
	30 代	7
	40 代	9
	50 代	5
	60 代	0
	70 代	0
	80 代	0
職種	購買・仕入れ	1
	調査・広告・宣伝	1
	製造・生産・品質管理	1
	経営・事務計画	1
	物流・配送	1
	技術開発・設計	2
	情報処理（システム）	5
	広報・編集	0
	営業・販売	4
	個人事業主・店主	1
	人事・総務・経理	3
	その他	6

表 11　開きやすい群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	53	1.00	-1.00	0.13	0.46
返報性	53	1.00	-0.67	0.01	0.39
権威	53	1.00	-0.67	0.16	0.48
一貫性	53	0.67	-1.00	-0.04	0.44
好意	53	1.33	-0.67	0.14	0.42
社会的証明	53	1.00	-1.33	-0.03	0.47

表 12　開きやすい群における属性情報

属性種別	属性	人数（人）
性別	男性	33
	女性	20
年齢層	10 代	0
	20 代	9
	30 代	16
	40 代	20
	50 代	6
	60 代	1
	70 代	1
	80 代	0
職種	購買・仕入れ	1
	調査・広告・宣伝	0
	製造・生産・品質管理	1
	経営・事務計画	3
	物流・配送	1
	技術開発・設計	4
	情報処理（システム）	3
	広報・編集	1
	営業・販売	11
	個人事業主・店主	2
	人事・総務・経理	13
	その他	13

4.5.4. 行動特性に関する分析結果

本項では，チャルディーニの各法則への行動特性の影響を分析するために，4.5.3 項で分類した開きにくい群の相対反応度と開きやすい群の相対反応度に有意差があるかを，次の帰無仮説と対立仮説により両側検定を実施した．有意水準は 5%とした．

帰無仮説：開きにくい群と開きやすい群では相対反応度に差がない．

対立仮説：開きにくい群と開きやすい群では相対反応度に差がある．

実験協力者群同士の比較となり，データに対応がないため，ウィルコクソンの順位和検定により検証した．

検定の結果を表 13 に示す．両群での相対反応度の有意差（$p<0.05$）が確認できたのは，返報性の法則のみであった．ただし，効果量 d は 0.58 であるため，開きにくい群と開きやすい群との間で反応度の差は中[59]であることが分かる．検出力は 0.76 である．以上より，「標的型メールを開きやすいかどうか」という行動特性によって，チャルディーニメールに対する相対反応度が異なる法則は一つのみであることが判明し，RQ2-2 については限定的に成り立つという結果になった．

表 13　開きにくい群と開きやすい群間での検定結果

	p 値(α)	効果量 d
希少性	0.706	0.04
返報性	0.049*	0.58
権威	0.886	0.11
一貫性	0.713	0.02
好意	0.534	0.19
社会的証明	0.596	0.10

*:p<0.05　†：p<0.1

4.5.5. 性格因子と行動特性に関する分析結果

本項では，チャルディーニの各法則への性格因子と行動特性による影響を分析するために，行動特性で分割した 2 つの実験協力者群に対し，それぞれ，性格因子スコアとチャルディーニの各法則の相対反応度との間で，スピアマンの順位相関係数を算出した．結果を表 14，表 15 に示す．有意水準は 5%とした．

表 14　開きにくい群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	-0.29	0.34 †	0.33	0.21	0.49*
返報性	0.00	0.12	-0.14	0.12	-0.15
権威	-0.40*	0.52**	0.28	0.29	0.27
一貫性	0.20	0.06	-0.18	-0.12	0.39*
好意	-0.03	0.20	0.19	0.32	0.00
社会的証明	0.04	0.22	0.00	0.08	0.11

**: p<0.01, *: p<0.05, †: p<0.1

表 15　開きやすい群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	0.31*	0.04	-0.29*	0.08	-0.21
返報性	0.14	0.05	0.07	-0.02	-0.03
権威	0.18	0.09	0.02	-0.04	-0.01
一貫性	0.01	-0.06	0.18	-0.01	0.25 †
好意	0.08	-0.02	-0.14	0.10	0.04
社会的証明	0.21	0.01	-0.01	0.00	0.04

**: p<0.01, *: p<0.05, †: p<0.1

相関分析の結果，開きやすい群と開きにくい群で類似の傾向（行動特性の違いが，チャルディーニの各法則に対する性格因子の影響に大きな差を生まない）を示すものと，それぞれの群で異なる傾向（行動特性に応じて，チャルディーニの各法則に対する性格因子の影響に差が生じる）を示すものがあった．

① 開きやすい群と開きにくい群で類似の傾向：
「誠実性スコア」と「一貫性の法則に対する相対反応度」との間に弱い正の相関(開きにくい群：r=0.39, p<0.05, 検出力=0.52, 開きやすい群：r=0.25, p<0.1, 検出力=0.44)がある．

② 開きにくい群のみに見られる傾向：

「情緒不安定性スコア」と「権威の法則に対する相対反応度」との間に弱い負の相関（r=－0.40, p<0.05, 検出力=0.54）がある．「外向性スコア」と「希少性の法則に対する相対反応度」との間に弱い正の相関（r=0.34, p<0.1, 検出力=0.41）がある．「外向性スコア」と「権威の法則に対する相対反応度」との間に正の相関（r=0.52, p<0.01, 検出力=0.81）がある．「誠実性スコア」と「希少性の法則に対する相対反応度」との間に正の相関（r=0.49, p<0.05, 検出力=0.75）がある．

③ 開きやすい群のみに見られる傾向：

「情緒不安定性スコア」と「希少性の法則に対する相対反応度」との間に弱い正の相関（r=0.31, p<0.05, 検出力=0.63）がある．「開放性スコア」と「希少性の法則に対する相対反応度」との間に弱い負の相関（r=－0.29, p<0.05, 検出力=0.57）がある．

①は，開きにくい群と開きやすい群に共通する傾向である．すなわち，人間の行動特性によらず性格因子の影響により現れる傾向であり，このため，4.5.2 項での分析結果が再度示された結果となった（ただし，検出力はそれぞれ 0.52, 0.54 であるので，両群の「誠実性スコア」と「一貫性の法則に対する相対反応度」の相関には相応の第二種の過誤が含まれうる）．

②と③は，それぞれ開きにくい群のみと，開きやすい群のみに見られる傾向である．4.2 節で「チャルディーニの法則に対する反応度は，個々人の性格因子と行動特性の両者に影響されること」と予想したが，今回の結果よりそれが支持される形となった（ただし，「外向性スコア」と「権威の法則に対する相対反応度」の検出力，「誠実性スコア」と「希少性の法則に対する相対反応度」の検出力を除き，検出力は 0.7 を下回るため，それぞれ相応の第二種の過誤が含まれうる）．

以下では，②と③の結果が，それぞれどのような理由で生じたのかを考察する．

②は，開きにくい群にのみ見られる傾向である．すなわち，「プレーンメールを開きにくい」という行動特性を有する人間のみに見られる性格因子の影響である．

一つ目の結果として，外向性スコアが高い人間（すなわち，社会性を有する傾向がある人間[22]）の中でも，プレーンメールを開かない人間だけが，希少性の法則のチャルディーニメールに対する相対反応度が高くなった（プレーンメールを開く人間は，希少性の法則のチャルディーニメールに対する相対反応度は高くならなかった）．すなわち，「外向性スコアが高い」という性格因子を持ち，かつ「プレーンメールを開かない」という行動特性を持つ人間だけに対して，希少性の法則が標的型メールを開かせる後押しをしたという結果が得られた．

ここで，後押しをされたからといって，メールを開いてしまうわけではないことに注意されたい．元々メールを開かない人間（プレーンメールに対する反応度が 1 や 2）は，標的型メールを開かせる後押しをされてもプレーンメールに対する反応度が 2 や 3 になる程度の場合もある．このように，反応度では，チャルディーニの各法則により引き起こされる後押しの効果を正しく評価することができない．相対反応度を用いて分析することで，後押し

の効果を明らかにすることができる．

社会性を有する人間がプレーンメール（チャルディーニの法則が含まれない疑似標的型メール）を開かなかったということは，その人間は「このメールを開かなくても，自分の社会的な立場に対する悪影響がない」と判断したのだと考えられる．しかし，希少性の法則が組み込まれることで焦燥感が煽られ，その判断が崩れて，標的型メールを開く後押しとなったのだろう．反対に，社会性を有する人間がプレーンメールを開くということは，「このメールを開かなければ，自分の社会的な立場に悪影響が生じる」と判断したのだと考えられる．そのため，希少性の法則によって焦燥感が煽られたところで，その判断は崩されず，標的型メールを開く後押しとはならなかったのだろう．このように，社会性を有する人間であっても，プレーンメールを開くか開かないかという行動特性の違いで，希少性の法則により引き起こされる後押しが発生するかが違ってくる．

２つ目の結果として，希少性の法則と同様に，外向性スコアが高い人間（すなわち，社会性を有する傾向がある人間[22]）の中でも，プレーンメールを開かない人間だけが，権威の法則のチャルディーニメールに対する相対反応度が高くなった（プレーンメールを開く人間は，権威の法則のチャルディーニメールに対する相対反応度は高くならなかった）．権威の法則のように自分より上の立場の人間からの指示が組み込まれることで，反応しなければ自分の社会的な立場に悪影響が及ぶと考えたため，標的型メールを開く後押しとなったのだろう．反対に，社会性を有する人間がプレーンメールを開くということは，「このメールを開かなければ，自分の社会的な立場に悪影響が生じる」と判断したのだと考えられる．そのため，権威の法則によって自分より上の立場からの指示が行われても，その判断は崩されず，標的型メールを開く後押しとはならなかったのだろう．

３つ目の結果として，誠実性スコアが高い人間（すなわち，計画性を有する傾向がある人間[22]）の中でも，プレーンメールを開かない人間だけが，希少性の法則のチャルディーニメールに対する相対反応度が高くなった（プレーンメールを開く人間は，希少性の法則のチャルディーニメールに対する相対反応度は高くならなかった）．計画性を有する人間は，自らの計画に基づいて判断する．希少性の法則を誘引するフレーズにより時間的制約を与えられることで，条件が変わり，計画を変更する可能性が生じる．今回の結果からは，「誠実性スコアが高い」という性格因子を持ち，かつ「プレーンメールを開かない」という行動特性を持つ人間は，時間的制約によって判断を変える傾向が強いことが分かった．一方で「誠実性スコアが高い」という性格因子を持ち，かつ「プレーンメールを開く」という行動特性を持つ人間は，時間的制約によって判断を変える傾向は見られなかった．

４つ目の結果として，情緒不安定性スコアが高い人間（すなわち，心配事に弱い傾向がある人間[22]）の中でも，プレーンメールを開かない人間だけが，権威の法則のチャルディーニメールに対する相対反応度が低くなった（プレーンメールを開く人間は，権威の法則のチャルディーニメールに対する相対反応度は低くならなかった）．心配事に弱い人間は，プレッシャーに過剰反応する．権威の法則を誘引するフレーズによって，過剰なプレッシャーを

感じ，標的型メールを開く方向への後押しを受けるはずだと思われるのだが，今回は違う結果が得られた．これは，性格因子や行動特性だけでは判断できないことを示唆しており，第三の要因が存在している可能性を表している．

③は，開きやすい群にのみ見られる傾向である．すなわち，「プレーンメールを開きやすい」という行動特性を有する人間のみに見られる性格因子の影響である．

１つ目の結果として，情緒不安定性スコアが高い人間（すなわち，心配事に弱い人間[22]）の中でも，プレーンメールを開く人間だけが，希少性の法則のチャルディーニメールに対する相対反応度が高くなった（プレーンメールを開かない人間は，希少性の法則のチャルディーニメールに対する相対反応度は高くならなかった）．心配事に弱い人間がプレーンメールを開くということは，その人間は「このメールを開かなければ，自分にとって何か悪影響があるのではないか」と心配に感じたのだと考えられる．ここに，希少性の法則が組み込まれることで焦燥感がさらに煽られ，標的型メールを開く後押しとなったのだろう．反対に，心配事に弱い人間がプレーンメールを開かないということは，「このメールを開かなくても，自分にとって悪影響は起きない」と判断したのだと考えられる．そのため，希少性の法則によって焦燥感を煽られても，その判断は崩されず，標的型メールを開く後押しとはならなかったのだろう．

２つ目の結果として，開放性スコアが高い人間（すなわち，独立心が強い人間[22]）の中でも，プレーンメールを開く人間だけが，希少性の法則のチャルディーニメールに対する相対反応度が低くなった（プレーンメールを開かない人間は，希少性の法則のチャルディーニメールに対する相対反応度は低くならなかった）．独立心が強い人間がプレーンメールを開いたということは，その人間は自分の強い意志で「このメールを開く」と判断したのだと考えられる．このような人間が，希少性の法則を組み込んだメールを見た場合，希少性の法則を誘引するフレーズの中に「メールを開かせよう」という送信者からの意図を感じるのではないだろうか．そのため，自らの意思決定に対して水を差されたと思い，メールを開くという気持ちが萎えたのだろう．反対に，独立心が強い人間がプレーンメールを開かなかったということは，その人間は自分の強い意志で「このメールを開かない」と判断したのだと考えられる．そのため，希少性の法則によって焦燥感を若干煽られても，その意思は崩されず，標的型メールを開く後押しとはならなかったのだろう．

以上の結果より，性格因子と行動特性の 2 軸によってチャルディーニの各法則の効果が異なることが判明し，RQ2-3 が成り立つことが示された．そして，RQ2-1 から RQ2-3 に対する分析結果から，RQ2 全体が成り立つことを示すことができた．ただし，今回の②の４つ目の分析結果からは，性格因子と行動特性以外の第３の要因が存在する可能性も暗示される．

4.6. 4 章のまとめ

ユーザ実験の結果, 標的型メールの文面において, チャルディーニの法則と, 個人の性格因子および行動特性との間には関係性があることが判明した. すなわち攻撃者は, OSINT ツールと AI ツールを用いて, 標的者の性格因子および行動特性に有効な文面を推定可能であると結論付けることができる.

まず, 攻撃者が標的者の名前を知っている場合, OSINT ツールによって, たとえばメールアドレスと所属組織を取得し, 図 20 のような擬態精度を高めた標的型メールを作成することができる. 自分の名前がメールの本文に含まれている, 差出人欄が所属組織のメールアドレスである, などから, 標的者は正規のメールであると感じやすくなる.

さらに攻撃者は, OSINT ツールと AI ツールによって標的者の性格因子や行動特性を調べることができる. 図 20 の標的型メールに対し, 心理操作効力を高めるために, チャルディーニの法則である希少性と権威を使用したメールをそれぞれ図 21 と図 22 に示す. プレーンメールを開きにくい傾向にあり, かつ誠実性スコアが高い人間であれば希少性は有効であるため, 標的者の行動特性と性格因子から希少性が有効であると判断できた場合には, 図 21 のような標的型メールを標的者に送ることによって, 攻撃者はさらに標的型攻撃の効果を高めることができる. プレーンメールを開きにくい傾向にあり, かつ外向性スコアが高い人間であれば権威は有効であるため, 標的者の行動特性と性格因子から権威が有効であると判断できた場合に, 権威を用いることでより効果的な標的型メール文面を作成することができる.

その他の法則についても, その法則の特性を活かして標的型メールを作成することが可能である. 少し文章を追加する程度で十分であることから, 攻撃者の費用対効果は非常に高いと考えられる.

標的型メールに組み込んで効果があるチャルディーニの各法則に影響がある要因として, 性格因子と行動特性以外に, 第 3 の要因が示唆されている.今後はその要因を特定する分析を行いたい.また, 4.3 節に示した反応度調査で利用したチャルディーニメールの文面は, 必ずしも単一のチャルディーニの法則が組み込まれた文面とはなっていない.「24 時間以内」という文言は, 暗に上司からの指示であるととらえることもでき, その場合には権威が含まれていると考えられる.正確にチャルディーニの法則を組み込んだメールを作成することは難しいが, 文面の作成についても再度検討していきたい.さらに, 実験協力者数を増やして追実験を行うことによって, 今回の実験よりも効果量や検出力が高い形での評価を行うと同時に, 実験目的に沿わない回答を除外する Instructional Manipulation Check(IMC)[80] などの手法によって回答の質を高めたうえで, 再現性を確認したい.

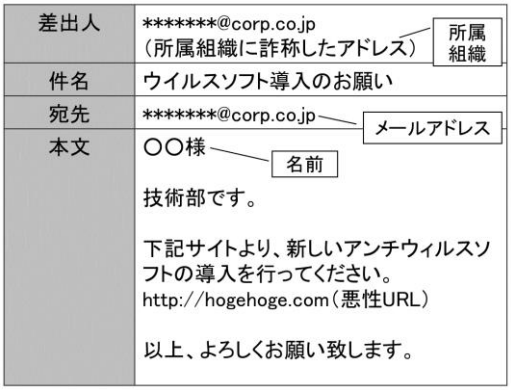

図 20　標的型メール例

差出人	*******@corp.co.jp （所属組織に詐称したアドレス）
件名	ウイルスソフト導入のお願い
宛先	*******@corp.co.jp
本文	〇〇様 技術部です。 下記サイトより、新しいアンチウィルスソフトの導入を行ってください。 http://hogehoge.com（悪性URL） ※ファイルは会社規定により本日のみ閲覧可能です。 以上、よろしくお願い致します。

図 21　希少性を利用した標的型メール例

差出人	*******@corp.co.jp （所属組織に詐称したアドレス）
件名	ウイルスソフト導入のお願い
宛先	*******@corp.co.jp
本文	〇〇様 技術部です。 XX部長より、新しいアンチウィルスソフトの導入の要請がありました。 下記サイトより、ソフトの導入を行ってください。 http://hogehoge.com（悪性URL） 以上、よろしくお願い致します。

図 22　権威を利用した標的型メール例

第5章　　個人に合わせたアラートによる心理操作効力を駆使した攻撃への対策

第 4 章より，個人の性格因子と行動特性に応じて有効なチャルディーニの法則が異なるため，攻撃者が個人に合わせて効果的な標的型メール文面を作成する危険性があることが判明した.

本章では，まず，個人に合わせたアラートシステムの構成を提案する．続いて，提案システムの主要な構成要素である，文面からチャルディーニの法則を検出する手法を示し，その実験結果を合わせて示す．さらに，個人の性格因子や行動特性に応じて表示するアラートを変更することが有効であるか，ユーザ実験により調査した結果を示す.

5.1. 個人に合わせたアラートシステムの構成

本節では，個々人に応じた標的型メール対策として個人に合わせたアラートシステムが有効であると考えられることを示し，そのシステム構成案と，本章で扱う範囲を示す.

メール文面からチャルディーニの法則をヨーロッパの金融機関における Security Operation Center(SOC)で収集したデータを実際のフィッシングメールを基に，フィッシングメールの本文中で用いられているチャルディーニの法則を検知する識別器を作成した研究がある[75]．本研究において作成した識別器は，メールがフィッシングメールかそうではないかを判断するのではなく，そのメール文面が，ユーザにとって引っ掛かりやすいメール本文であるかを推定する識別器である．本研究では，データに対して識別器を適用した結果を基に統計分析をした結果，一つのメール文中で用いられるチャルディーニの法則の種類の数が増えるほど，ユーザに影響を与えやすいことが示唆される結果を得た．本研究結果は，フィッシングドメインのテイクダウンや，ブラックリストの作成，顧客への注意喚起に活用できる．また，メールに用いられているチャルディーニの法則の度合を表示することで，インシデントレスポンスを効率化するといった新しい方法が期待できると述べている.

また，OSINT ツールと AI ツールによって攻撃者の能力が向上することを 4.6 項で示した．これらを踏まえ，攻撃者の能力が向上した場合における防御策としては，防御側も自組織内に所属する人物の性格因子や行動特性を把握することで，個々人に応じた標的型メール対策を実施することが考えられる.

個々人に応じた標的型メール対策として，自分にとって「騙されてしまいやすい」チャルディーニの法則が用いられているメールを受信した際に“肩を叩いてもらう”という方法が有効であると考えられる．これは，説得心理学における脅威アピールの考えに該当する[15]．そのためには，メールを受け取った際に，メール文面からチャルディーニの法則が使われている箇所を特定することと，ユーザの個人特性を踏まえてどのように提示するかを判断す

るシステムが必要となる．しかし，脅威アピールの既存研究においても，脅威の通知に効果がある場合と，そうではない場合があり，本研究で検討しているシステムにおいても，個々人の特性に応じて効果が異なると考えられる．例えば，アラートが必要以上に多い場合，人によってはアラートに慣れてしまい無視することがあり得る．

チャルディーニの法則が使われたメールが必ずしも標的型メールであるとも限らないので，ユーザに提示するアラートを絞ることで誤アラートを減らすことや，真に有効なアラートのみを提示することが必要となる．たとえば，権威の法則が使われたメールである場合に，そのことをアラートで通知することが有効である人物に対しては，アラートを提示し，権威の法則に対するアラートが有効ではない人物に対してはアラートを提示しない方法が効果的であると考えられる（図 23）．具体的には，個々人の性格因子や行動特性に応じて，アラートを挙げる閾値を調整する形となる．ここで，個々人において，4 章で示した「標的型メールにおいて有効なチャルディーニの法則」と，本章で示す「アラートを通知する際に提示すると有効なチャルディーニの法則」とでは，それぞれコンテキストが異なる．そのため，チャルディーニの法則の効果として，異なる結果が得られると考えられる．

ここで，対象者に対して直接チャルディーニの法則の効果を尋ねずに，性格因子や行動特性を媒介とするかを説明する．チャルディーニの法則を直接尋ねると，回答者が，良い解答をしようとするため正しく反応を測定することができない．そのため，性格因子や行動特性を通じて推定できるようにすることで，回答者によるバイアスを除いて，提示する効果を推定することが期待される．

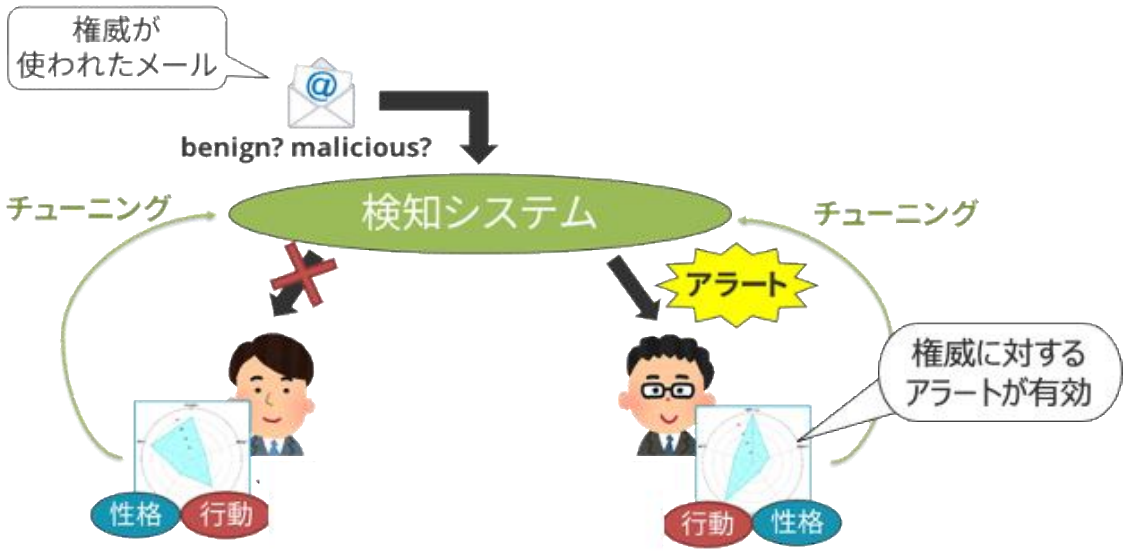

図 23　検知システムのチューニング[85]

5.1.1.アラートシステムの構成

本項では，本研究で提案するアラートシステムの構成を示す．図 24 にアラートシステムの構成と，アラートシステムの各構成要素，ユーザ，メールとの関係を示す．

アラートシステムは，個人特性検査部，メール文面分析部，アラート調整部，アラート出力部，個人特性 DB によって構成される．以下に，各構成要素の概要を示す．

個人特性検査部は，ユーザの個人特性をアンケートなどによって検査し，ユーザ ID と共に個人特性 DB に登録する．

メール文面分析部は，メールを入力として，メール文面からチャルディーニの法則を抽出し，抽出した文面の箇所と法則とをアラート調整部に渡す．

アラート調整部は，メールが届いたユーザ ID をキーとして，メールが届いた人物の個人特性を，個人特性 DB から抽出し，メール分析部から受け取った情報をもとに，該当する箇所に対するアラートの提示がその個人に有効なアラートであるかを判断した結果を，アラート出力部に渡す．

アラート出力部は，アラートを提示することが有効であるかの情報をアラート調整部から受け取り，ユーザに提示する．

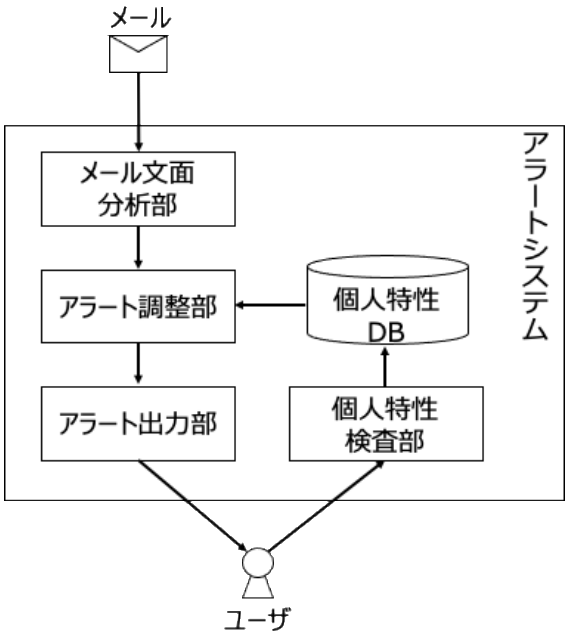

図 24　アラートシステムの構成と各要素との関係図

5.1.2.アラートシステムの動作

本項では，アラートシステムの動作を示す．本研究におけるアラートシステムは，準備フェーズと運用フェーズとで構成される．

準備フェーズでは，ユーザは，個人特性検査部によって性格や行動特性を検査し，検査結果を個人特性 DB に格納する．さらに，メールからチャルディーニの法則を抽出するための識別機を学習させる．学習した識別器は，メール文面分析部に置かれる．

運用フェーズでは，メールを受信した際にアラートシステムにメールが入力されるところから処理が始まる．まずメール文面からチャルディーニの法則を抽出し，抽出した文面の箇所と法則とを特定する．続いて，メールが届いたユーザ ID をキーとして，メールが届いた人物の個人特性を，個人特性 DB から抽出する．個人特性の情報と入力されたメール文面から抽出したチャルディーニの法則とその箇所をもとに，該当箇所に対してアラートを提示することが，その個人に対して有効であるかを判断する．有効である場合には，ユーザに対してアラートを提示する．

5.1.3.本研究での対象範囲

本研究においては，アラートシステムにおける重要な要素から検証した．まず，チャルディーニの法則を抽出できるかの検証を行うために，メール分析部の実現性を 5.2 にて検証した．続いて，個人に合わせたアラートが有効であるかの検証を行うために，アラート調整部の実現性を 5.3 にて検証した．個人特性検査部は，アラート調整部の検証を行うための実験において，性格検査と行動特性の取得方法を示した．本研究において未検証である箇所は，実際に届いたメールに対し，アラートを提示するアラート出力部である．アラート出力部の実装と評価は，今後の課題である．

5.2. チャルディーニの法則を抽出する方法

本研究は，心理操作効力の高い標的型メールに利用されると考えられる誘導手口（チャルディーニの法則）の有無をメールの文面から推定することを目的としている．チャルディーニの法則に利用されると考えられるキーワードを一意に定義することは容易ではなく，ルールベースによる特定は難しい．そこで，自然言語処理と機械学習を利用して，チャルディーニの法則に該当すると考えられる文面を推定する技術を提案する．機械学習によりチャルディーニの法則に該当すると考えられるパラグラフを推定するためには，チャルディーニの法則に該当するとラベル付けがされたパラグラフの教師データが必要である．

5.2.1.データセット

5.2.1.1. Enron Email Dataset

今回の実験でEnron メールデータセットを利用した[67][68]．提案手法では，メールの文章から，状態の推定を行う必要があり，そのためには学習された識別器が必要となる．このような識別器を用意するためには，学習データとして大量のメールが必要となる．そのため，自然な文章である大量のメールデータが必要である．

Enron メールデータセットは，約50 万通のビジネスメールを有しており，メールの自然言語処理分野では，データセットが登場した2004 年頃から頻繁に利用されている[67]．ここで, Enron について簡単に説明する．Enron とは，総合エネルギー取引とIT ビジネスを生業とする，全米でも有数の大企業であった．しかし巨額の粉飾決算が明るみになり，2001 年に倒産した．Enron メールデータセットは, E-Discovery を目的に, Federal Energy Regulatory Commission によって, Enron 社内のメールやり取りがWeb 上に公開されたもののことである．現在では，大量の自然言語によるビジネスメールであるため，メールに関する自然言語処理の研究において，データセットとして用いられている．用いられているデータセットにはいくつかの種類があるが，研究で最も用いられているデータセットは, Carnegie Mellon University (CMU) CALO Project dataset である．表 16に，Enron データセットの一部を示す．

表 16　Enron データセットの一部[68]

データ名称	データ数	特徴
FERC Enron Email Dataset	100万以上	PST 形式であるため，フォルダ構造や，添付ファイルの全てが保存されている
Carnegie Mellon University (CMU) CALO Project dataset	約50 万通	FERC dataset の派生．CALO Project 用に収集したデータセット．添付無．いくつかの感染した従業員からのメールと思われる分は削除されている．他にも，FERC Dataset と比較して，いつくかの最適化を施している．Enron データセットの中で，最も研究利用が多いデータセット．
EDO Enron Email PST Dataset	約50 万通	CALO Dataset をPST タイプに変更したもの．

5.2.1.2. Amazon Mechanical Turk（AMT）によるデータ収集

クラウドソースの Amazon Mechanical Turk（AMT）を利用して, Enron Email Dataset のメールのパラグラフごとチャルディーニの法則の教師データ（ラベル付きのデータセット）を作成した[69]．本来であれば，実際の不正なメールに対してラベル付けし学習データを作成すべきではあるものの，不正なメールの実サンプルの入手は非常に困難であるため，公開されている正常なメールのデータセットを利用した．

AMT Worker には, Akbar らのフローチャート[57]を参考に，どのような文面の場合にどのチャルディーニの法則に該当するかについて方針を文章で説明し，方針に従いラベル付けのタスクを行ってもらった．1 つのタスクは，1 件のメールに含まれるいくつかのパラグラフに対してどのチャルディーニの法則が該当するかを回答する作業である．1 つのパラグラフに複数のチャルディーニの法則が該当してもよい．AMT Worker には未加工のメールの本文を表示した．返信などの引用文は，可能な限り検出し，背景色を変えラベル付けの対象から外した．1 つのタスクに複数（3 人以上）の Worker を割り当てた．収集したデータセットの情報を表 17 に示す．

表 17　ラベル付けを行った Email の情報

項目	件数
Email 総数	22, 988
パラグラフ総数	122, 625

5.2.1.3. 一貫性のあるラベル付データセット作成

今回ラベル付けしたデータセットには，複数の異なるメールに短い同じパラグラフ（フレーズや簡単な文章）が含まれていた．例えば，メールの最初と最後の挨拶，署名，定型句，除外漏れの引用文などである．実際のメールにおいても，同様に異なるメールに同じパラグラフが含まれることがある想定し，データセット中の重複は除外せずそのままにした．一貫性のあるラベル付きデータセットを作成するために，データセット中の同一のパラグラフにおいて半数以上の Worker が同一パラグラフに同一のラベルを割り当てた場合に，そのラベルを採用した．表 18 にラベル付けされたデータセットの情報を示す．

表 18　ラベル付けされたデータセットの情報

項目	件数
全パラグラフ数	115039
権威（Authority）	6883
社会的証明（Consensus）	1151
一貫性（Consistency）	6166
好意（Liking）	2518
返報性（Reciprocity）	2155
希少性（Scarcity）	1139

5.2.2.前処理

識別モデルに入力する前に，パラグラフごとに，記号や特殊文字の除去，数字のタグ化などの基本的な前処理を行った．

5.2.3.識別モデル

各チャルディーニの法則に該当するかどうかの 2 値分類の識別器として，ベースライン方式と，ニューラルネットワークを利用した方式の二種類を比較のために作成する．

5.2.3.1. ベースライン方式

比較のために，従来の機械学習アルゴリズムを利用し，チャルディーニの法則の各ラベルを推定するベースラインモデルを作成した．事前実験において，比較的精度が高かった，Logistic Regression を採用した．Google が公開している学習済みの Word2Vec モデル（300 次元）[70][79]を用いて，パラグラフ中の単語から計算した Word2Vec の平均を特徴情報として利用した．Word2Vec は Google が公開しているモデルをそのまま利用した．Python[71]のライブラリである scikit-learn[72]および gensim[73]を利用し，Logistic Regression によ

る識別器を作成した.

Logistic Regression のパラメータである class_weight と C のみグリッドサーチで決定した．それ以外のパラメータはデフォルトのままである.

5.2.3.2. ニューラルネットワークを利用した方式

図 25 に示すニューラルネットワークを利用し，チャルディーニの法則の各ラベルを推定するモデル（ニューラルネットワークモデル）を作成した．順序関係を学習することができるリカレントニューラルネットワークの 1 つである Long short-term memory（LSTM）を利用しニューラルネットワークを構成した. Google が公開している学習済みの Word2Vec モデル（300 次元）を利用し, embed 層を構成し，パラグラフの単語ごとの分散表現を取得し，一層目の LSTM に入力する．学習時に embed 層は更新されない．パラグラフ前半の情報についても適切に学習を行うことを目的に，各時間の LSTM の出力に対して注意機構を用いて加重平均をとり出力層に入力する構成とした．陽性データと陰性データの割合をもとに class_weight を調整した．深層学習フレームワークの Chainer[74]を利用してニューラルネットワークモデルのプロトタイプを実装した.

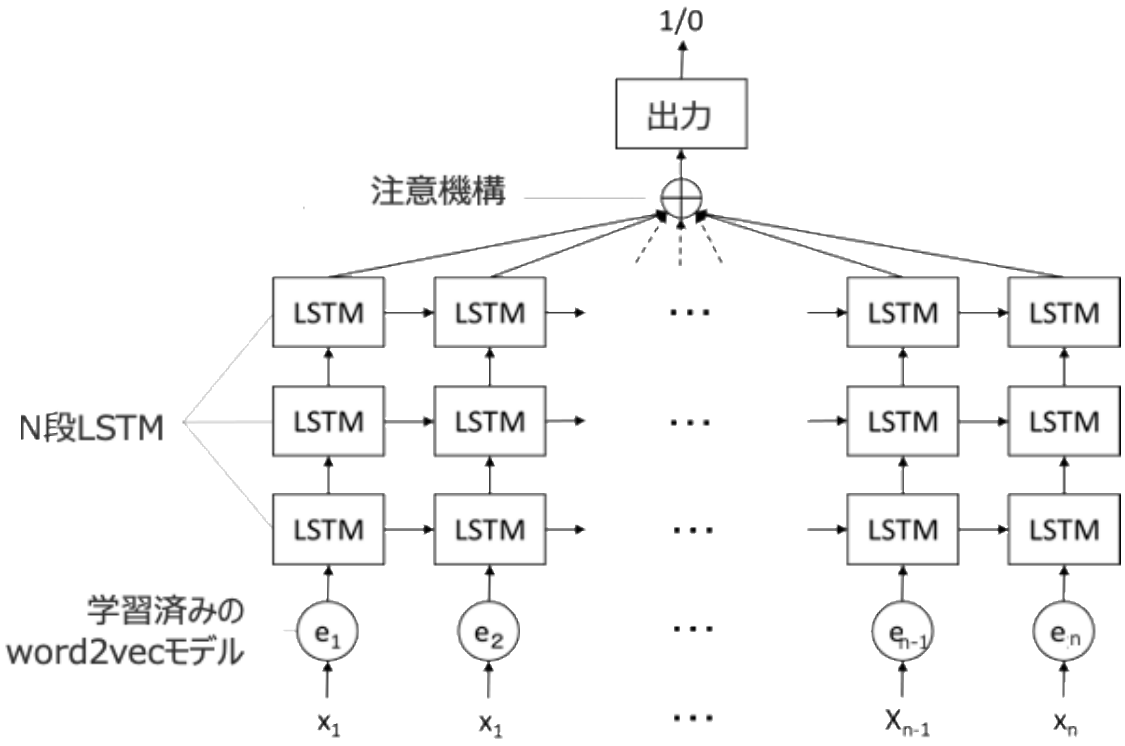

図 25 ニューラルネットワークのアーキテクチャ

5.2.4.精度の尺度

チャルディーニの法則のラベルごとに作成した 2 値分類モデルの正検知率（True Positive Rate：TPR）および誤検知率（False Positive Rate：FPR）の関係を表した Receiver Operating Characteristic（ROC）カーブから Area Under Curve（AUC）を算出した．AUC は 1.0 に近ければ近いほど，そのモデルの精度が高いことを意味している．AUC が 0.5 とは，当て推量で 2 択に答えているのと同程度の精度であると判断できる．5-fold cross validation を行い，5 回の検証の AUC とその平均をそれぞれ算出した[*9]．5-fold cross validation では，表 18 に示すデータセットを，ランダムに 5 等分し，3 つを訓練データ，1 つを検証データ，残り 1 つをテストデータとした．

ベースラインモデルについては，各 cross validation において，訓練データと検証データを使ったグリッドサーチによる検証を行い，AUC が最も高いパラメータを決定し，テストを行った．ニューラルネットワークモデルについては，各 cross validation において，epoch ごと訓練データと検証データを用いた検証を行い，AUC が最も高い epoch のモデルをテストに利用した．epoch の数は 20 とした．

5.2.5.評価結果

ベースラインモデルとニューラルネットワークモデルの評価結果を表 19 と表 20 にそれぞれ示す．表中 AUC は，各 cross validation におけるテストデータによる AUC の値である．また AUC の平均は，5 回分の cross validation の AUC の平均値である．

どちらのモデルに関しても，AUC の平均が 80%を超す結果が得られておらず，十分な精度とは言えないが，機械学習を用いることで，メールのパラグラフ中に存在する，チャルディーニの法則を特定する可能性を示唆することができたと考える．

ベースラインモデルとニューラルネットワークモデルとの比較では，社会的証明と希少性以外の法則において，ニューラルネットワークモデルの方がごくわずかではあるが AUC が高いことが確認された．

*9 今回のデータセットは陽性データと陰性データの数の偏りが非常に大きな不均衡データセットである．不均衡データセットにおいて Accuracy（正確度）を精度の尺度に利用することは適切ではない．

表 19　ベースラインモデルによる識別精度

	AUC (CV1)	AUC (CV2)	AUC (CV3)	AUC (CV4)	AUC (CV5)	AUCの平均
権威 (Authority)	0.7196	0.7343	0.7292	0.7430	0.7178	0.7288
社会的証明 (Consensus)	0.7374	0.7674	0.7723	0.7691	0.7511	0.7595
一貫性 (Consistency)	0.6892	0.7007	0.6981	0.6925	0.6959	0.6953
好意 (Liking)	0.7215	0.7822	0.7517	0.7670	0.7344	0.7514
返報性 (Reciprocity)	0.7551	0.7673	0.7862	0.7868	0.7737	0.7738
希少性 (Scarcity)	0.6857	0.7163	0.6958	0.6768	0.6594	0.6868

表 20　ニューラルネットワークによる識別精度

	AUC (CV1)	AUC (CV2)	AUC (CV3)	AUC (CV4)	AUC (CV5)	AUCの平均
権威 (Authority)	0.7263	0.7532	0.7541	0.7384	0.7343	0.7413
社会的証明 (Consensus)	0.7506	0.7310	0.7844	0.7175	0.7359	0.7439
一貫性 (Consistency)	0.6819	0.7036	0.7234	0.6862	0.6878	0.6966
好意 (Liking)	0.7395	0.7855	0.7758	0.7512	0.7636	0.7631
返報性 (Reciprocity)	0.7666	0.7658	0.8170	0.7683	0.7870	0.7809
希少性 (Scarcity)	0.6411	0.6634	0.6896	0.6771	0.6697	0.6682

5.2.6.考察

ベースラインモデルとニューラルネットワークモデルとの比較では，社会的証明と希少性以外の法則において，ニューラルネットワークモデルの方がごくわずかではあるが AUC が高いことが見て取れる．

表 18 を見ると社会的証明と希少性は他の法則と比べ，陽性データが少ない．ニューラルネットワークにおいては，データ量が精度に強く影響を与えることが知られており，陽性データの少なさがニューラルネットワークモデルの精度が低くなった要因の 1 つと考える．しかしながら，計算量のコストを考えると，ニューラルネットワークを利用する効果があまり得られなかったと考える．

陽性データの数の差に加え，利用する特徴情報やニューラルネットワークのアーキテクチャなどにも，チャルディーニの法則ごと精度に異なる影響が与えられていると考える．例えば，権威 (Authority) は，権威にかかわる単語の有無が重要な特徴となりうると考えられ

る．

希少性（Scarcity）は希少性に関わる情報（例えば，時間，数量などの数）の有無が重要な特徴となりうる．好意（Liking）や返報性（Reciprocity）は，言葉の言い回し，表現の柔らかさ，礼儀正しさなどのフレーズや文章レベルの情報の有無が重要な特徴となりうる．一方，社会的証明（Consensus）や一貫性（Consistency）は，過去に行ったことや他の人の発言や行動など，メールのやりとりなどの文脈が重要な特徴となりうる．ベースラインモデルもニューラルネットワークモデルも，全てのチャルディーニの法則に対して同じもモデルを利用しているため，チャルディーニの法則ごと得て不得手が現れたと考える．今後は，法則ごと適切なモデルについても検討していく予定である．

チャルディーニの法則は正常なメールにおいても使われることがある．そのためこの特徴だけでは不正なメールを特定するまでにはいたらないだろう．しかしながら，攻撃者の真意（攻撃を成功させたい）を考えると，チャルディーニの法則が不正なメールに利用される頻度は正常なメールのそれと比べると多いのではないかと推測する．今後は，正常なメールと不正なメールに含まれるチャルディーニの法則に該当する文章の割合がどの程度異なるのかを，本モデルを利用して検証していく．

5.3. アラートの効果検証

本節では，個人の性格因子や行動特性に応じて表示するアラートを変更することが有効であるか，ユーザ実験により調査した結果を示す.

5.3.1.リサーチクエスチョン

本章の冒頭で述べたように，巧妙な標的型メール攻撃に対しては，個人の特性に合わせたアラート調整が有効であると考えられる．そこで本節では，個々人の性格因子や行動特性が，組織内でのアンケート等によって収集し，そのデータを組織内部で活用可能であることを前提とし，「標的型メール文面にて使われているチャルディーニの法則への注意を促すアラート」と「個人の性格因子および行動特性」との関係性を，ユーザ実験により調査する.

本節でのリサーチクエスチョンとして，以下を掲げる.

RQ1：単純にアラートを提示することと，各種のチャルディーニの法則への注意を促すアラートとではメールを開くという行動の傾向は異なるのか.

RQ2：個々人で有効なチャルディーニの法則への注意を促すアラートは，性格因子と行動特性の両者によって異なる傾向を示すのか.

5.3.2.ユーザ実験

本節では，5.3.1 項で掲げたリサーチクエスチョンに答えるために，クラウドソーシングを通じて募集した実験協力者に対する性格検査と擬似標的型メールへのアラートに対する反応度調査を用いたユーザ実験を行い，その結果を分析する．本項では，ユーザ実験の手順を説明する.

5.3.2.1. 実験・分析の流れ

今回実施するユーザ実験および分析の流れを説明する.

A) 名字の入力：
実験協力者に自分の名字を入力してもらう．標的型メールでは，標的者の名前を文面に用いることで，標的者に正規のメールであると信じさせる手口が用いられている．そのため，実験で利用する疑似標的型メールの文面を実際の攻撃に似せるため，名字をメールの宛名に埋め込んだ．名字は，疑似標的型メールを作成するためだけに利用し，記録しない.

B) 属性情報の調査：
実験協力者に，性別，年齢層，職種，業務，IT 利用形態などの情報を入力してもらう.

C) 性格検査による性格因子（ビッグファイブ）の調査：
実験協力者に対し，5.3.2.3 に示す性格検査を実施する.

D) 擬似標的型メールのプレーンメールを用いた行動特性調査：
実験協力者に対し，5.3.2.4 に示す行動特性の内，プレーンメールを開きやすいかどうかの行動特性の調査を実施する．

E) 疑似標的型メールの開封意図への注意を促すアラートが表示されたアンケートによる反応度調査：
実験協力者に 5.3.2.6 に示す反応度調査を実施する．

F) RQ1 に対する分析：
全実験協力者に実施した(A)から(E)の実験で得られたデータに対し，ウィルコクソンの順位和検定とスピアマンの順位相関係数を用いて 5.3.5 に示す分析を実施する．本分析の一部では，実験協力者の集合を二つの群に分けて分析を行う．

G) RQ2 に対する分析：
全実験協力者に実施した(A)から(E)の実験で得られたデータに対し，本分析では，実験協力者の集合を性格因子と行動特性によって分割し，分析を行う．

5.3.2.2. 実験協力者

本ユーザ実験では，クラウドソーシングのランサーズ[63]を利用して一週間の期間で集まった 433 名の実験協力者を集め，実験を行った．実験協力者への支払いは，1 名あたり税込み 165 円とした．

実験協力者数を決定した手順は以下のとおりである．実験の実施に先立ち，検出力分析を行った．検出力，有意水準，効果量の 3 つを次のように定めることで，十分なサンプルサイズを決定した．有意水準と検出力は，文献[59]などで適正値として示唆されている 5%, 0.8 とした．効果量を大として，文献[59]をもとに各検定に対する効果量の値（ウィルコクソン検定：0.8，相関分析：0.5）を決定した．その理由は,4 章の実験と同様に，チャルディーニの法則が標的型メールの開封率を上げる効果や，標的型メールの開封率における性格因子および行動特性とのチャルディーニの相関は大きいと考えたためである．その結果，ウィルコクソン符号付き順位和検定（片側検定）のサンプルサイズは 12 名，ウィルコクソンの順位和検定（両側検定）では 27 名，スピアマンの順位相関係数（両側検定）では 29 名となる．(F)の実験では，2 つの実験協力者群が必要となるため，29 名の 2 倍である 58 名がサンプルサイズとして必要な人数となる．外れ値の除外やデータの分割によりサンプルサイズの実数が減少することに備え，実験協力者の人数を 400 名以上に設定した．

実験協力者は，あらかじめ以下の(ア)～(オ)に関する説明を受け，同意をした上で，ユーザ実験に参加している．

(ア) 本ユーザ実験で収集した情報は，個人が特定できない状態に加工したうえ，学術目的で利用すること．

(イ) 名字の入力を求めるが，名字は実験中の質問時にのみ利用し，記録しないこと．

(ウ) ユーザ実験の質問に最後まで回答していない場合には報酬を受け取ることができ

ないこと.

(エ) 「問題文をしっかり読んでない回答」と実験後に実験実施者により判断された場合には報酬を受け取ることができないこと.

(オ) 会社員を対象とした実験のため, 会社員以外は回答しないこと.

(オ)の条件を追加した理由は, 標的型メールは特定の企業など組織をターゲットとして送信されるため, 会社員の経験があるほうが, 今回の実験趣旨にあった回答が得られると考えたためである. ただし, Web アンケートの性質上, 参加者の正確な職業を知ることはできない. そのため本ユーザ実験の結果には, 会社員以外の参加者の回答も含まれている可能性がある.

5.3.2.3. 性格検査

本節で実施する実験は, 4.3.3 節にて説明した性格検査と同一の手法によって算出するため, ここでの詳細は省略する.

5.3.2.4. 行動特性の調査

本節においても, 4.3.4 節と同様に,「そもそもチャルディーニの法則を用いていない標的型メールを開いてしまうかどうか」という行動特性に対して特に焦点を当てる. さらに本節ではアラートに対する反応を検証したいため, 行動特性として「アラートに対して反応しやすいかどうか」も追加する. 本節の実験において,「プレーンメール（チャルディーニの法則を組み込んでいない標的型メール）への反応度」が高いか低いか,「シンプルなアラートへの反応度」が高いか低いかによって, 各実験協力者の行動特性が 4 つに分類される. 行動特性によるデータの分割については 5.3.5.3 にて示す.

5.3.2.5. アラートの反応度を尋ねる質問作成手順

アラートの反応度を尋ねる質問データセットを作成する手順を以下に示す.

① 日本サイバー犯罪対策センター(JC3)[62]で公開されている, 実際の標的型攻撃で利用された標的型メールの文面から無作為に 10 種類のメール（オリジナルメールと呼ぶ）を選択する.

② オリジナルメールからチャルディーニの法則が利用されていると考えられる部分を, 4.3.5.1 での手順と同様に削除し, チャルディーニの法則が使われていない 10 種類の疑似標的型メール（プレーンメールと呼ぶ）を作成する.

③ 各プレーンメールに対して, チャルディーニの各法則（希少性, 返報性, 権威, 一貫性, 好意, 社会的証明）の全てを取り入れた擬似標的型メール（オールインメールと呼ぶ）を作成する. オールインメールの生成手順の詳細は後述する.

④ オールインメールに対して各種アラートをつけた質問を生成する. 各種アラートは, 単純にチャルディーニの法則が使われていることへの注意を促すもの（シンプルなアラ

ート）と，オールインメールに記載されている各チャルディーニの法則に対して注意を促すもの，の合計 7 種類の質問が作成される．

ここで，オールインメールの作成手順と，生成される 7 種類の質問の具体例を示す．

まず，オールインメールの作成手順を以下に示す．

① 4.3.5.1 にて生成したチャルディーニメールから，追加したチャルディーニの法則に関わる文言を抽出する．この処理により，各チャルディーニの法則について，3 種類のチャルディーニ文言が得られる．

② 各プレーンメールについて，出現させるチャルディーニの法則の順番をランダムに設定する．

③ 各チャルディーニの法則について，出現させるチャルディーニ文言を 1 から 3 の間でランダムに決定する．

④ 各プレーンメールにおいて，手順②で決定した出現するチャルディーニの法則の順番で，手順③で決定した出現させるチャルディーニ文言を入れ込み，オールインメールを生成する．この際，著者らで協議の上，自然な文面となるように調整する．

プレーンメールを元に生成したオールインメールとの対応例を表 21 に示す．

続いて，生成される 7 種類の質問の例を，実際に実施したアンケートにおける画面である図 26，図 27 に示す．図 26 は，オールインメールに対して，チャルディーニの法則における好意の法則の箇所に注意を促すことを示す文言を含んだバルーンを，該当箇所に示している．図 27 は，オールインメールに対して，メール全体に添付ファイルの開封を誘導する文言が含まれていることを示している．以後，図 27 のようなアラートを本書ではシンプルなアラートと呼ぶ．

表 21　プレーンメールとオールインメールの対応例

プレーンメール	オールインメール	チャルディーニの法則順番の出現順とチャルディーニ文言 ID
{NAME}様 XLS 版にて送付致します. 添付ファイルのご確認，宜しくお願いいたします.	{NAME}様 XLS 版にて送付致します. 添付ファイルのご確認，宜しくお願いいたします. ＊いつも申し訳ございません. ＊前にもして頂いた通りです. ＊ファイルは会社規定により本日のみ閲覧可能です. 他の皆様と同じように，{NAME}様も，宜しくお願いいたします. 貴社の部長より，対応の要請がありました. 私の方でも一度確認しました.	好意:3 一貫性:3 希少性:3 社会的証明:1 権威:1 返報性:2
{NAME}様 おつかれさまです. 修繕依頼書・発注書が来ましたので添付します. ご対応願います.	{NAME}様 おつかれさまです. 修繕依頼書・発注書が来ましたので添付します. ご対応いただけますと大変助かります. ＊他の皆様には，既に確認頂いております. 今日中に，よろしくお願いいたします. 依頼された件，こちらは対応しましたので，そちらでもご対応願います. ＊担当責任者から確認して頂くよう連絡がありました. 前回と同じように，よろしくお願いします.	好意:2 社会的証明:3 希少性:2 返報性:1 権威:3 一貫性:1

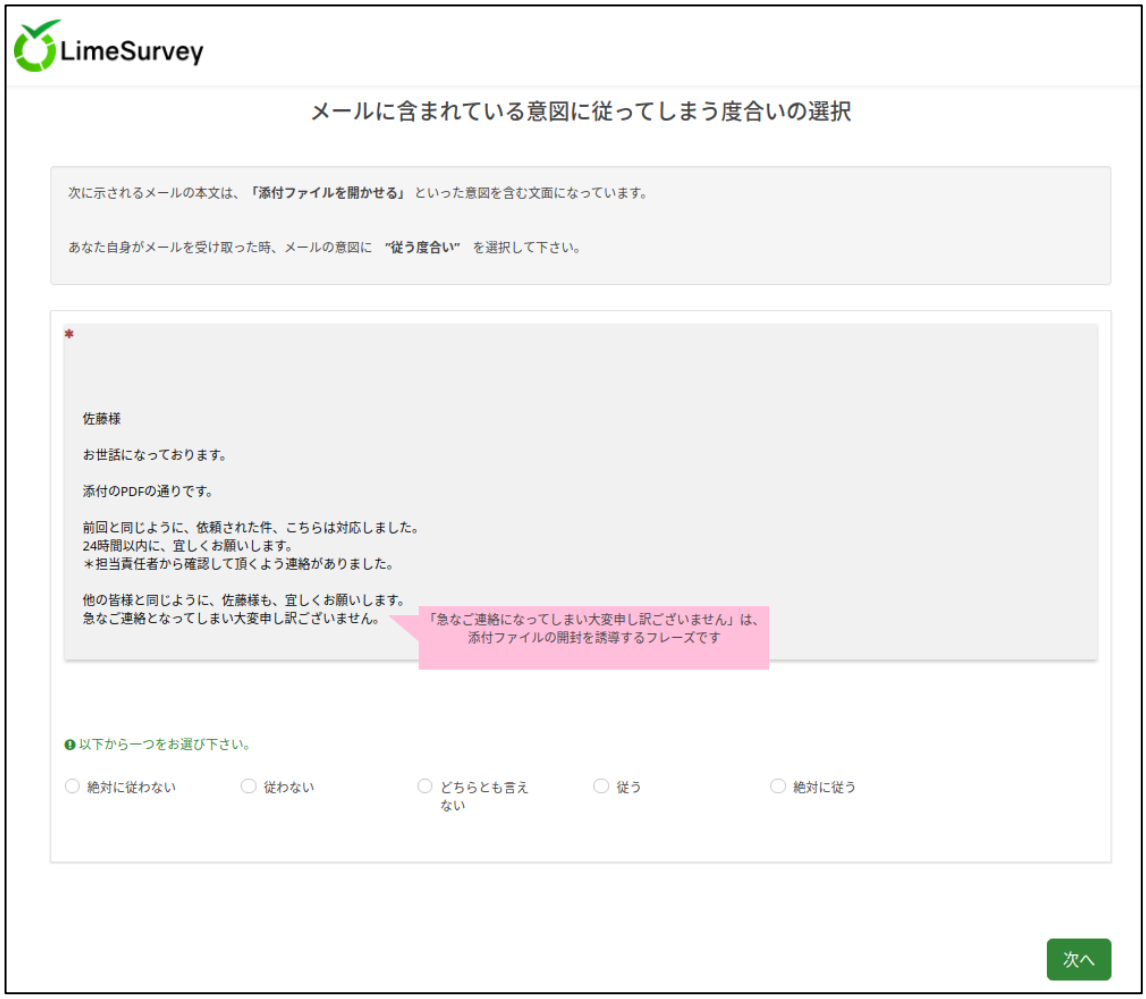

図 26　質問例（好意の法則に対して注意を促すアラート）

図 27　質問例（メール文面全体に対して注意を促すアラート（シンプルなアラート））

5.3.2.6. 実験手順

5.3.2.5 で作成した質問を用いて，反応度調査を実施する．本節での実験では，回答の一貫性を確認するために，三種類の質問に対して同一の質問を出現させる．同一の質問に対する回答が，著者らの想定している範囲内に存在しているかによって，回答が一貫しているかどうかの判定を行う．回答が一貫しているかどうかの判定結果は，5.3.3.4 に示すように，外れ値の判定として利用する．

ここで，10 種類のプレーンメールによって構成されているプレーンメール集合をPとして，プレーンメールp_iを，$P \ni p_i (0 \leq i \leq 9)$と定義する．続いて，各プレーンメールから生成される 10 種類のオールインメール集合をMとして，オールインメールm_iを$M \ni m_i (0 \leq i \leq 9)$と定義する．さらに，オールインメールに対して，各種のアラートを付した質問を$q_{ij}\ (0 \leq i \leq 9)(0 \leq j \leq 6)$とする．チャルディーニの希少性，返報性，権威，一貫性，好意，社会的証明の各法則に番号を振り，j 番目の法則（1≦j≦6）と呼ぶことにする．ここで，j=0 の時，シンプルなアラートを意味するものとする．さらに，プレーンメールを用いた質問をq_{i7}とする．

① プレーンメール集合Pから，重複のないようランダムに 3 つの要素を抽出したものをQ^Pとする．ここで，$n(Q^P) = 3$である．

② 実験で提示するアラートの 7 種類（シンプル，チャルディーニの各法則の合計 7 種類）について，それぞれオールインメールを三つずつ，重複のないようランダムに抽出したものをQ^Mとする．ここで，$n(Q^M) = 21$である．

③ Q^Mの集合から，重複のないようランダムに 3 つの要素を抽出したものをQ^Cとする．この時，抽出した要素q_{ij}の添字によって構成される集合Lを定義する．集合Lに記録される要素cは，抽出した要素がq_{ab}である場合，(a, b)となる．ここで，$n(Q^C) = 3, n(L) = 3$である．

① 集合$Q^p \cup Q^M \cup Q^C$の要素から，q_{ij}を重複のないようランダムに 1 つ抽出し，提示する．実験協力者は，提示されたq_{ij}において，「メールに書かれている指示に従ってしまう度合い」を「1：絶対に従わない」，「2：従わない」，「3：どちらともいえない」，「4：従う」，「5：絶対に従う」の 5 件法[*10]で，メールごとに回答する．集合$Q^p \cup Q^M \cup Q^C$に含まれる要素全てに対して回答を行う．

反応度調査において，プレーンメールの後にアラートへの反応を尋ねる質問が出てくるかを決めてしまうと，順序効果（プレーンメールの文面を見たことがアラートへの反応を尋ねる質問への回答に影響を与える）が発生する．これを防ぐために，集合$Q^p \cup Q^M \cup Q^C$を全て集めたのちに，重複のないようランダムに抽出することで実験協力者への質問を選択している．

[10] 本実験では，メール文面に対してアラートを提示することで，標的型メールに対する反応がどのように変化するかを観測しようとしている．その変化の度合いを適度な粒度で分析するために本実験では 5 件法を用いた．

5.3.3.前処理

実験によって得られたデータは，前処理を実施した後に，リサーチクエスチョンに回答するための分析に適用する．本項では，本実験で実施した各種前処理を説明する．なお，性格因子スコアと行動特性の算出は，5.3.2.3，5.3.2.4 にて説明した．本項で実施する前処理によって外れ値を除外したデータ件数は 357 件となった．得られた全データの相対反応度及び属性情報の統計値を表 22，表 23 に示す．本項以後，357 件のデータセットに対して処理や分析を実施した結果を示す．

表 22　全実験協力者の相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
シンプル	357	2.00	-1.67	0.01	0.49
希少性	357	2.00	-2.33	-0.05	0.50
返報性	357	3.00	-1.67	0.05	0.51
権威	357	2.00	-1.67	-0.08	0.52
一貫性	357	2.67	-2.67	-0.02	0.54
好意	357	2.33	-2.00	-0.03	0.54
社会的証明	357	2.00	-1.67	0.01	0.49

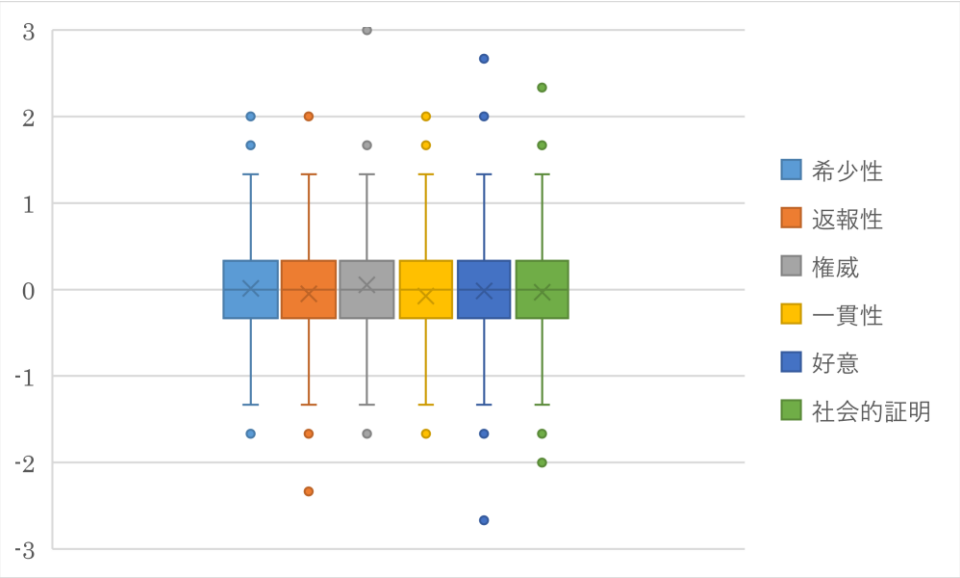

図 28　全実験協力者の相対反応度の箱髭図

表 23　全実験協力者の属性情報

属性種別	属性	人数（人）
性別	男性	204
	女性	153
年齢層	10 代	1
	20 代	56
	30 代	133
	40 代	122
	50 代	37
	60 代	8
	70 代	0
	80 代	0
職種	購買・仕入れ	9
	調査・広告・宣伝	5
	製造・生産・品質管理	31
	経営・事務計画	13
	物流・配送	8
	技術開発・設計	28
	情報処理（システム）	31
	営業・販売	65
	広報・編集	8
	人事・総務・経理	72
	商品企画・開発	9
	基礎・技術研究	4
	個人事業主・店主	5
	企業等の経営者・役員	1
	その他	68

5.3.3.1. アラート反応度の算出

5.3.2.1 の項目 E での回答結果の内，オールインメールを三つずつ，重複のないようランダムに抽出したものであるQ^Mへの回答結果をもとに，各種アラート反応度$r_i(0 \leq i \leq 6)$の算出方法を説明する．

① 得られた質問への回答結果から，Q^Mに対する回答結果Vのみを抽出する．この時，Q^Cへの回答と区別する方法は，5.3.3.4 にて後述する．

② 得られた回答結果において，各アラート種別（シンプル，チャルディーニの各法則）の 7 種それぞれに対する回答結果$v_{ij}(0 \leq i \leq 6)(0 \leq j \leq 2)$の結果を抽出する．

③ 各種アラート反応度を以下の式のように計算する．

$$r_i = \frac{1}{3}\sum_{j=0}^{2} v_{ij}$$

5.3.3.2. 相対反応度の算出

それぞれの実験協力者ごとに，シンプルなアラートの反応度 r_0を除く各アラート種別の反応度$r_i(1 \leq i \leq 6)$から，シンプルなアラートの反応度 r_0を引いた値を，各アラート種別に対する相対反応度 $r'_i = r_i - r_0(1 \leq i \leq 6)$と定義する．今回の実験では，チャルディーニの法則への注意を促すアラートを提示することで，標的型メールを開きにくくなるか否かを確認することが目的である．そのため，相対反応度が，「標的型メールに使われているチャルディーニの法則への注意を促すことで，標的型メールが開封されやすくなるか否かを確認する」にあたっての尺度となる．

5.3.3.3. 回答時間による外れ値除去

実験結果のうち，実験協力者が 5.3.2.1 の項目 A～E への回答に要した総回答時間の分布において，極端に大きな値および小さい値を外れ値として除外した．外れ値を求める際には回答時間の四分位数を用いた．第 1 四分位数より四分位範囲の 1.5 倍以上小さい値，あるいは第 3 四分位数より四分位範囲の 1.5 倍以上大きい値を外れ値とした．実験協力者の回答時間の平均は 12 分 9 秒であり，最も短い回答時間は 3 分 52 秒，最も長い回答時間は 4 時間 54 分 37 秒であった．となった．なお本ユーザ実験は，Web アンケート形式であり，実験協力者が実験実施者の想定しているとおりに実験を実施しているかどうかを把握することはできないことに注意されたい．

5.3.3.4. 質問の回答一貫性チェックによる外れ値除去

一貫性チェックの質問で利用した添字集合Lを用いて，$Q^M \cup Q^C$に対する回答から二度質問している項目の回答を得て，それぞれの回答を比較し，質問の回答に一貫性があるかの

判断を行う．３つの一貫性チェックの質問に対してそれぞれ一貫性の判断を行い，３つのうちのどれか一つでも「一貫性がない」と判断された回答は，外れ値として除外する．

ここで，本論文では，一貫性があるかどうかの判断を，「メールに書かれている指示に従ってしまう度合いを尋ねる同一の質問について，一回目と二回目，それぞれの質問における回答の変化が小さい場合」を一貫性があると判断する．本論文における，一貫性がある回答の定義を表 24 に示す．なお，各数字の意味は，「1：絶対に従わない」，「2：従わない」，「3：どちらともいえない」，「4：従う」，「5：絶対に従う」である．

表 24　一貫性があると判断する反応度の範囲

一回目の回答	二回目の回答
1	1, 2, 3
2	1, 2, 3, 4
3	2, 3, 4
4	2, 3, 4, 5
5	3, 4, 5

5.3.3.5. 質問への回答が全て同一のものは除外

5.3.2.1 の項目 D, E で実施される 27 件の質問は，全て 5 件法（「1：絶対に従わない」，「2：従わない」，「3：どちらとも言えない」，「4：従う」，「5：絶対に従う」）によって回答されている．ある実験協力者において，D, E の回答全てが同一の選択肢を選んでいる場合，その実験協力者の回答を外れ値として扱う．

5.3.4.RQ1 に対する分析

RQ1「単純にアラートを提示することと，各種のチャルディーニの法則への注意を促すアラートとではメールを開くという行動の傾向は異なるのか」に答えるために，シンプルなアラートが提示される場合とチャルディーニの法則への注意を促すアラートが提示される場合とで，標的型メールを開く傾向が異なるかを分析した[76][77][78]．分析は，シンプルなアラートに対する反応度と，各チャルディーニの法則の注意を促すアラートに対する反応度に有意差があるかを，次の帰無仮説と対立仮説により両側検定を実施した．今回，有意水準は 5%とした．

帰無仮説：シンプルアラートの提示とチャルディーニの法則に応じたアラートの提示とで標的型メールに対する反応は変わらない

対立仮説：シンプルアラートの提示とチャルディーニの法則に応じたアラートの提示とで標的型メールに対する反応は変わる

各実験協力者にシンプルアラートとチャルディーニの各法則に対するアラートの提示に対する反応度を聞いているので，二つの反応度は対応したデータ（対標本から得られたデータ）である．このため，ウィルコクソン符号付き順位和検定により検証した．

検定の結果を表 25 に示す．希少性の法則，好意の法則，社会的証明の法則に対しては，有意差を確認することができず，帰無仮説を棄却できなかった．権威の法則，一貫性の法則では有意差（$p<0.05$）をそれぞれ確認することができ，帰無仮説を棄却することができた．ただし,権威の法則，一貫性の法則における効果量 d は 0.1 以下であるため，プレーンメールとチャルディーニメールとの間で反応度の差は小さい[59]ことが分かる．返報性の法則，権威の法則，一貫性の法則における検出力は 0.17，0.30 であるため，相応の第二種の過誤が含まれる．以上より,いくつかのチャルディーニの法則については，メール本文内でのその法則の利用に注意を促すアラートを提示することと，法則の種類に関わらず全てのメールに統一のアラートを提示することとの間に，限定的ではあるが違いがあることを示すことができ，RQ1 が部分的に成り立つことを示すことができた．

表 25　シンプルアラートとチャルディーニの各法則に対するアラート間での検定結果

	p 値（α）	効果量 d
希少性	0.389	0.014
返報性	0.099*	0.053
権威	0.026*	0.054
一貫性	0.004*	0.079
好意	0.785	0.018
社会的証明	0.743	0.032

*: $p<0.05$

5.3.5.RQ2 に対する分析

RQ2「個々人で有効なチャルディーニの法則への注意を促すアラートは，性格因子と行動特性の両者によって異なる傾向を示すのか」に答えるために，各種のチャルディーニの法則への注意を促すアラートに応じて標的型メールへの開きやすさに影響があるかを分析した[76][77][78]．分析は，(RQ2-1)性格因子によって各アラートの効果が異なるか，(RQ2-2)行動特性によって各アラートの効果が異なるか，(RQ2-3)性格因子と行動特性の 2 軸によって各アラートの効果が異なるか，の 3 つの観点から実施する．

5.3.5.1. 性格因子スコアの算出

並川らの性格検査によって性格因子スコアを算出した[61]．実験協力者ごとに各性格因子に対応する複数の質問に対して，それらへの回答の点数をそれぞれ算術平均し，これを各実験協力者の 5 つの性格因子のスコアとする．得られた全データの統計値を表 26 に示す．

表 26　性格因子の統計値

	N	最大値	最小値	平均	標準偏差
情緒不安定性	357	5.00	1.00	3.45	0.93
外向性	357	5.00	1.00	2.95	0.93
開放性	357	5.00	1.00	3.25	0.74
調和性	357	5.00	1.00	3.38	0.74
誠実性	357	5.00	1.14	3.23	0.77

5.3.5.2. 性格因子に関する分析結果

RQ2-1「性格因子によって各アラートの効果が異なるか」に答えるために，性格因子スコアと各チャルディーニの法則アラートの相対反応度との間で，スピアマンの順位相関係数を算出した．実験協力者全体（357 名のデータセット）に対して分析を行った．有意水準は 5%とした．

算出した相関係数を表 27 に示す．どの関係においても相関が見られなかったため，RQ2-1 は成り立たず，性格因子のみでは各アラートの効果を表すことができないことが示された．以降で示すように，行動特性を考慮することで，アラートごとに傾向が異なるため，行動特性を考慮する必要があることが本結果より導かれた．

表 27　全データにおける性格因子とチャルディーニの法則アラートとの相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
シンプル	0.070	0.054	0.012	-0.018	0.009
希少性	-0.074	0.055	-0.027	0.070	0.048
返報性	0.085	-0.086	-0.085	-0.131	-0.081
権威	-0.057	0.034	0.030	0.018	0.003
一貫性	0.008	-0.056	0.063	0.017	0.075
好意	0.056	0.064	0.077	0.005	-0.004
社会的証明	-0.002	-0.062	-0.111	0.013	-0.016

**: p<0.01, *: p<0.05, †: p<0.1

5.3.5.3. 行動特性に基づく実験協力者の分割

RQ2-2, RQ2-3 に答えるための分析を行うにあたり，その準備として，「標的型メールを開きやすいかどうか」，「アラートに対して反応しやすいかどうか」という行動特性(5.3.2.4)に着目して，実験協力者を 4 つの群に分割した．今回は，プレーンメール（チャルディーニの法則を組み込んでいない標的型メール）への反応度が 3 以下の実験協力者を「プレーンメールを開きにくい群」とし，3 より大きい実験協力者を「プレーンメールを開きやすい群」，シンプルなアラートへの反応度が 3 以下の実験協力者を「シンプルアラートメールを開きにくい群」，3 より大きい実験協力者を「シンプルアラートメールを開きやすい群」として，それぞれの場合の組み合わせとして，第一象限（プレーンメールを開きやすく，シンプルアラートメールを開きやすい），第二象限（プレーンメールを開きにくく，シンプルアラートメールを開きやすい），第三象限（プレーンメールを開きにくく，シンプルアラートメールを開きにくい），第四象限（プレーンメールを開きやすく，シンプルアラートメールを開きにくい）の 4 種類に分類した．分類を象限により図解したものを図 29 に示す．結果，第一象限の回答数は 145 件，第二象限の回答数は 100 件，第三象限の回答数は 83 件，第四象限の回答数は 29 件となった．行動特性によって分類された各群それぞれの相対反応度及び属性情報の統計値を表 28～表 35 に示す．

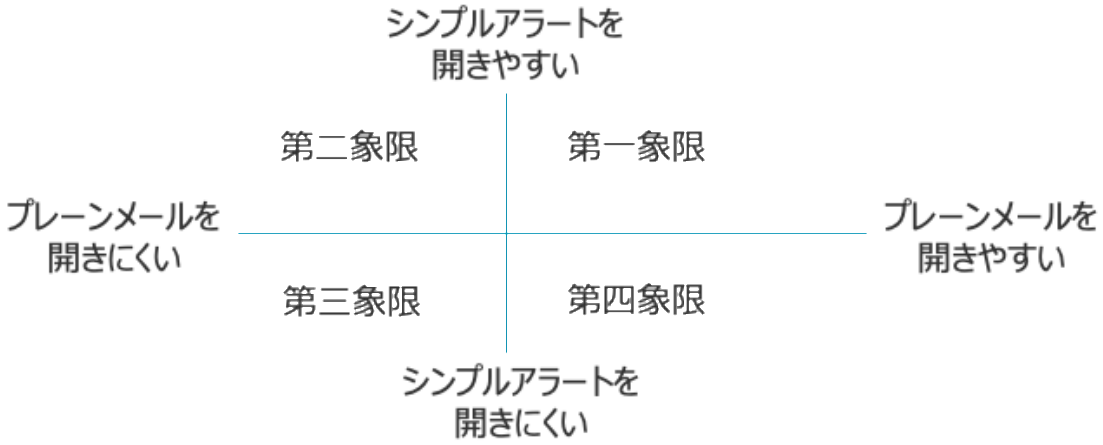

図 29　行動特性による分類の図解

表 28　第一象限の群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	145	1.00	-1.67	-0.02	0.43
返報性	145	1.00	-2.33	-0.10	0.48
権威	145	1.33	-1.67	-0.05	0.48
一貫性	145	1.00	-1.33	-0.14	0.42
好意	145	1.00	-2.67	-0.09	0.48
社会的証明	145	1.00	-2.00	-0.07	0.50

表 29　第二象限の群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	100	1.67	-1.67	-0.02	0.51
返報性	100	0.67	-1.33	-0.19	0.46
権威	100	1.33	-1.33	0.06	0.47
一貫性	100	0.67	-1.67	-0.21	0.52
好意	100	0.67	-1.33	-0.11	0.46
社会的証明	100	1.00	-2.00	-0.11	0.57

表 30　第三象限の群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	83	1.67	-1.00	0.04	0.54
返報性	83	1.33	-1.00	0.09	0.47
権威	83	1.67	-1.00	0.14	0.49
一貫性	83	1.33	-1.33	0.05	0.55
好意	83	2.00	-1.67	0.06	0.59
社会的証明	83	1.33	-1.00	0.07	0.50

表 31　第四象限の群における相対反応度の統計値

	N	最大値	最小値	平均	標準偏差
希少性	29	2.00	-0.67	0.21	0.54
返報性	29	2.00	-0.67	0.28	0.62
権威	29	3.00	-0.67	0.28	0.71
一貫性	29	2.00	-0.33	0.31	0.58
好意	29	2.67	-0.33	0.43	0.68
社会的証明	29	2.33	-1.00	0.15	0.71

表 32　第一象限の群の属性情報

属性種別	属性	人数（人）
性別	男性	72
	女性	73
年齢層	10 代	1
	20 代	20
	30 代	67
	40 代	47
	50 代	9
	60 代	1
	70 代	0
	80 代	0
職種	購買・仕入れ	7
	調査・広告・宣伝	3
	製造・生産・品質管理	8
	経営・事務計画	6
	物流・配送	0
	技術開発・設計	6
	情報処理（システム）	8
	営業・販売	29
	広報・編集	4
	人事・総務・経理	34
	商品企画・開発	5
	基礎・技術研究	1
	個人事業主・店主	2
	企業等の経営者・役員	0
	その他	32

表 33　第二象限の群の属性情報

属性種別	属性	人数（人）
性別	男性	59
	女性	41
年齢層	10 代	0
	20 代	25
	30 代	38
	40 代	26
	50 代	7
	60 代	4
	70 代	0
	80 代	0
職種	購買・仕入れ	1
	調査・広告・宣伝	0
	製造・生産・品質管理	10
	経営・事務計画	3
	物流・配送	5
	技術開発・設計	10
	情報処理（システム）	14
	営業・販売	18
	広報・編集	2
	人事・総務・経理	17
	商品企画・開発	2
	基礎・技術研究	2
	個人事業主・店主	1
	企業等の経営者・役員	1
	その他	14

表 34　第三象限の群の属性情報

属性種別	属性	人数（人）
性別	男性	55
	女性	28
年齢層	10 代	0
	20 代	8
	30 代	18
	40 代	40
	50 代	14
	60 代	3
	70 代	0
	80 代	0
職種	購買・仕入れ	0
	調査・広告・宣伝	1
	製造・生産・品質管理	8
	経営・事務計画	2
	物流・配送	1
	技術開発・設計	10
	情報処理（システム）	9
	営業・販売	13
	広報・編集	1
	人事・総務・経理	17
	商品企画・開発	1
	基礎・技術研究	1
	個人事業主・店主	2
	企業等の経営者・役員	0
	その他	17

表 35　第四象限の群の属性情報

属性種別	属性	人数（人）
性別	男性	18
	女性	11
年齢層	10 代	0
	20 代	3
	30 代	10
	40 代	9
	50 代	7
	60 代	0
	70 代	0
	80 代	0
職種	購買・仕入れ	1
	調査・広告・宣伝	1
	製造・生産・品質管理	5
	経営・事務計画	2
	物流・配送	2
	技術開発・設計	2
	情報処理（システム）	0
	営業・販売	5
	広報・編集	1
	人事・総務・経理	4
	商品企画・開発	1
	基礎・技術研究	0
	個人事業主・店主	0
	企業等の経営者・役員	0
	その他	5

5.3.5.4. 行動特性に関する分析結果

各チャルディーニの法則への注意を促すアラートへの行動特性の影響を分析するために,4 つの群に分割した各群における相対反応度に有意差があるかを，各群を総当たりで，A 群，B 群として，次の帰無仮説と対立仮説により両側検定を実施した．有意水準は 5%とした．

帰無仮説：A 群と B 群では相対反応度に差がない．

対立仮説：A 群と B 群では相対反応度に差がある．

これは，例えば第一象限の群と第二象限の群とで相対反応度に有意差があるかを検証する際には，帰無仮説と対立仮説は次のようになる．

帰無仮説：第一象限の群と第二象限の群では相対反応度に差がない．

対立仮説：第一象限の群と第二象限の群では相対反応度に差がある．

実験協力者群同士の比較となり，データに対応がないため，ウィルコクソンの順位和検定により検証した．

検定の結果を表 36，表 37 に示す．第一象限の群と第二象限の群間では，相対反応度の有意差（$p<0.05$）は確認できなかったが，それ以外の組み合わせにおいては相対反応度の有意差が確認できたチャルディーニの法則が存在した．ただし，効果量 d は 0.37 から 0.93 であるため，行動特性で分割した群間における相対反応度の差は中から大[59]であることが分かる．また，これらの検出力を表 38 に整理して示す．0.8 を下回るものについては，相応の第二の過誤がありうる．以上より，「プレーンメールを開きやすいかどうか」と，「アラートに対して反応しやすいかどうか」という二種類の行動特性によって，チャルディーニアラートに対する相対反応度が異なる法則が複数あることが判明し, RQ2-2 が成り立つことが示せた．

表 36　行動特性で分割した群間での検定結果（第一象限と第二象限，第一象限と第三象限，第一象限と第四象限）

	第一象限と第二象限		第一象限と第三象限		第一象限と第四象限	
	p 値(α)	効果量 d	p 値(α)	効果量 d	p 値(α)	効果量 d
希少性	0.80	0.01	0.53	0.12	0.03*	0.46
返報性	0.16	0.19	0.01*	0.41	0.00*	0.68
権威	0.08	0.22	0.00*	0.38	0.02*	0.53
一貫性	0.50	0.15	0.01*	0.38	0.00*	0.88
好意	0.98	0.04	0.02*	0.28	0.00*	0.88
社会的証明	0.63	0.08	0.17	0.29	0.48	0.36

*:p<0.05

表 37　行動特性で分割した群間での検定結果（第二象限と第三象限，第二象限と第四象限，第三象限と第四象限）

	第二象限と第三象限		第二象限と第四象限		第三象限と第四象限	
	p 値(α)	効果量 d	p 値(α)	効果量 d	p 値(α)	効果量 d
希少性	0.68	0.12	0.05	0.43	0.13	0.31
返報性	0.00*	0.61	0.00*	0.86	0.24	0.34
権威	0.21	0.16	0.25	0.36	0.74	0.23
一貫性	0.00*	0.48	0.00*	0.93	0.07*	0.46
好意	0.03*	0.31	0.00*	0.92	0.02*	0.57
社会的証明	0.07	0.35	0.28	0.41	0.96	0.13

*:p<0.05　†：p<0.1

表 38　行動特性で分割した群間での検定結果において有意差がある結果に対する検出力

	第一象限と第二象限	第一象限と第三象限	第一象限と第四象限	第二象限と第三象限	第二象限と第四象限	第三象限と第四象限
希少性	N/A	N/A	0.59	N/A	N/A	N/A
返報性	N/A	0.83	0.90	0.98	0.98	N/A
権威	N/A	0.76	0.72	N/A	N/A	N/A
一貫性	N/A	0.77	0.99	0.88	0.99	0.53
好意	N/A	0.50	0.99	0.54	0.92	0.73
社会的証明	N/A	N/A	N/A	N/A	N/A	N/A

5.3.5.5. 性格因子と行動特性に関する分析結果

各チャルディーニの法則のアラートに対する性格因子と行動特性による影響を分析するために，行動特性で分割した 4 つの実験協力者群（第一象限（プレーンメールを開きやすく，シンプルアラートメールを開きやすい)，第二象限（プレーンメールを開きにくく，シンプルアラートメールを開きやすい)，第三象限（プレーンメールを開きにくく，シンプルアラートメールを開きにくい)，第四象限（プレーンメールを開きやすく，シンプルアラートメールを開きにくい)）に対し，それぞれ，性格因子スコアと各チャルディーニの法則のアラートの相対反応度との間で，スピアマンの順位相関係数を算出した．結果を表 39~表 42 に示す．有意水準は 5%とした．

表 39　第一象限の群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	-0.05	-0.09	-0.11	-0.03	0.04
返報性	0.02	-0.11	-0.09	-0.08	-0.10
権威	0.01	-0.07	-0.05	-0.09	-0.07
一貫性	0.06	-0.05	-0.01	-0.03	0.05
好意	-0.04	-0.04	0.04	0.01	0.06
社会的証明	-0.03	-0.06	-0.05	-0.12	-0.11

**: p<0.01, *: p<0.05, †: p<0.1

表 40　第二象限の群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	-0.12	-0.02	-0.06	0.21*	-0.01
返報性	0.02	-0.23*	-0.14	-0.30**	-0.06
権威	-0.23*	-0.03	0.15	0.09	-0.04
一貫性	-0.13	-0.11	0.11	-0.02	0.04
好意	-0.08	0.06	0.19	-0.04	-0.03
社会的証明	-0.02	-0.15	-0.10	0.20	0.12

**: p<0.01, *: p<0.05, †: p<0.1

表 41　第三象限の群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	-0.06	0.05	-0.03	-0.03	-0.10
返報性	0.03	0.02	-0.03	0.06	0.07
権威	-0.04	0.11	0.01	0.01	0.09
一貫性	-0.03	-0.16	-0.01	0.04	0.16
好意	-0.02	0.00	-0.07	0.06	0.03
社会的証明	-0.03	-0.02	-0.15	0.02	-0.11

**: p<0.01, *: p<0.05, †: p<0.1

表 42　第四象限の群における性格因子とチャルディーニの法則との相関係数

	情緒不安定性	外向性	開放性	調和性	誠実性
希少性	-0.11	0.02	0.07	0.26	0.39*
返報性	0.09	-0.16	-0.07	0.11	0.22
権威	0.03	0.11	0.07	0.21	0.20
一貫性	0.23	-0.03	-0.01	-0.30	-0.16
好意	0.38*	-0.06	-0.09	-0.18	-0.07
社会的証明	0.07	-0.07	-0.04	0.00	0.08

**: p<0.01, *: p<0.05, †: p<0.1

相関分析の結果，第二象限の群と第四象限の群で相関が確認できた.

① 第二象限の群に見られる傾向：

「情緒不安定性」と「権威の法則」との間に弱い負の相関(r =-0.23, p<0.05, 検出力=0.66)がある.「外向性」と「返報性の法則」との間に弱い負の相関(r =-0.23, p<0.05, 検出力=0.62)がある.「調和性」と「希少性の法則」との間に弱い正の相関(r =0.21, p<0.05, 検出力=0.54)がある.「調和性」と「返報性の法則」との間に弱い負の相関(r =-0.30, p<0.01, 検出力=0.87)がある.

② 第四象限の群に見られる傾向：

「情緒不安定性」と「好意の法則」との間に弱い正の相関(r =0.38, p<0.05, 検出力=0.55)がある.「誠実性」と「希少性の法則」との間に弱い正の相関(r =0.23, p<0.05, 検出力=0.56)がある.

第二象限の群である「プレーンメールでは開封せずにシンプルアラートでは開封する人」というのは，チャルディーニの法則を示すフレーズが強く効いてしまって，シンプルアラートでは開封を制止できない人（プレーンメールでは開封しないのに，チャルディーニの法則を示すフレーズメールだと開封してしまう．更に，シンプルアラートによって注意するだけでは，チャルディーニの法則を示すフレーズメールに添付されているファイルの開封を思い止まらせることができない人）ということである．

そのような人の中で，情緒不安定性の性格因子を持つ人は，権威のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封の抑止効果が高いという結果が得られた．情緒不安定性が高い人物はストレスに弱い人物であるので，権威の法則のように，「権威ある存在からのプレッシャー」に対する部分には特に敏感であることが予想される．権威のチャルディーニの法則を示すフレーズに対するアラートによって，その敏感な部分の文面を疑う気持ちが芽生え，「シンプルアラートでは制止しきれなかった添付ファイルの開封」が抑止されたのだと考えられる．

外向性の性格因子を持つ人は，返報性のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封の抑止効果が高いという結果が得られた．外向性が高い人物は社会的な人物であるので，返報性のチャルディーニの法則を示すフレーズのように「相手からの恩」に対する部分には特に敏感であることが予想される．返報性のチャルディーニの法則を示すフレーズに対するアラートによって，その敏感な部分の文面を疑う気持ちが芽生え，「シンプルアラートでは制止しきれなかった添付ファイルの開封」が抑止されたのだと考えられる．

調和性の性格因子を持つ人は，返報性のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封の抑止効果が高いという結果が得られた．調和性が高い人物は利他的な人物であるので，返報性のチャルディーニの法則を示すフレーズのように「相手からのほどこし」に対する部分には特に敏感であることが予想される．返報性のチャルディーニの法則を示すフレーズに対するアラートによって，その敏感な部分の文面を疑う気持ちが芽生え，「シンプルアラートでは制止しきれなかった添付ファイルの開封」が抑止されたのだと考えられる．

一方，調和性の性格因子を持つ人は，希少性のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封を促されてしまうという結果が得られた．シンプルアラートでは開封を抑止できないほどチャルディーニの法則を示すフレーズが効いてしまう人に対しては，希少性のチャルディーニの法則を示すフレーズに対するアラート表示は逆効果であるため，シンプルアラートと比較して開きやすくなったと考えられる．

第四象限の群である「プレーンメールでは開封するがシンプルアラートでは開封しなくなる人」というのは，シンプルアラートに含まれるチャルディーニの法則を示すフレーズ

があったとしても，シンプルアラートを提示するだけで効果がある人（プレーンメールでは開封するが，シンプルアラートによって注意するだけで，チャルディーニの法則を示すフレーズメールに添付されているファイルの開封を思い止まらせることができる人）ということである．

そのような人の中で，情緒不安定性の性格因子を持つ人は，好意のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封を促されてしまうという結果が得られた．このような人は，プレーンメールでさえ開封してしまうため，当然チャルディーニメールはそれ以上に開封してしまう人だが，シンプルアラートを表示するだけで開封しなくなる人である．しかし，好意のチャルディーニの法則を示すフレーズに対するアラート表示だけは，むしろチャルディーニの法則に対する効果が増幅され，シンプルアラートと比較して開く傾向が現れたと考えられる．

誠実性の性格因子を持つ人は，希少性のチャルディーニの法則を示すフレーズに対するアラート表示が，シンプルアラートを提示する場合よりも添付ファイル開封を促されてしまうという結果が得られた．しかし，希少性のチャルディーニの法則を示すフレーズに対するアラート表示だけは，むしろチャルディーニの法則に対する効果が増幅され，シンプルアラートと比較して開く傾向が現れたと考えられる．

今回の分析では，第一象限の群と第三象限の群ではアラートとの相関が現れなかった．

第一象限の群である「プレーンメールで開封し，シンプルアラートでも開封する人」というのは，シンプルアラートでは開封を制止できない人（プレーンメールを開封する．更に，シンプルアラートによって注意するだけでは，チャルディーニの法則を示すフレーズメールに添付されているファイルの開封を思い止まらせることができない人）ということである．このような人物は，今回のようなアラートを提示するだけではメールの開封を止めることができない．そのため，アラートについての教育を事前に実施し，アラートの意味を理解させることや，アラートの提示方法を変えることが必要になると考えられる．

第三象限の群である「プレーンメールで開封せず，シンプルアラートでも開封しない人」というのは，シンプルアラートによってメールを開封しない人（プレーンメールを開封せず，シンプルアラートによって注意するだけで，チャルディーニの法則を示すフレーズメールに添付されているファイルの開封を思い止まらせることができる人）ということである．このような人物は，チャルディーニの法則のアラートを提示することよりも，シンプルアラートを提示することのほうが有効であると考えられる．

5.4. 5 章のまとめ

本章では，個人に合わせたアラートシステムの構成を提案し，その中でも主要な構成要素である，文面からチャルディーニの法則を検出する手法を示した．さらに，個人の性格因子や行動特性に応じて表示するアラートを変更することが有効であるか，ユーザ実験に

より調査した結果を示した．今回の実験で相関が見られなかった特性を有する人物に対しては，現在有効なアラート提示手段がない．また，アラートの提示により，むしろ開封行動に結びついてしまうような，逆効果になる人間が存在しうることが分かった．

本論文では，メール文面からチャルディーニの法則を抽出する手法の検証に，大規模な実験データを利用するために，英語のメールデータセットである Enron データセットを利用した．今後，日本語のメールデータセットに対しても同様にラベル付けを実施し，英語のメール文面と同様にチャルディーニの法則が抽出できることを示したい．さらに，チャルディーニの法則を抽出したのち，ユーザの特性に合わせたアラートを提示するアラートシステムを実環境に導入し，現実の標的型メール被害を未然に防ぐことに貢献したい．

第6章　　まとめと今後の展望

6.1. まとめ

本研究では，サイバーセキュリティの Weakest Link である「人」に着目し，標的型メールの脅威分析と，その対策を示した．

2 章では不審メール（標的型メールや，特定個人を狙わずに同一の内容のメールを多数の送信先にばら撒くフィッシングメール）に対する対策と，説得に関わる心理学の研究，人を対象としたセキュリティの既存研究を示した．

3 章では，擬態精度の脅威と対策について示した．攻撃者は，OSINT ツールと AI ツールを用いて，標的者の性格因子および行動特性を推定し，標的の特性との相関関係が判明しているチャルディーニの法則を利用することで，標的に有効な文面を推定し作成することが可能である．

擬態精度への対策としては，それぞれの脅威に対して個人や企業がどれだけ情報を発信しているか，を踏まえた対策として，擬態精度は防御側でも OSINT を実施することでどれだけの情報が漏洩しているかを特定することによる注意喚起する方法を示した．

4 章では，心理操作効力の脅威を示した．ユーザ実験の結果，標的型メールの文面において，チャルディーニの法則と，個人の性格因子および行動特性との間には関係性があることが判明した．すなわち攻撃者は，OSINT ツールと AI ツールを用いて，標的者の性格因子および行動特性に有効な文面を推定可能であると結論付けることができる．

5 章では，心理操作効力を高める攻撃への対策として，個々人の特性に合わせたアラート提示を行うシステムを提案し，アンケート実験により，個々人の特性に合わせたアラート提示の有効性を評価した．実験の結果，性格因子と行動特性を考慮すると，アラートが有効である人と，そうではない人が現れたため，個々人の特性に合わせたアラートの提示を行うことが効果的であることを示した．

本研究の貢献は，「説得のされやすさ（チャルディーニの法則に対する感受性）を性格因子と行動特性のコンビネーションによって分析した」ことと，「攻撃者による脅威をあらかじめ把握することで，標的者に応じた有効な対策を配備することができることを示した」ことの 2 点である．

6.2. 巧妙な標的型メール対策の実装への課題

本節では，擬態精度と心理操作効力を高める巧妙な標的型メール対策の実装への課題を示す．

まず，擬態精度を高める攻撃対策の課題について述べる．

本論文では，攻撃者が OSINT によって収集することができる情報をもとに標的に到達する可能性があるメール文面を提示する対策を示した．このような対策は，現在既に開示している情報をもとにして，どのようなメールを攻撃者が送信することができるか，という観点からの対策である．さらに進めた対策として，インターネットに公開する前に，公開する予定の情報をもとに，どのようなメール文面が攻撃者によって送信されうるかを予測する対策が考えられる．このような方法をとることで，攻撃者に情報が渡る前に対策を考えることができ，攻撃者による被害を未然に防ぐ可能性が高まると考えられる．

次に心理操作効力を高める攻撃対策の課題について述べる．

今回の実験結果では，人の性質によってはアラートの提示が有効に働かない人や，むしろ開封行動に結びついてしまう人も現れた．これらの人物に対しては現在のアラート提示方法では有効ではなく，文言や提示の方法を変えることなどが必要であると考えられる．

今回の研究は人間の特性と標的型メールの開封意図への反応や，アラートへの反応との関係性を一部明らかにすることができた．今後は，これらの関係性の分析を，性格因子や行動特性のみならず，他の観点からも研究を進めたい．

さらに，アラートの提示方法も重要になると考えられる．説得心理学の防衛機構や予告の性質を踏まえて，どのようにアラートを提示することが有効に働くかを検討したい．さらに，実際の組織での運用を見据えた仕組みづくりも検討していきたい．これは，アラートを提示する前に，組織の人物に対してアラートの意味を教育することで効果が変わるか，といった検討をしていきたい．

本論文では，メール文面からチャルディーニの法則を抽出する手法の検証に，大規模な実験データを利用するために，英語のメールデータセットである Enron データセットを利用した．今後，日本語のメールデータセットに対しても同様にラベル付けを実施し，英語のメール文面と同様にチャルディーニの法則が抽出できることを示したい．

チャルディーニの法則はもともと営業の分野で提唱された法則であり，説得を試みる際には正常なやりとりにおいても現れるものである．そのため，正常なやりとりにおいてもチャルディーニの法則が付与されていることだけを根拠にアラートをあげてしまうと，誤検知が多くなりアラートの見逃しに繋がってしまう．今後は，実際に攻撃で用いられるメールを入手して分析し，正常なメール文面と攻撃で用いられるメール文面とで，用いられるチャルディーニの法則の傾向を分析し，ユーザに対して正確にアラートを提示できるように検討を進めたい．

今回の実験では，を示すために 5 章の RQ2「個々人で有効なチャルディーニの法則への

注意を促すアラートは，性格因子と行動特性の両者によって異なる傾向を示すのか」を示すための分析手法として，(RQ2-1)性格因子によって各アラートの効果が異なるか，に対してはスピアマンの順位相関係数，(RQ2-2)行動特性によって各アラートの効果が異なるか，に対してはウィルコクソンの順位和検定，(RQ2-3)性格因子と行動特性の 2 軸によって各アラートの効果が異なるか，に対してスピアマンの順位相関係数を利用した．これらの分析の妥当性を示すために，他の分析方法によっても今回の結果と同じ結論が得られることを示す必要がある．適用する分析としては，二元配置分散分析により，性格因子と行動特性の交互作用効果が認められるかを調べることや，検証的因子分析や共分散構造分析によりチャルディーニの各法則の因子が，性格因子と行動特性であることや，このモデルの妥当性を検証したい．

今回の実験結果の一般性を論ずるために，他の企業や国でも同様の結果が得られるのかを検証していく．具体的には，企業ごとの比較として，実験協力者を収集する際に，実験協力者が所属する企業規模や業種が異なるように収集し，本研究と同様のアンケート実験を行うことで，企業間の比較を実施する．国ごとの比較としては，いくつかの国を対象に，英語のメール文面によってアンケート実験を行うことで，英語に対しての反応の一般性を論じることができる．さらに，実験対象とする国の母国語によるメール文面も作成し，英語のみならず，母国語によるメールに対する反応を検証することで，実験結果の一般性を論ずることが期待できる．

6.3. 今後の展望

企業においては現在もメールは多く利用されており，標的型メール攻撃の脅威は依然として考慮する必要がある．一方で昨今ではチャットツールを導入する企業もあり，大手企業においてのコミュニケーションのうち，14.2%がチャットツールによって行われているという調査結果がある[81]．スピーディにコミュニケーションができるようになった，という回答が 41.9%あるなど，チャットツールの利便性が示唆されており，チャットツールを利用したコミュニケーションは今後増えていくと考えられる．攻撃者もメールだけではなくチャットツールを駆使することが考えられるため，今後はメールだけではなく，チャットツールにおいてもソーシャルエンジニアリングを考慮する必要があると考えられる．

さらに，対話型のチャットボットなど，AI の進歩により，あたかも人間とやりとりをしているかのようなツールの開発が進んでいる[83]．今後このようなツールの発展により，ソーシャルエンジニアリングの対話部分すらも自動化される危険性があり，「人」を守ることがますます重要になると考えられる．

さらに今後，リモートワークなどによって，従業員の端末を利用する BYOD が進むことも考えられる．そのため，従業員がばらまき型の攻撃に対しても引っかからないように注意する必要がある．ばらまき型の手法の一つとして，フィッシングメールにおけるアラートのようなメール（Amazon のアカウントが停止しました，など）が考えられる．このようなメールが到達した際にも，アラートに対する反応をあらかじめ特定しておくことで，偽アラートに対して注意を促すことで，フィッシングメールに気づくことが期待される．

今後，標的型攻撃において人を対象とした攻撃がなくなることはなく，その手口も発展していくと考えられる．本研究によって，攻撃者による脅威をあらかじめ把握することで人への有効な対策を提案することに貢献することができた．本研究は，著者が知る限り，個々人の特性に応じてセキュリティシステムの防御方法を変える考えを具体化した初めての研究である．本研究が，先に述べた，Weakest Link である人に着目した研究の足掛かりとなることを期待する．

参考文献

[1]. 警察庁：令和２年上半期におけるサイバー空間をめぐる脅威の情勢等について，（オンライン），入手先⟨https://www.npa.go.jp/publications/statistics/cybersecurity/data/R02_kami_cyber_jousei.pdf⟩ (参照 2020-11).

[2]. 日本経済新聞：標的型メール攻撃，過去最多ペース，上半期 3900 件,（オンライン），入手先⟨https://www.nikkei.com/article/DGXMZO64479980R01C20A0CR8000/⟩（参照 2020-11).

[3]. 内閣サイバーセキュリティセンター：日本年金機構における個人情報流出事案に関する原因究明調査結果,（オンライン），入手先⟨https://www.nisc.go.jp/active/kihon/pdf/incident_report.pdf⟩, (参照 2020-11).

[4]. 総務省：サイバー攻撃（標的型攻撃）対策防御モデルの解説,（オンライン），入手先⟨https://www.soumu.go.jp/main_content/000495298.pdf⟩, (参照 2020-11).

[5]. 情報処理推進機構：2020 情報セキュリティ白書，情報処理推進機構 (2020).

[6]. 情報処理推進機構：サイバー情報共有イニシアティブ（J-CSIP（ジェイシップ））,（オンライン），入手先⟨https://www.ipa.go.jp/security/J-CSIP/index.html⟩ (参照 2020-11)

[7]. 情報処理推進機構： サイバーレスキュー隊 J-CRAT（ジェイ・クラート），（オンライン），入手先⟨https://www.ipa.go.jp/security/J-CRAT/index.html⟩ (参照 2020-11)

[8]. TrendMicro： COMBATING MALICIOUS EMAIL AND SOCIAL ENGINEERING ATTACK METHODS, （オンライン），入手先⟨https://www.trendmicro.tw/cloud-content/us/pdfs/business/datasheets/ds_social-engineering-attack-protection.pdf⟩ (参照 2021-2)

[9]. MITRE： ATT&CK, MITRE（オンライン），入手先⟨https://attack.mitre.org/⟩ (参照 2020-11)

[10]. TrendMicro： 標的型サイバー攻撃, TrendMicro （オンライン），入手先⟨https://www.trendmicro.com/ja_jp/security-intelligence/research-reports/threat-solution/apt.html⟩ (参照 2020-11)

[11]. LAC：Cyber GRID View vol.1 日本における標的型サイバー攻撃の事故実態調査レポート,（オンライン），入手先⟨https://www.lac.co.jp/lacwatch/pdf/20141216_cgview_vol1_d001t.pdf⟩ (参照 2020-11)

[12]. National Institute of Standards and Technology： Computer Security Incident Handling Guide(SP800-61r2), （オンライン），入手先⟨https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf⟩ (参照 2020-11)

[13]. 情報処理推進機構： 標的型攻撃／新しいタイプの攻撃の実態と対策, （オンライン），

入手先⟨https://www.ipa.go.jp/files/000024542.pdf⟩ (参照 2020-11)
[14].クリストファー・ハドナジー (著), 成田光彰 (訳)ソーシャル・エンジニアリング，日経 BP（2012）.
[15].深田 博己：説得心理学ハンドブック―説得コミュニケーション研究の最前線，北大路書房（2002）.
[16].Rajivan, P., & Gonzalez, C.：Creative persuasion: a study on adversarial behaviors and strategies in phishing attacks. Frontiers in psychology, 9, 135 (2018).
[17].Goel, S., Williams, K., & Dincelli, E.：Got phished? Internet security and human vulnerability. Journal of the Association for Information Systems, 18(1), 2 (2017).
[18].Acquisti, A., Gross, R. and Stutzman, F.D.: Face recog- nition and privacy in the age of augmented reality, Jour- nal of Privacy and Confidentiality, Vol.6, No.2, pp.1–20 (2014).
[19].Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S. and Dabbish, L.D.: Anonymity, privacy, and security online, pp.1–35, Pew Research Center (2013).
[20].Ball, L.D., Ewan, G. and Coull, N.J.: Undermining- social engineering using open source intelligence gather- ing, Proc. 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012), pp.275–280, SciTePress-Science and Technology Publications (2012).
[21].Goldberg, L.R.: An alternative “description of person- ality”: The big-five factor structure, Journal of person- ality and social psychology, Vol.59, No.6, pp.1216–1229 (1990).
[22].Rothmann, S. and Coetzer, E.P.: The big five personality dimensions and job performance, SA Journal of Industrial Psychology, Vol.29, No.1, pp.68–74 (2003).
[23].IBM:Personality Insights, （オンライン），入手先⟨ https://www.ibm.com/jp-ja/cloud/watson-personality-insights>(参照 2020-11).
[24].ロバート・B・チャルディーニ(著), 社会行動研究会(訳): 影響力の武器[第三版]:なぜ，人は動かされるのか, 誠信 書房 (2014).
[25].Wright, R.T., Jensen, M.L., Thatcher, J.B., Dinger, M. and Marett, K.: Research note – influence techniques in phishing attacks: An examination of vulnerability and resistance, Information systems research, Vol.25, No.2, pp.385–400 (2014).
[26].Alkı ̧s, N. and Temizel, T.T.: The impact of individual differences on influence strategies, Personality and Individual Differences, Vol.87, pp.147–152 (2015).
[27].Modic, D., Anderson, R. and Paloma ̈ki, J.: We will make you like our research: The development of a susceptibility-to-persuasion scale, PloS ONE, Vol.13, No.3, pp.1–21 (2018).
[28].NTT TechnoCross Corporation：CipherCraft/Mail, （オンライン），入手先⟨https://www.ntt-tx.co.jp/products/ccraftmailtypeh/⟩ (参照 2020-11).
[29].Symantec：About Disarm, （オンライン），入手先⟨https://help.symantec.com/cs/SMG_10_6_6/SMG/v85068372_v125807409/Acerca-de-la-

desactivaci?locale=EN_US〉(参照 2020-11).

[30]. Duman, S., Kalkan-Cakmakci, K., Egele, M., Robertson, W., & Kirda, E. ：Emailprofiler: Spearphishing filtering with header and stylometric features of emails. In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC) (Vol. 1, pp. 408-416). IEEE.(2016)

[31]. Gascon, H., Ullrich, S., Stritter, B., & Rieck, K.：Reading between the lines: content-agnostic detection of spear-phishing emails. In International Symposium on Research in Attacks, Intrusions, and Defenses (pp. 69-91). Springer, Cham.(2018)

[32]. Ho, G., Cidon, A., Gavish, L., Schweighauser, M., Paxson, V., Savage, S., Voelker, G., & Wagner, D.：Detecting and characterizing lateral phishing at scale. In 28th USENIX Security Symposium (USENIX Security 19) pp. 1273-1290 (2019).

[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, 58(12), 1890-1900 (2017).

[34]. 寺田剛陽, 津田宏, 片山佳則, & 鳥居悟. ：IT 被害に遭いやすい心理的・行動的特性に関する調査. マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, 2014, 1498-1505 (2014).

[35]. 片山佳則, 寺田剛陽, 鳥居悟,& 津田宏：ユーザー行動特性分析による個人と組織の IT リスク見える化の試み. SCIS (Symposium on Cryptography and Information Security)2015, 4D1-3 電子情報通信学会 (2015).

[36]. 片山佳則, 寺田剛陽, 鳥居悟, & 津田宏. ：利用者の行動特性分析に基づくセキュリティリスク判定技術の試作. 人工知能学会全国大会論文集 第 29 回全国大会 (2015) (pp. 4H13-4H13). 一般社団法人 人工知能学会(2015).

[37]. Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A.：Breaching the human firewall: Social engineering in phishing and spear-phishing emails. arXiv preprint arXiv:1606.00887 (2016).

[38]. Williams, E. J., Hinds, J., & Joinson, A. N.：Exploring susceptibility to phishing in the workplace. International Journal of Human-Computer Studies, 120, 1-13 (2018).

[39]. Best, C.：OSINT, the Internet and Privacy. EISIC (2012).

[40]. Edwards, M., Larson, R., Green, B., Rashid, A., & Baron, A.：Panning for gold: Automatically analysing online social engineering attack surfaces. Computers & Security (2016).

[41]. Silic, M., & Back, A.：The dark side of social networking sites: Understanding phishing risks. Computers in Human Behavior, 60, 35-43 (2016).

[42]. Singh, A., Thaware, V.：WIRE ME THROUGH MACHINE LEARNING, Black Hat USA 2017, Black Hat (2017).

[43]. 岩田一希, 中村嘉隆, 稲村浩, 高橋修.：標的型メール攻撃対策訓練における訓練メール自動生成のための受信メール分析手法の検討. マルチメディア, 分散協調とモバイ

ルシンポジウム 2016 論文集, 819-825 (2016).
[44]. IntelTechniques : Buscador Investigative Operating System, （オンライン）, 入手先⟨https://inteltechniques.com/buscador/⟩ (参照 2020-11)
[45]. 奥村紀之, 金丸裕亮, & 奥村学. : 感情判断と Big Five を用いたブログ著者の性格推定に関する調査. 人工知能学会全国大会論文集, 29, 1-4 (2015).
[46]. IntelTechniques, Updated OSINT Flowcharts, （オンライン）, 入手先⟨https://inteltechniques.com/blog/2018/03/06/updated-osint-flowcharts/⟩ (参照 2020-11)
[47]. Justin, N. : OSINT Framework, （オンライン）, 入手先⟨https://osintframework.com/⟩ (参照 2020-11)
[48]. Radu, T. : 8 best email extractor software for Windows 10, windowsreport（オンライン）, 入手先⟨https://windowsreport.com/email-extractor-software/⟩ (参照 2020-11)
[49]. Maldevel : EmailHarvester, GitHub（オンライン）, 入手先⟨https://github.com/maldevel/EmailHarvester⟩ (参照 2020-11)
[50]. Johnny L. : Google Hacking for Penetration Testers, Blackhat（オンライン）, 入手先⟨https://www.blackhat.com/presentations/bh-europe-05/BH_EU_05-Long.pdf⟩ (参照 2020-11)
[51]. BISHOPFOX : Google Hacking Diggity Project, （オンライン）, 入手先⟨http://www.bishopfox.com/resources/tools/google-hacking-diggity/⟩ (参照 2020-11)
[52]. pipl, pipl, （オンライン）, 入手先⟨https://pipl.com/⟩ (参照 2020-11)
[53]. Maltego, Maltego CE, （オンライン）, 入手先⟨https://www.maltego.com/products/⟩ (参照 2020-11)
[54]. Kali Tools, Metagoofil Package Description , （オンライン）, 入手先⟨https://tools.kali.org/information-gathering/metagoofil⟩ (参照 2020-11)
[55]. Ioannis, K. : Creepy, GitHub（オンライン）, 入手先⟨https://github.com/ilektrojohn/creepy⟩ (参照 2020-11)
[56]. Tinfoleak, TINFOLEAK.COM,（オンライン）, 入手先⟨https://tinfoleak.com/⟩ (参照 2020-11)
[57]. Akbar, N.: Analysing persuasion principles in phishing emails, Master's thesis, University of Twente, pp.1–105(2014).
[58]. Egelman, S., & Peer, E. : The myth of the average user: Improving privacy and security systems through individualization. Proceedings of the 2015 New Security Paradigms Workshop pp. 16-28 (2015).
[59]. Cohen, J.: A power primer, Psychological bulletin, Vol.112, No.1, pp.155–159 (1992).
[60]. 和田さゆり:性格特性用語を用いた Big Five 尺度の作成, 心理学研究, Vol.67, No.1, pp.61‐67 (1996).
[61]. 並川 努, 谷 伊織, 脇田貴文, 熊谷龍一, 中根 愛, 野口 裕之:Big Five 尺度短縮版の開

発と信頼性と妥当性の検 討, 心理学研究, Vol.83, No.2, pp.91‐99 (2012).
[62]. 日本サイバー犯罪対策センター(JC3):日本サイバー犯罪 対策センター(JC3)(オンライン), 入手先 （オンライン），入手先⟨https://www. jc3.or.jp/index.html⟩(参照 2020-11).
[63]. Lancers:Lancers, （オンライン），入手先⟨https://www. lancers.jp/⟩⟩ (参照 2020-11).
[64]. Carifio, J. and Rocco, P.: Resolving the 50-year debate around using and misusing Likert scales, Medical educa- tion, Vol.42, No.12, pp.1150–1152 (2008).
[65]. Bhakta, R., Harris, I.：Semantic analysis of dialogs to detect social engineering attacks, Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (2015).
[66]. Yuki Sawa, Ram Bhakta and Ian G. Harris：Detection of Social Engineering Attacks Through Natural Language Processing of Conversations, 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)
[67]. William W. Cohen：Enron Email Dataset, Carnegie Mellon University（オンライン），入手先⟨https://www.cs.cmu.edu/~./enron/⟩ (参照 2020-11).
[68]. EnronData.org：EnronData, （オンライン），入手先 ⟨https://enrondata.readthedocs.io/en/latest/⟩ (参照 2020-11).
[69]. Amazon Mechanical Turk：Amazon Mechanical Turk , （オンライン），入手先 ⟨https://www.mturk.com/⟩ (参照 2020-11).
[70]. Google：word2vec, （オンライン），入手先⟨https://code.google.com/archive/p/word2vec/⟩ (参照 2020-11).
[71]. Python：Python,（オンライン），入手先⟨https://www.python.org/⟩ (参照 2020-11).
[72]. Scikit-learn.org：scikit-learn Machine Learning in Python, （オンライン），入手先 ⟨https://scikit-learn.org/⟩ (参照 2020-11).
[73]. Gensim：Genism topic model for humans, （オンライン），入手先 ⟨https://radimrehurek.com/gensim/⟩ (参照 2020-11).
[74]. Preferred Networks：Chainer: A flexible framework for neural networks, （オンライン），入手先⟨https://chainer.org/⟩ (参照 2020-11).
[75]. Van, D, H, A, and Luca, A.: Cognitive tiaging of Phshing Attack. 28th Usenix Security Symposium, 1309-1326 (2019).
[76]. 東京大学教養学部統計学教室: 統計学入門, 東京大学出版会(1991)
[77]. 東京大学教養学部統計学教室: 自然科学の統計学, 東京大学出版会(1992)
[78]. 山田 剛史: R によるやさしい統計学, オーム社(2008)
[79]. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean.：Distributed Representations of Words and Phrases and their Compositionality, Proceedings of NIPS (2013).
[80]. Oppenheimer, D. M., Meyvis, T., & Davidenko, N.：Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of experimental social psychology, 45(4), 867-872 (2009).
[81]. 伊藤忠テクノソリューションズ株式会社：大手企業のビジネスチャットツールの利用

状況調査, （オンライン）, 入手先⟨https://www.ctc-g.co.jp/news/press/doc/20190919a_detail.pdf⟩ (参照 2020-11)
[82]. 八藤後茉央, 高田豊雄, 小倉加奈代:人間の脆弱性を利用した標的型攻撃への防御手法の検討, 第 79 回全国大会講演論文集, Vol.2017, No.1, pp.605‐606 (2017).
[83]. Wu, Xianchao, et al. "りんな: 女子高生人工知能." 言語処理学会 2016 年次大会. 仙台, 日本 (2016).
[84]. 西川弘毅, 上原航汰, 山本匠, 河内清人, 西垣正勝：標的型メールにおける心理操作テクニックと性格特性および行動特性との関係性分析, 情報処理学会論文誌, Vol.61, No.3, pp.591‐607(2020)
[85]. いらすとや：いらすとや, （オンライン）, 入手先⟨https://www.irasutoya.com/⟩(参照 2021-2)

謝辞

本研究を進めるにあたり，指導教員としてご丁寧，的確，かつ，きめ細やかなご指導を賜り，常に励まし続けてくださった静岡大学 西垣正勝教授に心より御礼申し上げます.
博士論文審査委員として，事前審査・本審査を通じて，本論文に関して数多くのご助言をしてくださった，静岡大学 竹内勇剛教授，近藤淳教授，大木哲史准教授に深謝申し上げます.

本研究を進めるにあたり，静岡産業大学 漁田武雄教授には認知心理学の観点からご助言をいただきました.御礼申し上げます

会社の先輩であり，入社当時から研究や業務について親身にご指導いただき，研究活動を指導いただいた三菱電機株式会社 情報技術総合研究所　山本匠殿に，心から感謝いたします.
本研究の立ち上げから共に研究に取り組み，多大に尽力いただいた西垣研究室の元学生である上原航汰殿に，深く感謝いたします.

実験環境の構築方法の助言をいただきました西垣研究室の北川沢水殿に感謝申し上げます.

企業に属する身でありながら日々の業務と並行しての学位取得について，ご高配を賜った三菱電機株式会社　中川路哲男殿，米田健殿，河内清人殿に，深く感謝いたしますとともに，心から御礼申し上げます.

大学時代の恩師である東京理科大学 岩村惠市教授には，研究者としての基礎をご指導いただきました．深く感謝いたします.

最後に，いつも暖かく見守り応援してくれる両親，弟，祖父母，叔母に感謝します.

発表論文等

A 学位論文申請資格に関わる論文

1) 西川弘毅, 上原航汰, 山本匠, 河内清人, 西垣正勝：標的型メールにおける心理操作テクニックと性格特性および行動特性との関係性分析, 情報処理学会論文誌, Vol.61, No.3, pp.591‐607(2020)

B 学位論文内容に関わる論文（未発表論文も含む）

1) **Hiroki Nishikawa**, Takumi Yamamoto, Bret Harsham, Ye Wang, Kota Uehara, Chiori Hori, Ayako Iwasaki, Kiyoto Kawauchi, Masakatsu Nishigaki: Analysis of Malicious Email Detection using Cialdini's Principles. In 2020 15th Asia Joint Conference on Information Security (AsiaJCIS) (pp. 137-142). IEEE.

2) Kota Uehara, **Hiroki Nishikawa**, Takumi Yamamoto, Kiyoto Kawauchi, Masakatsu Nishigaki: Analysis of the relationship between psychological manipulation techniques and personality factors in targeted emails, Proceedings of 2019 International Conference on Broad-Band Wireless Computing, Communication and Applications, pp.338-351 (2019.10)

3) Kota Uehara, **Hiroki Nishikawa**, Takumi Yamamoto, Kiyoto Kawauchi, Masakatsu Nishigaki, Analysis of the Relationship between Psychological Manipulation Techniques and Personality Factors in Targeted Emails(BWCCA2019).

4) Kota Uehara, Kohei Mukaiyama, Masahiro Fujita, **Hiroki Nishikawa**, Takumi Yamamoto, Kiyoto Kawauchi, Masakatsu Nishigaki: Basic Study on Targeted E-mail Attack Method Using OSINT, Proceedings of 2019 International Conference on Advanced Information Networking and Applications, pp.1329-1341 (2019.3).

C その他の論文

なし

D 口頭発表など

1) 山本匠, Bret Harsham, Ye Wang, 西川弘毅, 上原航汰, Chiori Hori, 岩崎亜衣子, 河内清人, 西垣正勝：メールにおける誘導手口の推定手法に関する検討, 情報処理学会研究報告, 2019-CSEC-86-63, pp.1-7 (2019.7).

2) 西川弘毅, 山本匠, 上原航汰, 西垣正勝, 河内清人：標的型メールにおける誘導手口の考察, 情報処理学会研究報告, 2019-SPT-32-34, pp.1-6 (2019.3).

3) 西川弘毅, 山本匠, 河内清人, 西垣正勝：チャットボットを用いた犯罪手口の収集・通知システムの提案, 暗号と情報セキュリティシンポジウム2019予稿集, 2F2-5 (2019.1).
4) 西川弘毅, 上原航汰, 山本匠, 河内清人, 西垣正勝：インテリジェンスを利用する標的型メールと標的型メールに対するインテリジェンスを利用した防御に関する検討, コンピュータセキュリティシンポジウム2018論文集, pp.654-658 (2018.10).
5) 上原航汰, 井上佳祐, 本多俊貴, 西川弘毅, 山本匠, 河内清人, 西垣正勝： OSINTと人間の心理を利用した標的型メール攻撃に対するインテリジェンスを活用した防御に関する基礎検討, コンピュータセキュリティシンポジウム2018論文集, pp.647-653 (2018.10).
6) 西川弘毅, 山本匠, 河内清人, 西垣正勝：攻撃者のメール送信状態推定による不審メール検知技術の提案, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム2018論文集, pp.1298-1302 (2018.7).