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

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

上原 航汰[†1]　井上 佳祐[†1]　本多 俊貴[†1]
西川 弘毅[†1,2]　山本 匠[†2]　河内 清人[†2]　西垣 正勝[†1]

概要：近年は，企業や個人に関する多くの情報が OSINT（Open Source Intelligence）によって容易に入手できる状況にあり，標的型メール攻撃をはじめとするソーシャルエンジニアリングの脅威は従前よりも格段に大きくなっている．OSINT は標的型メールを正規のメールに擬態させるための情報収集のために行われているが，標的型メールの受信者を心理的に誘導するためにも OSINT が利用できることが分かってきた．すなわち，攻撃者による OSINT の悪用は，標的型メールの擬態精度・心理操作効力を高めるために非常に有効であり，攻撃者がより巧妙な標的型メールを作成するにあたっての大きな手段となる．しかし，OSINT は攻撃者だけでなく，正規ユーザ（防御側）にも恩恵を与えるものである．攻撃者が攻撃力を高めるために標的者の情報を利用するように，防御側も自分自身の情報を積極的に利用して防御力を高める方法を確立することによって，OSINT を悪用する攻撃者に対抗する手段と成すことができる．企業や個人が外部に公開している情報は，彼ら自身に関する全情報の内の一部に過ぎない．したがって，インテリジェンス（OSINT）を悪用する攻撃者に対して，防御側もインテリジェンスを持って対抗するアプローチは，防御側が有利な結果となることが期待される．その具現化が本研究の目的である．本稿では，OSINT 攻撃者が用いる攻撃法をそのまま防御法として逆用する方法に焦点を当てる．具体的には，攻撃者が OSINT を利用してより巧妙な標的型メールを作成する手順について論じた上で，その攻撃手順を活用して標的型メールを検知・峻別する方法を示す．

キーワード：ソーシャルエンジニアリング，標的型メール，OSINT，チャルディーニの法則，主要 5 因子性格検査

1. はじめに

近年，標的型メール攻撃の被害が急増している．標的型メール攻撃はソーシャルエンジニアリングの典型例の一つであり，標的者を騙すことで対象者に被害を与える（例えば，情報や金銭を搾取する，PC を不正に操る）．標的型メール攻撃を成功させるためには，標的者に標的型メールを正規のメールと信じ込ませることが不可欠であり，このため攻撃者は，より巧妙な標的型メールを作成しようと試みる．

今日では，企業やユーザが自身の情報をオウンドメディアやソーシャルメディアに自ら発信することが当たり前の時代になってきた．今や，Web 上や SNS 上には企業や個人に関する情報が氾濫しており，公開されているパーソナル情報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可能であるとも報告されている[1][2]．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 ツールを利用して標的者の所属組織・上司・友人の名前・メールアドレス・出来事・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，擬態精度の高い標的型メール（標的者にとって正規のメールとの区別がつきにくい標的型メール）を標的者ごとに作成することが可能である[3][4]．

更に，近年の研究からは，SNS 上の情報がユーザの心理的な傾向や情動をも推測し得ることが判明してきており，ツイートやブログ記事からユーザの主要 5 因子（ビッグファイブ）性格検査を実施する AI の実運用が既に始まっている[5]．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[6][7]ことが知られており，その際，ユーザの性格因子に応じて心理操作による誘導の受けやすさが異なる[8]ことが報告されている．以上より，攻撃者は，OSINT ツールを利用して標的者の SNS 情報を取得し，上述の性格検査 AI や心理操作テクニックを悪用することによって，心理操作効力の高い標的型メール（標的者が誘導されやすい標的型メール）を標的者ごとに作成することが可能である．

このように，攻撃者による OSINT の悪用は，標的型メールの擬態精度・心理操作効力を高めるために非常に有効であり，攻撃者がより巧妙な標的型メールを作成するにあたっての大きな手段となる．しかし，OSINT は攻撃者だけでなく，正規ユーザ（防御側）にも恩恵を与えるものであり，本来，攻撃者のみが有利になるものではない．攻撃者が攻撃力を高めるために標的者の情報を利用するように，防御側も自分自身の情報を積極的に利用して防御力を高める方法を確立することによって，OSINT を悪用する攻撃者に対抗する手段と成すことができる．その具現化が本研究の目的である．本研究では，企業やユーザが各々有している全情報の活用を「ASINT（All Source Intelligence）」と呼称する．

一般的に，企業やユーザが所有している情報の中には非公開情報（社外秘情報・プライバシ情報）も多く含まれており，企業やユーザが外部に公開している情報は，彼ら自身に関する全情報の内の一部に過ぎない．すなわち，攻撃

†1 静岡大学
Shizuoka University
†2 三菱電機株式会社
Mitsubishi Electric Corporation

者が OSINT によって入手できる標的者の情報は, 標的者が ASINT によって利用できる全情報中の部分集合である．したがって，インテリジェンス（OSINT）を悪用する攻撃者に対して，防御側もインテリジェンス（ASINT）を持って対抗するアプローチは，必ず防御側が有利な結果となることが期待される．

ASINT を利用した防御は，非公開情報（OSINT では取得できない情報）を利用して OSINT 攻撃を防ぐ方法と，情報量で勝る ASINT の利を活かして OSINT 攻撃者が用いる攻撃法をそのまま防御法として逆用する方法に大別できる．本稿では，後者に焦点を当て，OSINT 攻撃者が利用する攻撃法を ASINT 防御に逆用する方法について論ずる[a]．具体的には, 攻撃者が OSINT を利用してより巧妙な標的型メールを作成する手順（以下，攻撃手順）について論じた上で，その攻撃手順を活用して標的型メールを検知・峻別する方法を示す. 攻撃者が OSINT を利用して攻撃手順を実行することによって「標的者が騙されるであろう標的型メール」を作成するのに対し，防御側は ASINT を利用して攻撃手順を実行することによって「攻撃者が作成するであろう標的型メール」を推測する．攻撃者と防御側は同一の攻撃手順を用いるが, 防御側（ASINT）の情報量が攻撃者（OSINT）の情報量を凌駕するため，防御精度が攻撃精度を上回ると考えられる．

2. 課題

2.1 関連研究

OSINT を利用した標的型メール攻撃に関する先行研究を，擬態精度と心理操作効力の観点から概説する．前者については，OSINT によってソーシャルエンジニアリングの脅威が高まることが文献[4]にて報告されている．著者らも文献[3]にて，攻撃者が擬態精度の高い標的型メールを作成するにあたって，OSINT ツールを活用して標的者の情報を芋づる式に取得していく様子を状態遷移図としてモデル化した．後者については，OSINT を利用した標的者の性格推定に関する先行研究と，その性格に応じた心理操作に関する先行研究に大別される．

OSINT を利用した標的者の性格推定に関する先行研究については，SNS に投稿された情報からユーザの性格の主要 5 因子（ビッグファイブ）を推定する研究が盛んに行われており，その顕著な成果として IBM Watson Personality Insights が既に実運用されている[5]. Personality Insights は，ユーザの Twitter 上のツイートや, その人物が執筆したブログ等の記事を入力することによって，当該ユーザのビッグファイブのスコアを出力する[5]．既往研究より，心理テストを通じて測定された被験者のビッグファイブと当該被験者が書いた文章の間には相関関係があることが知られている．Personality Insights は，数千人に及ぶユーザに対してビッグファイブに関する心理テスト(アンケート)を実施し，各ユーザのアンケート結果と SNS 文書・SNS 投稿（ツイートあるいはブログ記事)の相関を機械学習した AI である．Personality Insights は API モジュールとして公開されており，ユーザの SNS 文章を入力すると当該ユーザのビッグファイブのスコアを出力する．

ユーザの性格に応じた心理操作に関する先行研究については，文献[7][9]においてフィッシングメールにおけるチャルディーニの法則（詳細については 3.1 節にて説明する）の悪用に関する研究が行われている．Wright らは．被験者である大学生の集団に，チャルディーニの 6 つの法則を利用したフィッシングメールと，チャルディーニの法則を利用していないメールを送り，その反応率（フィッシングメールに書かれている指示に従ってしまう割合）の違いを比較する実験を行った[7]．実験の結果，チャルディーニのどの法則を利用したフィッシングメールも，利用していないメールと比較して被験者の反応率が高く，フィッシングメールにおけるチャルディーニの法則の効果が確認された結果となった．Akbar らは，メールの本文中にチャルディーニの各法則が組み込まれているか否かを判別するためのフローチャートを開発し，チャルディーニの法則が利用されているフィッシングメールが実際にどの程度存在するのか調査を行った[9]．その結果，Akbar が調査したフィッシングメールデータセット中 96.1%に「権威」の法則が，41.1%に「希少性」の法則がそれぞれ用いられており，また，その他の法則についても高い割合でフィッシングメールに用いられていることが明らかとなった．Akbar の調査は，現在のフィッシングメールにおいてチャルディーニの法則が広く利用されていることを示したが，標的型メールにもチャルディーニの法則が有効に作用することは容易に予想される．

また，ユーザの性格因子に応じてチャルディーニの法則の影響の受けやすさが異なることも，既往研究から明らかになっている[8]．Alkış らは，チャルディーニの法則は万人に通用する法則だと述べた上で，その影響の「受けやすさの度合い」は各ユーザの性格因子に左右されるとし，検証実験を行った．具体的には，ビッグファイブに関する心理テスト（アンケート）を被験者に実施し，その後チャルディーニの法則に対する反応率を捕捉するためのアンケート（例えば希少性に関しては,「珍しい製品（数が少ないもの）が大量生産品よりも価値があると思うか」という質問に対し，1（強く同意する）～5（強く同意しない）の 5 段階で回答してもらう）を行い，各ユーザのビッグファイブの各スコアとチャディーニの各法則に対する反応率の相関を調べた．その結果，ビッグファイブとチャルディーニの各法則の反応率の間に，いくつか有意な相関があることが明らかになった．

a 前者の「非公開情報（OSINT では取得できない情報）を利用して OSINT 攻撃を防ぐ方法」については，著者らによる別の発表にて取り扱う．

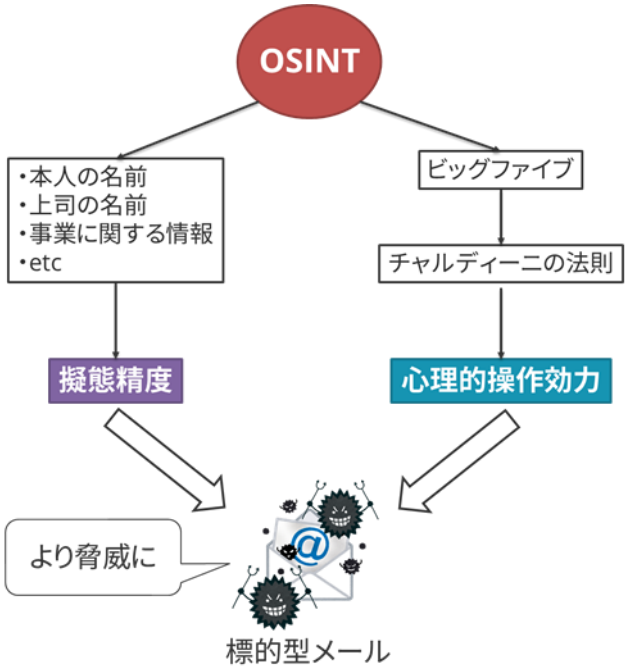

図 1　OSINT を利用した標的型メール攻撃の増進

2.2 本稿の位置付け

2.1 節で説明した先行研究の結果を集約すると，「攻撃者が OSINT ツールを利用して標的者の SNS 情報を収集し，Personality Insights を利用して標的者の性格因子を推測した上で，標的者に効果的なチャルディーニの法則を組み込んだ標的型メールを作成する」というシナリオが暗示される．著者らは文献[3]において，OSINT による標的型メールの擬態精度の増進に対する警戒を促したが，現在の AI 技術の進歩により，OSINT は標的型メールの心理操作効力を高めるための手段としても活用可能である（図 1）．

このように，OSINT は標的型メール攻撃の脅威を深刻化させる．この問題に対し，防御側も自分自身の情報を積極的に利用して防御力を高める方法を確立することが本研究の目的である．本研究では，企業やユーザが各々有している全情報の活用を「ASINT（All Source Intelligence）」と呼称する．本稿では，特に心理操作効力の観点に焦点を当て，OSINT 攻撃者が利用する攻撃法を防御に逆用する方法について論ずる．防御側が ASINT を利用して攻撃手順を実行することによって，「攻撃者が OSINT を利用して作成する標的型メール」を推測することが可能となる．攻撃者が OSINT によって入手できる標的者の情報は，標的者が ASINT によって利用できる全情報中の部分集合であるため，防御精度が攻撃精度を上回ることが期待される．以降では，攻撃者が OSINT を利用して標的型メールの心理操作効力を増進させる手順（以下，攻撃手順）ついて説明した後（3 章），その攻撃手順を活用して標的型メールを検知・峻別する方法を示す（4 章）．

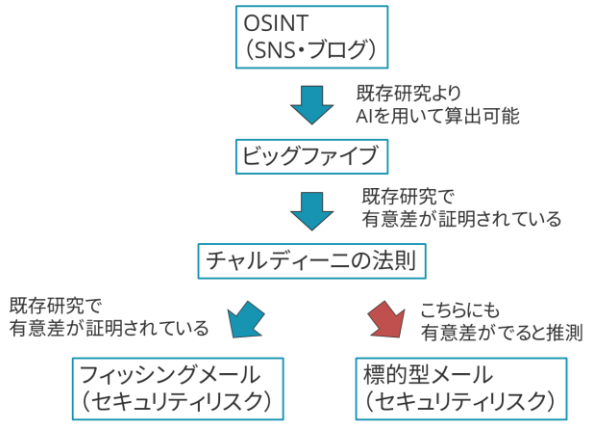

図 2　各要素との関係性

3. OSINT を利用した標的型メールの作成手順

OSINT，ビッグファイブ，チャルディーニの法則，フィッシングメール，標的型メールの関係性を図 2 に示す．

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．パーソナリティを理解する上での包括的かつ明瞭なモデルであると言われ，医療や消費嗜好調査等多くの領域で用いられている．ビッグファイブは以下の 5 つの要素に基づき，性格を数値化する．

- 知的好奇心（Openness）
- 誠実性（Conscientiousness）
- 外向性（Extroversion）
- 協調性（Agreeableness）
- 感情起伏（Neuroticism）

3.2 チャルディーニの法則

チャルディーニの法則は，チャルディーニによって提唱された，相手を自分の思い通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には，好意，返報性，社会的証明，一貫性，権威，希少性の 6 つが存在する．以下に，それぞれの概要を示す．

- 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」という心理法則である．必ずしも相手のことを知っている必要はなく，「好ましい雰囲気」や「丁寧な口調」等も好意の法則に含まれる．

- 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）」という心理法則である．一方的に押し付けられた恩であっても返報性が現れる．すなわち，恩を受けた本人が嬉しいか，嬉しくないかに関わらず，何か相手にお返しをしなくてはいけないという心理が働く．

- 社会的証明（Social Proof）

「周囲の動きに同調したくなる」という心理法則である．

「皆がやっているから自分もやる」という気持ちから生じる心理であり，人間は自分以外の誰か（第三者）の行動を物事の判断基準にしてしまう．

- 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）」という心理法則である．人間は自ら決めたことに対して，それを正当化する傾向にある．すなわち，過去に経験したような事態に出くわすと，その時と同じ行動を取ろうとする．「表明した約束を守ろうとする」気持ちも，一貫性の法則に含まれる．

- 権威（Authority）

「肩書きや経験などの“権威”を持つ者に対して，信頼を置いてしまう」という心理法則である．自分より立場が上の人物や，目上と感じる人物，特定の分野の専門家には自然と従う心理が生じる．

- 希少性（Scarcity）

「限られたものほど，価値があると感じてしまう」という心理法則である．差し迫った時間的制約があるものや，数が少ないものに対して，それがなくなってしまう前に早く取得しなければならないと思う心理が働く．

3.3 OSINT によるビッグファイブの取得

分析者が，ある人物のブログ記事から，その人物のビッグファイブの正しいスコアを推定することは，一般的に難しいことが知られている[10]．これは，分析者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，分析者によって分析対象者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 情報を取得できたとしても，これを標的型攻撃に利用することは困難であったと言える．

しかし，近年の AI の発展により，SNS 情報（ツイートやブログ記事）からその人物のビッグファイブを推定できるようになってきた．機械であれば，人間のようなバイアスに捕らわれずに，客観的に（まさに機械的に）ビッグファイブを推定することができる．

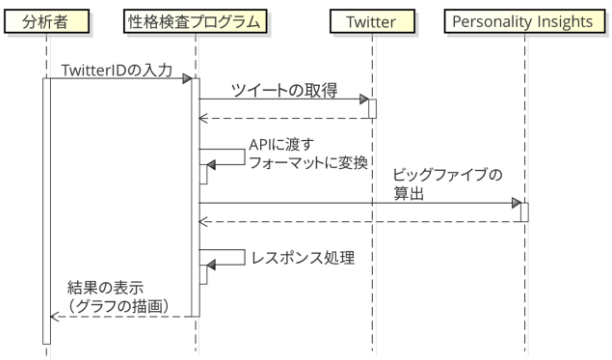

図 3　性格検査プログラムのシーケンス図

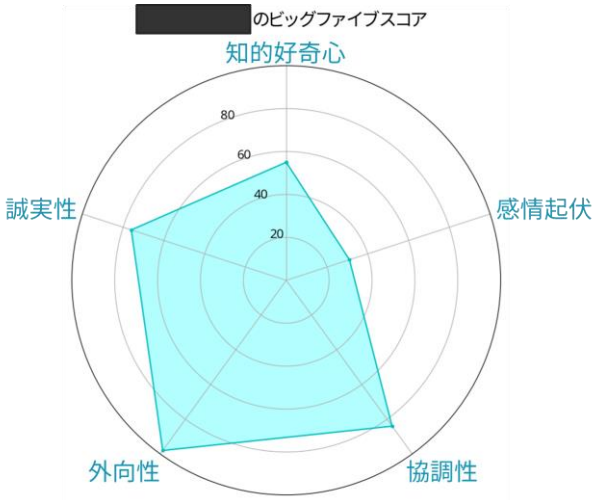

図 4　性格検査プログラムの実行結果

代表的なものに，IBM Watson の Personality Insights が挙げられる[5]．Personality Insights は IBM Watson Developer Cloud に API が公開されており，誰でも利用可能な状況にある[11]．著者らは，当該 API を用いてユーザのツイートからそのユーザのビッグファイブを推定する性格検査プログラムを作成した．プログラムのシーケンスを図 3 に示す．このプログラムにあるユーザの Twitter ID を入力することによって，得られた結果が図 4 である．なお，プライバシ保護の観点から図 4 における Twitter ID の表示は黒塗りにしている．

3.4 ビッグファイブとチャルディーニの法則の関係

チャルディーニの法則は万人に通用する法則であるが，その影響の受けやすさの度合いは性格に左右される．既存研究によって示されたビッグファイブの各スコアとチャルディーニの各法則の相関を表 1 に示す．例えば，ビッグファイブの外向性のスコアが高い人は，返報性，希少性，好意に関するチャルディーニの法則の影響を受けやすく，誠実性のスコアが高い人は，返報性，好意，一貫性の影響は受けやすいが好意の影響は受けにくい．したがって，攻撃者は，前節に説明した方法を用いて標的者のビッグファイブを取得し，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，次節に説明する方法を用いて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標的者ごとに作成することが可能である．

3.5 標的型メールへのチャルディーニの法則の利用

既存研究より，チャルディーニの法則を用いたフィッシングメールは，用いていないフィッシングメールに比べ攻撃成功率が高いことが知られている[7]．チャルディーニの法則の利用はソーシャルエンジニアリング全般に効果が見込まれるため，標的型メール攻撃を成功させたい攻撃者は，同様のアプローチで標的型メールを作成すると考えられる．

著者らの先行研究[3]に述べたように，攻撃者は OSINT

ツールを利用して標的型メールの擬態精度を高めることができる．図 5 は，攻撃者が標的とする人物の名前から OSINT によってメールアドレスと所属組織を取得した上で作成した標的型メールの例である．このメールは標的者の所属組織の技術部になりすましたものであり，アンチウイルスソフトの導入と称し，悪性 URL のクリックを促している．このメールに対し，攻撃者が更に（標的者の性格因子に合わせた）チャルディーニの社会的証明の法則と権威の法則を適用したメールをそれぞれ図 6 と図 7 に示す．

3.2 節で説明した通り，社会的証明の法則は「周囲の動きに同調したくなる」という人間の心理である．これを悪用して，攻撃者は「他の社員は既にアンチウイルスソフトの導入を完了している」とほのめかしている（図 6）．権威の法則は「肩書きや経験などの“権威”を持つ者に対して，人は信頼を置いてしまう」という人間の心理である．これを悪用して，攻撃者は所属組織の役職者の名前を使っている（図 7）．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 で役員の名前を取得するのは比較的容易であると言える．これらの文面があることで，メールを受け取ったユーザはメールの指示に従ってしまう可能性が高まると考えられる．

チャルディーニのその他の法則についても，その法則に関する人間の心理を活かした文面を追記することによって，標的型メールの心理操作効力を高めることが可能である．わずかな文面を追加するだけである程度の心理操作効力が期待されることから，攻撃者のコスト対効果は非常に高いと考えられる．

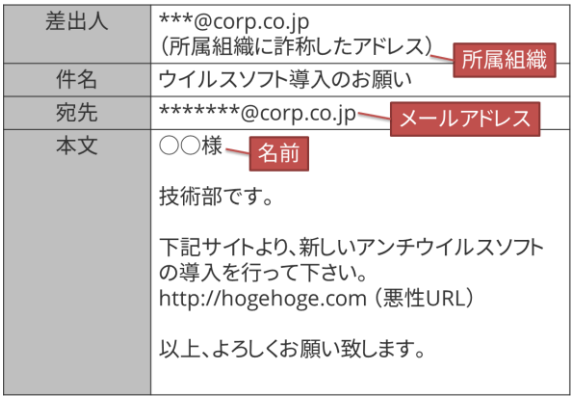

図 5　標的型メール例

図 6　社会的証明の法則を利用した標的型メール例

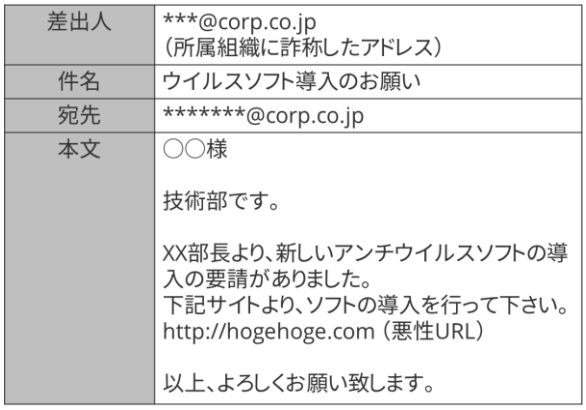

図 7　権威の法則を利用した標的型メール例

4. ASINT を利用した標的型メールの防御手順

4.1 防御側の ASINT の利用

3 章で説明した攻撃手順を用い，攻撃者は OSINT を利用して心理操作効力の高い標的型メールを作成することが可能である．しかし，本来，情報から得られるインテリジェンスは攻撃者だけでなく，正規ユーザにも恩恵を与えるものである．攻撃者が攻撃力を高めるために標的者の情報（OSINT）を利用するように，防御側も自分自身の情報（ASINT）を積極的に利用して防御力を高める方法を確立することによって，OSINT を悪用する攻撃者に対抗する手段と成す．

攻撃者が OSINT によって得られる情報と，防御側が

表 1　ビッグファイブとチャルディーニの各法則の相関[8]

	Reciprocation	Scarcity	Authority	Consensus	Liking	Commitment
Extraversion	0.139	0.255			0.121	
Agreeableness	0.197		0.252	0.101	0.289	0.147
Conscientiousness	0.139		0.234		−0.121	0.310
Neuroticism	0.112	0.166				
Openness			−0.196	−0.222	−0.122	0.124

ASINT によって得られる情報には差が存在する．通常，企業では厳格な情報管理を行っており，組織外に公開される情報は IR 情報等のみである．また，SNS では，実名をユーザ名として用いていないユーザや，情報の公開範囲を知人や友人のみに限定しているユーザが少なくない．すなわち，OSINT で得られる個人情報というのは組織や個人に関する全情報の一部であり，防御側が利用できる ASINT 情報は，攻撃者が OSINT で得られる情報より多い（図 8）．したがって，インテリジェンス（OSINT）を悪用する攻撃者に対して，防御側もインテリジェンス（ASINT）を持って対抗するアプローチは，必ず防御側が有利な結果となることが期待される（図 9）．

本稿では，3 章で述べた攻撃手順をそのまま防御法として逆用する方法を 2 例説明する．攻撃者と防御側は同一の手順を用いるが，上述の通り防御側（ASINT）の情報量が攻撃者（OSINT）の情報量を凌駕するため，防御精度が攻撃精度を上回ると期待できる．

4.2 心理操作の可能性を有するメールの検知

メールの本文中にチャルディーニの法則が利用されている文面が含まれていた際にはアラートを発するシステムである．本システムは，3 章で説明した攻撃手順を構成する自然言語処理や機械学習を組み合わせることによって作成できる．例えば，以下が考えられる．

共起検査型：

チャルディーニの各法則が利用されたメール文に現れるキーワードのリストを保持しておき，受信メールの本文中におけるキーワードの出現によってアラートを発行する．

機械学習型：

チャルディーニの各法則が利用されたメールを大量に用意し，教師ラベルとともに機械学習を実施することによってアラート推定器を構築する．

ただし，チャルディーニの法則が用いられているメールが必ずしも悪意のあるメールとは限らないため，本システムの運用においては誤アラートが増える可能性がある．

4.3 個人に応じたアラートの峻別

チャルディーニの各法則に対する反応率が個人の性格に応じて異なることを考慮すると，その対策もまた個々人に合わせてチューニングすることが有効であると考えられる．ユーザにとって，自分が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に直面した時にこそアラートが発せられるシステムであることが望ましい．また，ユーザの慢心やシステムの形骸化を防ぐためには，深刻度に応じてアラートを峻別するような対処を講じる必要もある．以下に，アラートの発生頻度が高い場合と低い場合のアラート峻別システムについて述べる[b]．

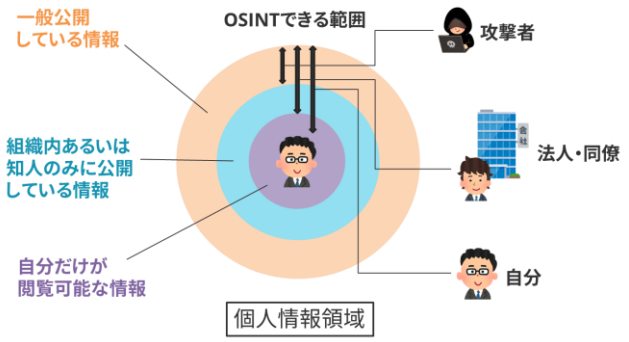

図 8　得られるインテリジェンスの差

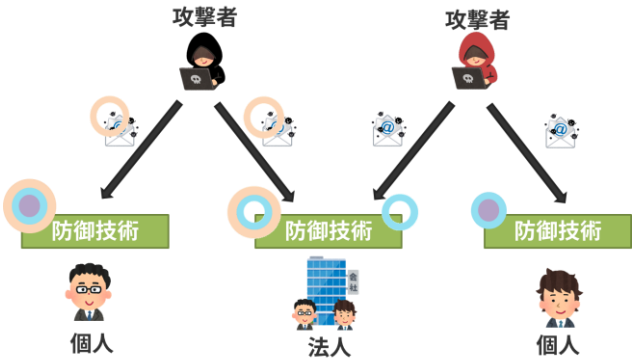

図 9　攻撃者と防御側のインテリジェンスの利用

アラートの発生頻度が高い場合：

アラートの発生頻度が高い状況においては，ユーザはたくさんのアラートに埋もれてしまうことになるため，ユーザに提示するアラートを絞ることが重要である．例えば，チャルディーニの希少性の法則に強い性格因子を有する人物と弱い人物に対し，希少性の法則が使われた同じ文面のメールが送られてきた際に，強い人物は冷静に対処できることを期待しアラートをあげないが，弱い人物は盲信して対処してしまう可能性があるのでアラートをあげる，という形でアラートを峻別することが可能である（図 10）．

アラートの発生頻度が低い場合：

アラートの発生頻度が少ない状況においては，チャルディーニの法則の利用が疑われるメールに対してはすべてアラートをあげるが，ユーザの性格因子に応じた「騙されてしまい易さ（深刻度）」に合わせて青色・黄色・赤色のラベルをメールに付与する，といった運用が効果的であると考えられる（図 11）．

b ここではチャルディーニの法則のみに焦点を当てているが，アラートの深刻度を測る尺度は他にもあり，実際にはそれら複数の尺度を用いてアラートの深刻度が決められることになる．

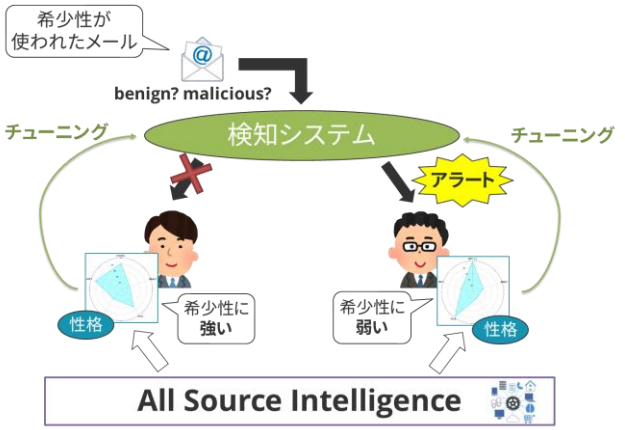

図 10　アラート発生頻度が高い状況における検知システムの運用

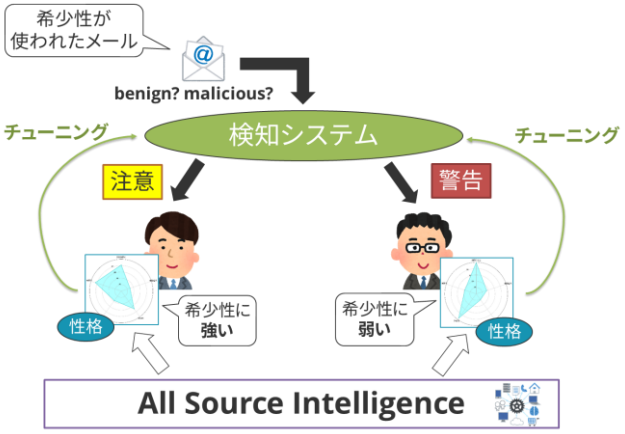

図 11　アラート発生頻度が低い状況における検知システムの運用

5. まとめと今後の課題

本稿では，攻撃者が OSINT を利用することによって，標的型メールの擬態精度だけでなく，心理操作効力を高めることができることを明らかにした. 攻撃者は OSINT ツール，性格検査 AI，チャルディーニの法則を用いて，容易に標的者ごとに標的型メールの文面を調整することができる．具体的には，ユーザが騙されてしまうチャルディーニの法則を推定し利用することによって，標的型メールの擬態精度だけでなく，心理操作効力を高めることができることを述べた．そして，この問題に対し，防御側もインテリジェンス（ASINT）を持って対抗するアプローチを提案した．本稿ではこの内，攻撃者が用いる攻撃手順を，そのまま防御に利用して標的型メールを検知・峻別する方法を示した．防御側（ASINT）の情報量が攻撃者（OSINT）の情報量を凌駕するため，防御精度が攻撃精度を上回ると期待される．

今後は，チャルディーニの法則の検知システムや個人に合わせたアラート峻別システムの具体設計を進めていく．また，本稿で述べた以外の形での ASINT を活用した OSINT 攻撃対策について検討していく．更に，提案方式を拡張することによって，ASINT 活用の目的をソーシャルエンジニアリング攻撃全体に対する対策の強化へとつなげていきたい．

研究倫理の記載

本研究の実施にあたって行った OSINT は，インターネット上のサーバに負荷をかけないよう十分に配慮して実行した．著者らが作成した性格検査プログラム（図 3）における Twitter ID に対するツイートを取得するモジュールも同様の配慮の下に実装している．

参考文献

[1] Acquisti, A., Gross, R. and Stutzman, F.: Face recognition and privacy in the age of augmented reality, Journal of Privacy and Confidentiality, Vol.6, No.2, pp.1–20 (2014).
[2] Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., Dabbish, L.: Anonymity, privacy, and security online, Pew Research Center (2013).
[3] 上原航汰，向山浩平，藤田真浩，西川弘毅，山本匠，河内清人，西垣正勝 : OSINT を利用した標的型メール攻撃手法に関する基礎検討，コンピュータセキュリティシンポジウム 2017（CSS2017）論文集，pp.222-229（2017）.
[4] Ball, L.D., Ewan G. and Coull, N.J.: Undermining-social engineering using open source intelligence gathering, Proc. of 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012), SciTePress-Science and Technology Publications, pp.275-280 (2012).
[5] IBM : Watson Personality Insights, IBM-Japan（オンライン），入手先<https://www.ibm.com/watson/services/personality-insights/>（参照 2018-07-09）.
[6] Cialdini, R.B.: Influence (Vol. 3), HarperCollins (1987).
[7] Wright, R.T., Jensen, M.L., Thatcher, J.B., Dinger, M., 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).
[8] Alkış, N. and Temizel, T.T.: The impact of individual differences on influence strategies, Personality and Individual Differences, Vol.87, pp.147-152 (2015).
[9] Akbar, N.: Analysing persuasion principles in phishing emails, Master's thesis, University of Twente (2014).
[10] 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，Vol.29，pp.1-4（2015）.
[11] IBM: IBM Watson Developer Cloud, IBM Cloud (online), available from <https://www.ibm.com/watson/developercloud/> (accessed 2018-07-10).