Computer Security Symposium 2018
22 - 25 October 2018

# 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

西川弘毅 [1,2]　上原 航汰 [2]　山本匠 [1]　河内清人 [1]　西垣正勝 [2]

**概要**：標的型攻撃メールや，BEC(ビジネスメール詐欺)といった，メールによるサイバー犯罪の被害は，依然として脅威となっている．攻撃者は，OSINT(Open Source INTelligence)や HUMINT (HUMan INTelligence)といった，様々な手法を用いて標的の情報を収集し，標的に対して効果的な文章のメールを送りつけてくる．一方，防御側が得られるインテリジェンスは，攻撃者よりも多く，その差を用いて攻撃を防御することが可能になると考えられる．本稿では，攻撃者が様々なインテリジェンスを駆使することにより可能となる攻撃と，それに対抗するために防御側でも活用できるインテリジェンスとして，OSINT，PRINT(Proprietary Inteligence)，EXINT(EXperimental INTelligence)を活用する方策を示す．

**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，Intelligence

## 1. はじめに

標的型メール攻撃は依然として社会の脅威である[1]．標的型メール攻撃に対する様々な対策が考案されてはいる一方，攻撃者は，システムの堅牢化に伴って，最も脆弱である”人”に対する攻撃に力を入れており，標的型メール攻撃の脅威が下がることはない．標的型メール攻撃はソーシャルエンジニアリングの一つであり，メールによって攻撃対象者を騙すことで，金銭や情報の窃取といった攻撃者の目的を達成する．標的型メール攻撃を成功させるためには，攻撃対象者に，標的型メールを正規のメールと信じ込ませることが必要である．そのため攻撃者は，攻撃の対象者に関する情報を収集し，さらに自然なメールのやり取りを装うことで，メールの信憑性を高めようと試みる．

攻撃者は，標的組織に対して詐欺等の攻撃を成功させるために，公開情報からの情報収集である OSINT(Open Source INTelligence)や，人間に対して接触し，情報を得る HUMINT (HUMan INTelligence)といった，様々な手法を用いて標的に関する情報を収集し，攻撃に利用可能な形式であるインテリジェンスを収集する．攻撃者は，収集したインテリジェンスを利用し説得することで，標的に情報や金銭の窃取やマルウェア感染を引き起こす．本稿では，特にインテリジェンスを活用した標的型攻撃メールに着目する．攻撃者は，インテリジェンスを活用することで，標的が信じてしまうような内容のメールを作成し，標的に対して送信することができる．このように，攻撃者にとって，確度の高い攻撃を実施するためにはインテリジェンスは不可欠であり，防御側は，攻撃者がインテリジェンスを活用してくることを前提として防御手段を講じる必要がある．

しかし，防御側においても，このようなインテリジェンスは，自身の身を守るために活用可能である．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，後者の，ASINT を利用した防御手法の全体像を示す．防御側で利用可能なインテリジェンスとしては多種多様なものがあると思われるが，本稿では，攻撃者も利用する OSINT，組織が保有するスケジュール等の情報を利用する PRINT(PRoprietary INTeligence)，これまでに行われた攻撃事例等の情報を利用する EXINT(EXperimental INTelligence)に着目して論ずる．

本稿の構成を次に示す．まず，2 章では，関連する研究として，既存の不審メール技術を紹介する．3 章では，攻撃者と防御側とがインテリジェンスを活用した際の全体像を示す．4 章では，攻撃者がインテリジェンスを活用した際に行われる攻撃をモデル化する．5 章では，防御側で，OSINT，PRINT，EXINT を活用することで可能となる防御

1 三菱電機株式会社　情報技術総合研究所
Mitsubishi Electric Corporation Information Technology R&D Center
2 静岡大学
Shizuoka University

手段を示す.

## 2. 関連研究

本章では，既存の不審メール検知技術について説明する. CipherCraft/Mail[2]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．しかし，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[3]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[4]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 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.

## 3. 攻撃と防御の相互関係

本章では，攻撃者が対象組織に対して，インテリジェンスを活用することで，どのように攻撃を試みるか，さらに防御側は，インテリジェンスを活用することで，どのように攻撃を防ぐのか，その概要を示す.

インテリジェンスを活用することで，対象組織への物理的侵入や，電話による詐欺等も考えられるが，本稿では特に，標的型メール攻撃に焦点を当てて議論する.

攻撃者は，インターネットや対象組織等の人物等に対し，OSINT や HUMINT 等の手法を実施することで，攻撃に活用な情報であるインテリジェンスを入手する．攻撃者は，入手したインテリジェンスに加え，知識や技術として保持している詐欺の技術を駆使し，標的型メールを，攻撃対象者に送信する.

ここで防御側が，送信された標的型メールをどのように検知し，防ぐのかを考える．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，攻撃者が OSINT によって入手できる標的の情報は, 標的が 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.

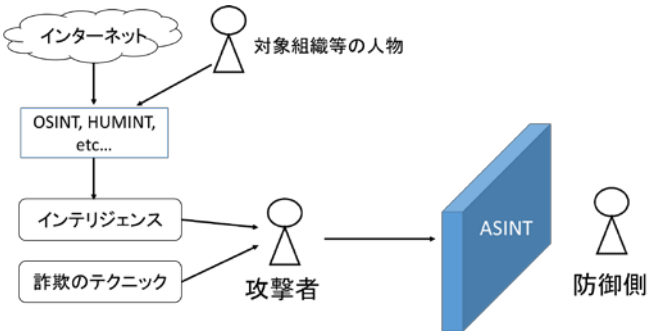


図 1　攻撃者のインテリジェンスや詐欺のテクニックを利用した巧妙な攻撃を ASINT により防ぐイメージ図

攻撃者が，インテリジェンスや詐欺のテクニックを駆使して攻撃を試みることに対し，ASINT により攻撃を防ぐ防御側の図を図 1 に示す.

## 4. 攻撃者によるインテリジェンス活用モデル

本章では，攻撃者が，インテリジェンスを活用することで可能となる攻撃モデルを示す．本章では，攻撃者が OSINT を活用することで，標的の選定から，信じ込ませるための文章作成が可能となることを示す．また，心理学的特性や，やり取りを通して，より標的が騙されやすくなることを利用した攻撃についても示す.

### 4.1 OSINT による攻撃モデル

攻撃者は，標的がつい開いてしまうような巧妙な標的型メールを作成するために，OSINT を活用する．OSINT によって，標的がどのような人間であるかを特定し，どのようなメールを送ることが有効であるかを知ることができるためである.

著者らは，OSINT が標的型メールを正規のメールに模倣するための情報収集に活用される可能性について言及した[5]．具体的には，OSINT を半自動的に実行する OSINT ツールにより，攻撃対象者の所属組織やメールアドレス，所属組織の上司の名前，友人関係などが取得することで，攻撃者がこれらの情報をメールに組み入れ盛り込み，標的型メールを正規のメールだと攻撃対象者に信じ込ませることが可能であることを明らかにした.

### 4.2 チャルディーニの法則による攻撃

チャルディーニの法則とは，社会心理学者ロバート・B・チャルディーニが提唱した，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」という人間の心理である.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”“丁寧な口調”も好意の法則に含まれる.

- 返報性（Reciprocation）

「人から受けた恩は,返さなければならない（返したくなる）」という人間の心理である.その恩というのは,一方的に押し付けられたものであっても有効である.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」という人間の心理である.つまり,“皆がやっているから自分もやる”といったような心理である.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のである.

- 一貫性（Commitment and Consistency）

「自分の行動に一貫性を持たせたい」という人間の心理である.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る傾向を示す.「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,人は信頼を置く」という人間の心理である.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」という人間の心理である.差し迫った時間的な制限や,数が少ないものに対して,早く行動しなければならないと思う心理が働く.

既存研究により,これらの心理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おり,より標的型メール攻撃を成功させたい攻撃者は,同じように心理的要因にアプローチする標的型メールを作成すると考えられる.

### 4.3 やり取りを伴う攻撃モデル

やり取りを伴う攻撃のモデルとしては,ビジネスメール詐欺や,やり取り型攻撃がある.本節は,それぞれの攻撃の概要を示す.

#### 4.3.1.1. ビジネスメール詐欺

ビジネスメール詐欺(BEC)は,巧妙な騙しの手口を利用して,メールによって金銭をだまし取る攻撃である.具体的には,攻撃者が,攻撃対象とする企業の重役や取引関係のある企業の担当者になりすまし,攻撃対象企業の決済担当者にメールを送り,攻撃者の口座へ送金するように誘導する.

巧妙なビジネスメール詐欺においては,技術的な対策が難しく,一人ひとりが手口を理解し,“怪しさ”を見抜くことが重要である,と言われている[6].

#### 4.3.1.2. やり取り型攻撃

本節では,IPA の資料[7]を参考にし,攻撃者が標的とやり取りを行うことで信頼を得た後,感染行動に移るやり取り型攻撃の説明を行う.

資料[7]中に,「「やり取り型攻撃」とは,一般の問い合わせ等を装った無害な「偵察」メールの後,ウイルス付きのメールが送られてくるという,標的型サイバー攻撃の手口の一つです.」と,記載されている.攻撃者は,対象とする組織の外部向け窓口等に対して,返信せざるを得ないメールを送りつける.対象から返信があると,辻褄の合う会話をしながら,マルウェアである添付ファイルを開かせ,組織へのマルウェア感染を試みる.やり取り型攻撃のイメージを図 2 に示す.

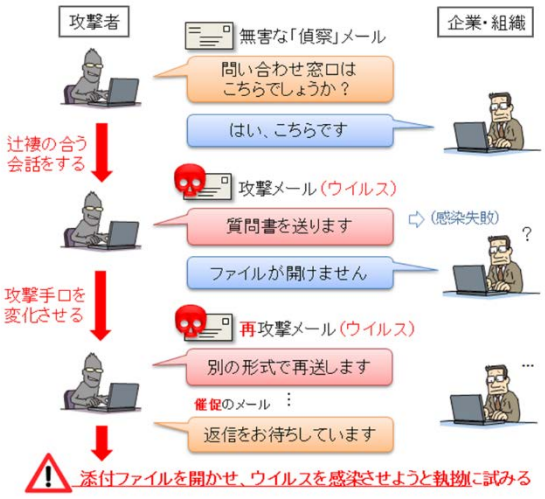


図 2 やり取り型攻撃の例

## 5. 防御側が活用するインテリジェンス

攻撃者が OSINT によって得られる情報と,防御側が ASINT によって得られる情報には差が存在する.例えば,SNS を実名ではなく,ニックネーム等の匿名でやっている者や,公開範囲を適切に設定して,知人や友人のみにしかアクセスを許可していない者もいるだろう.すなわち,OSINT で得られる個人情報というのは個人情報領域の一部であり,ASINT がそれらを包含していることから,防御側が得られる個人の情報というのは,攻撃者が OSINT で得られる情報より多い.図 3 に,得られる情報量の差を表現した図を示す.

防御側が,企業やユーザが有していうる全情報の活用で

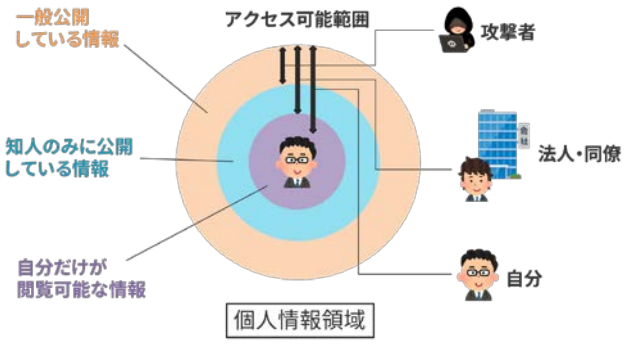


図 3　得られる情報量の差

ある ASINT の具体例として，OSINT，PRINT，EXINT の三つがある．

本章では，各インテリジェンスを利用することで，どのような防御が可能となるかを説明する．

**5.1.1 OSINT を利用した防御方法**

本節では，防御側が OSINT によって公開情報を収取することにより取ることができる防御方法について示す．

「攻撃者が攻撃に資する情報を抽出する手法」を防御側もそっくりそのまま利用することによって，攻撃者が自分にどういう攻撃メールが送信してくる可能性があるのか，自分はどういう攻撃メールに引っかかりやすいのか，等の「自衛に資する情報」を知ることが可能となる．

OSINT だけではなく，後述する PRINT や EXINT と組み合わせることで，防御側は，攻撃者が得ることのできる「攻撃に資する情報」よりも多くの情報を，精度よく収集することが可能となる．例えば，SNS を実名ではなく，ニックネーム等の匿名でやっている者や，公開範囲を適切に設定して，知人や友人のみにアクセスを許可している場合，本人であれば，それらの情報も活用することができる．

OSINT を活用した防御方法として，ある個人が騙されてしまいやすいチャルディーニの法則に応じて，メールのアラートレベルをチューニングする方法がある[a].

本例を示すために，危険であると判断されアラートが付されたメールが届いた場合を想定する．まず，メールに用いられているチャルディーニの法則を推定する．これは，文章からチャルディーニの法則を推論可能な識別器をあらかじめ用意することで対応することができる．続いて，OSINT を利用することで，個人の弱点となるチャルディーニの法則を推定する．これらの情報を利用し，個人ごとにメールのアラートを強調する等の調整を行い，多発するアラートによるシステムの形骸化を防止することができる．

**5.1.2 PRINT を利用した防御方法**

本節では，当該組織が占有している様々な情報を利用した PRINT の考え方と，それを応用した防御方法について示す．

当該組織のディレクトリ情報や社員のスケジュールといった組織内で管理されている情報を活用して，防御に役立てることができる．ここで，組織内の情報は一般的に部外秘扱いで管理されており，攻撃者にとっては入手しにくい情報であるため，これを防御に利用することによって効果的な防御が期待できる．

PRINT を用いた防御方法の具体例として，社内のある人物から，標的が関連するプロジェクトに関するメールが届いたことを想定する．普段からやり取りがある人物であれば問題はないが，例えばその人物はプロジェクトに関係がないことや，メール送信時刻に，勤怠システムを見るに社内にいないことが判明した場合，不審と判断することができる．

また，それ以外にも，当該人物から，メールが送信された際に，これまで受信したメールの文脈と異なるメールの場合も不審であると判断できる．例えば，それ以前に添付ファイル付のメールを受信していないにも関わらず,「先ほどの添付ファイルは誤りでした.」と，本文に記載され，添付ファイルがついたメールが届いた場合や，やり取りが無いにも関わらず,「資料を送付します.」と，突然メールが送られてきた場合を考える．これらの場合でも，組織が占有しているディレクトリ情報や社員のスケジュール，メールサーバ等の PRINT を利用することで，不審であると判断できる．

**5.1.3 EXINT を利用した防御方法**

本節では，これまでの攻撃情報に関する情報を基にする EXINT を利用した不審メールの検知について示す．

まず，これまでの攻撃メールのやり取りに関する情報をデータベース化，可能であればモデル化を実施する．その後，データベースに登録された攻撃メールのやり取りに関する情報，或いはやり取りのモデルを利用することで，攻撃を検知する手法である．

著者らは，EXINT の具体例として，メールを受信した際に，該当メールと関連するメールによってスレッドを作成し，そのスレッドが，予め登録したやり取りと同一である場合に，不審であるとアラートをあげる手法を提案した[8]. 図 4 に，本手法の模式図を示す．

a 本手法の詳細な議論は，著者らによる別の発表にて取り扱う．

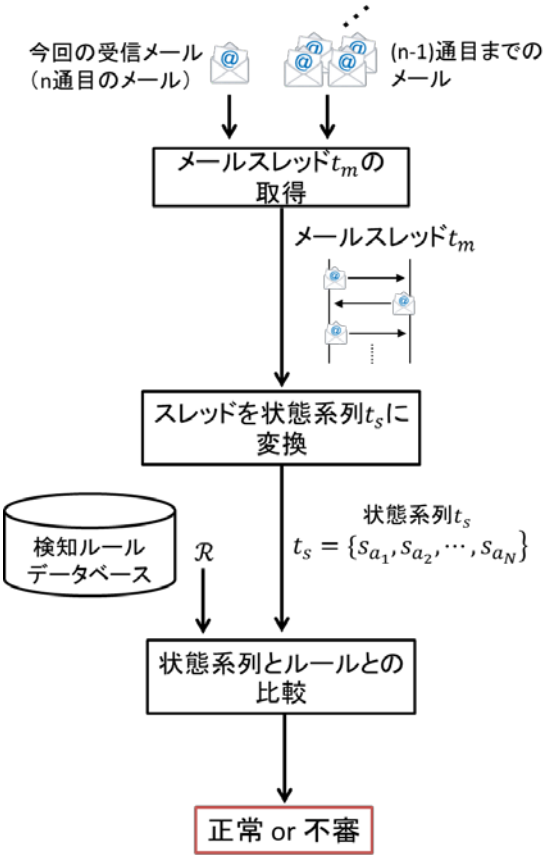


図 4　EXINT によるやり取りを伴う攻撃検知手法

## 6. おわりに

本稿では，攻撃側がインテリジェンスを利用した際に行われる攻撃に対抗するために，防御側でもインテリジェンスを活用することが有効であることを述べた．特に本稿では，防御側は，公開情報から情報を収集する OSINT，組織が保有するスケジュール等の情報を利用する PRINT，これまでに行われた攻撃事例等の情報を利用する EXINT により，攻撃を検知・緩和させる戦略を示した．

今後，各コンポーネントを実装し，評価を行う．

## 参考文献

[1]　IPA，情報セキュリティ 10 大脅威 2018, https://www.ipa.go.jp/security/vuln/10threats2018.html
[2]　CipherCraft/Mail, https://www.ntt-tx.co.jp/products/ccraftmailtypeh/
[3]　Disarm, https://support.symantec.com/en_US/article.HOWTO93096.html
[4]　Sevtap Duman, Kubra Kalkan Cakmakciy, Manuel Egelez, William Robertson and Engin Kirda, “EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails”, Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual.
[5]　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. (2017). 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論文集, 2017, 222-229.
[6]　IPA, 【注意喚起】偽口座への送金を促す“ビジネスメール詐欺”の手口, https://www.ipa.go.jp/security/announce/20170403-bec.html
[7]　IPA，組織外部向け窓口部門の方へ：「やり取り型」攻撃に対する注意喚起 ～ 国内 5 組織で再び攻撃を確認 ～, https://www.ipa.go.jp/security/topics/alert20141121.html.
[8]　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", DICMO2018, pp. 1298 - 1302