Vol.2017-DPS-170 No.33
Vol.2017-CSEC-76 No.33
2017/3/3

やり取り型の標的型メールを自動で行うシミュレータの検討

西川　弘毅[†1]　山本　匠[†1]　木藤 圭亮[†1]　河内　清人[†1]

特定の組織や人を対象とする標的型メール，特にやり取り型の攻撃が新たな脅威として想定されている．標的型メールを予防する手段として，標的型メールの訓練がある．しかし，やり取り型のメールによる訓練を実施できる高度な技術者は少ない一方で，対象者は技術者よりも多いため，訓練を十分に実施することができない問題がある．本稿では，現在送信するメールの状態が，偵察・攻撃・催促のどの状態であるか，訓練者から受信したメールにより判断することで，自動でやり取り型攻撃のシミュレーションを行う手法を提案する．

Automatic Exchange-type Attack Simulator

Hiroki Nishikawa[†1] Takumi Yamamoto[†1] Keisuke Kito[†1] Kiyoto Kawauchi[†1]

Spear phishing email, especially exchanging email with a target is supposed to be a new threat. Spear phishing email training is a typical method to prevent malware infection by opening spear phishing email. However, on the one hand there are little engineers who can conduct the training, on the other hand there are many trainees who need to receive the training. Therefore we could not serve sufficient the training. In this paper, we propose Automatic Exchange-type Attack Simulator which judges sending mail based on a state of the system what is next mail.

1. はじめに

特定の組織や人を対象に機密情報の窃取等の攻撃を行う標的型攻撃は深刻な脅威となっている．その中でも，感染経路にメールを用いる，標的型メール手法は重大な脅威の一つである．トレンドマイクロによれば，標的型攻撃において，感染経路にメールを使うものは，いまだに 76%以上存在している[1].

標的型メールによる攻撃を予防する手段の一つとして，標的型メールの訓練を行うツールやサービスがある[2-11]. これらは, 訓練者に対して, 実際の標的型メールを模した，訓練用のメールを送信することで，対象者を訓練することを想定している．訓練用のメールは，実際の標的型メールで使われるような特徴を含むように題名や本文が作られており, 訓練用の添付ファイルや URL を開くように誘導する. 訓練者が添付ファイルや URL を開くと, 訓練であることを訓練者に通知し，不審なメールを開かないような教育につなげることができる. このように, 訓練者は訓練を通じて，実際の標的型メールはどのようなものか，また標的型メールを受信した際にはどのような対応を行えばよいかを訓練することができる.

しかし，最近では，何度か標的とやり取りを行った後，マルウェアに感染させるようなメールを送信する，やり取り型の標的型メール（以下，やり取り型攻撃）が報告されている[12]．このような巧妙な攻撃は，細心の注意を払っていても感染してしまう危険性がある.

IPA の報告[12]によれば，このような攻撃は，既存の攻撃と比較して多くは報告されていない．しかし，これは攻撃が高度であるため攻撃に気づいていない可能性も高く，実際には報告よりも多く存在する可能性がある.

一方，自動で文章を生成することでスピアフィッシングを行うようなツイートを生成する技術も公開されている[13]．このように，攻撃者の能力が向上し，やり取り型攻撃のような高度な攻撃も簡単に行われる危険性が高まっている．そのため，やり取り型攻撃に対しても，メール訓練を行うことが必要であると筆者らは考えている.

しかし，既存の標的型メール訓練手法では，やり取りを行うことができないため，やり取り型攻撃の訓練を行うことができなかった．また，ダミーメールで利用する文章をひな形として予め用意する必要があったため，状況にあった文章を自動で生成することができなかった.

更に，やり取り型攻撃をシミュレートできるような高度な技術を持つ技術者は少なく，人手では多数の訓練対象者に訓練を提供できない問題があった.

そこで，訓練者にやり取り型攻撃の脅威を体験させ，教育することを目的した，自動でやり取り型攻撃のシミュレーションを行う手法を提案する．やり取り型攻撃の分析をした結果，攻撃の開始状態と終了状態，標的の出方を伺う偵察状態, 標的に対して添付ファイル付きのメールや, URL が本文に書かれたメールを送信する攻撃状態，メールへの返信が来ないことへの催促状態，の 5 つの状態を遷移することが分かった．この知見を元に，現在の状態が，受信したメールに応じてどのように遷移するかを，メールの特徴から学習することで，次に送るべきメール文章を適切に生成できる手法を提案する.

本稿の構成は，2 章で関連研究について示し，3 章では，やり取り型攻撃について説明する．4 章では，提案するやり取り型攻撃シミュレーションの動作を説明する．5 章では，提案手法を実装するための課題等について考察する.

[†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.

Vol.2017-DPS-170 No.33
Vol.2017-CSEC-76 No.33
2017/3/3

2. 関連研究

標的型メールに対する訓練として，訓練用のメールを生成し，訓練対象者に送信するようなツールやサービスがある[2-11].

これらの訓練は，訓練提供者によって，概ね次のような手順により実施される.

1． 訓練対象に対して，訓練メールを送信する．訓練メールには，添付ファイルがあるか，本文中に URL が記載されていて，これらを開くように誘導する文章が書かれている.

2． 訓練対象が添付ファイルや URL を開くと，訓練メールであり，怪しいファイル・URL を開いてしまったことを訓練対象に通知するとともに，誰が，いつ開いたかの情報を収集する.

3． 収集した情報と，必要であれば訓練対象から収集したアンケート結果を元に，訓練対象の組織を評価し，報告書を作成する.

このような標的型メール訓練では，やり取りを行うことができないため，やり取り型攻撃の訓練を行うことができなかった．また，ダミーメールで利用する文章をひな形として予め用意する必要があったため，状況にあった文章を自動で生成することができなかった.

一方，John らは，マルコフモデルや RNN によって文章を自動で生成し，Twitter 上でスピアフィッシングを行う手法を提案している[12]．文献の中で，人間が行う場合には一分間で 1.67 のツイート，クリックされた数が 49 であったが，本手法では一分間で 6.85 ツイート，クリックされた数が 275 と，人手で行うより効率的に攻撃が実行可能であることを示している.

このように，技術の発展に伴い攻撃者の能力が拡大されることを考えると，やり取り型攻撃のような高度な攻撃も簡単に行われる危険性が高まっている.

3. やり取り型攻撃とは

本章では，IPA の資料[12]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する.

「「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，ある．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 1 に示す.

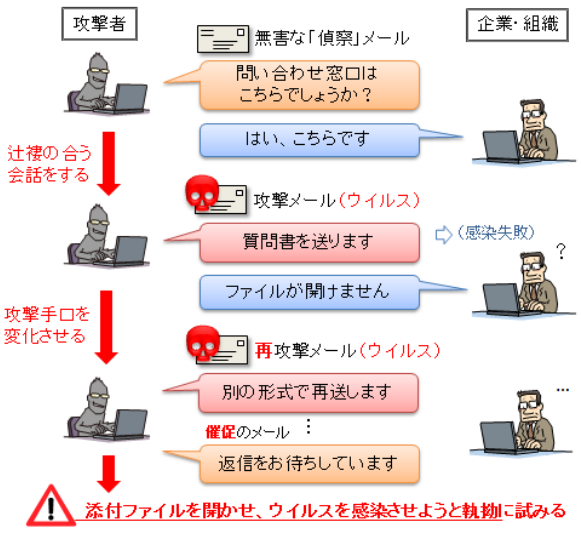

図 1 やり取り型攻撃のイメージ[12]

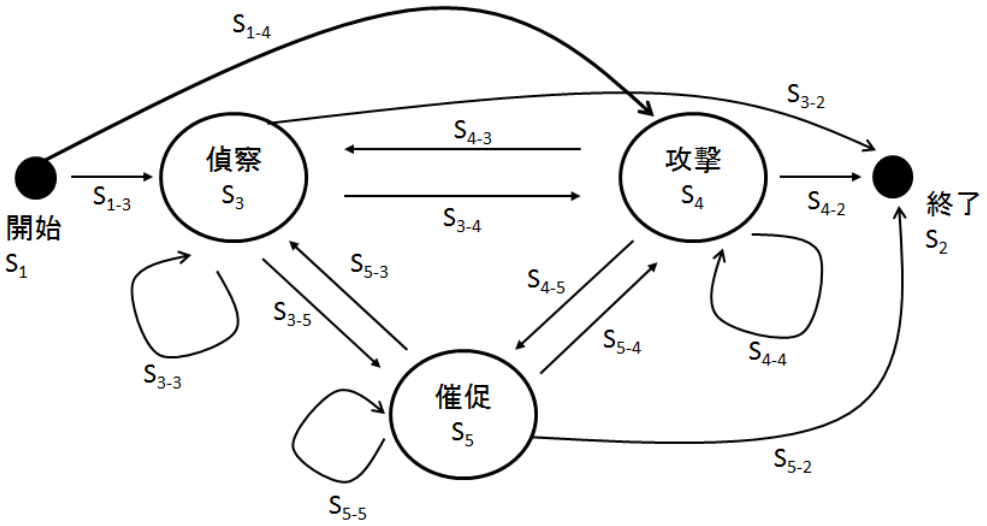

図 2 やり取り型攻撃の状態遷移

Vol.2017-DPS-170 No.33
Vol.2017-CSEC-76 No.33
2017/3/3

4. 提案手法

4.1 状態の定義

文献[12]に示される例を元に，やり取り型攻撃を分析した結果，攻撃の開始状態と終了状態，標的の出方を伺う偵察状態，標的に対して添付ファイル付きのメールや，URL が本文に書かれたメールを送信する攻撃状態，メールへの返信が来ないことへの催促状態，の 5 つの状態を遷移することが分かった．この分析に基づいた，やり取り型標的型メール攻撃の状態遷移を図 2 に示す．

ただし，S1，S2，S3，S4，S5 は，各々，開始，終了，偵察，攻撃，催促，の状態を表している．また，S1-3 や S3-3 等は，状態から状態への遷移を表しており，このことを本稿では遷移状態と呼ぶ．

4.2 訓練の流れ

訓練は大きく，次の三つのフェーズに分かれて実施される．

①登録フェーズ

訓練対象の氏名や所属，メールアドレスといった情報を入力する．更に，訓練対象に適したメールのやり取りを収集する．

②学習フェーズ

登録フェーズで収集したメールを元に，遷移状態の学習と，メールを生成するための，メール生成モデルを生成する．

③訓練フェーズ

学習フェーズで生成した，遷移状態，メール生成モデル，登録フェーズで入力した訓練対象の情報，を元に訓練対象に対して訓練を実施する．

次節より，各フェーズについて詳細に説明する．

4.3 登録フェーズ

登録フェーズでは，訓練対象の情報を入力と，学習データとなるメールの収集を行い，訓練を開始できるように準備を行う．ここで，訓練対象の情報とは，訓練者の氏名や，所属する組織名，メールアドレスや，訓練時にメールを送る元として詐称する氏名や組織名，メールアドレス，更に訓練のシナリオ等を指定する．例えば訓練のシナリオとして，製品の質問受付窓口に対する訓練や，リクルート業務に携わる者に対する訓練が考えられる．登録フェーズで登録する情報の例を表 1 に示す．

続いて，設定した訓練対象に適切なメールを用意し，学習データとする．この学習データは，次のフェーズである学習フェーズで利用される．

ここで，訓練の元となるメールの収集については，考察で述べる．

表 1　登録する属性情報名の例

属性情報名	登録内容
訓練者名	山田花子
訓練者名（読み）	やまだ　はなこ
訓練者組織名	XY 商事
訓練者組織名（読み）	えっくすわいしょうじ
訓練者メールアドレス	yamada@xyShoji.co.jp
攻撃元名	佐藤太郎
攻撃元名（読み）	さとう　たろう
攻撃元組織名	ab 運送
攻撃元組織名（読み）	えーびーうんそう
攻撃元メールアドレス	sato@abunso.co.jp

4.4 学習フェーズ

学習フェーズでは，登録フェーズで入力した情報を元に，状態から状態への遷移を示す遷移状態を学習し，更にメール生成モデルを生成する．

学習フェーズでは，次の 4 つの処理が順に実施される．

① メールの振り分け

登録フェーズで用意した学習用のメールを，遷移状態ごとに振り分ける．

② 特徴ベクトル算出

各メールの特徴から，それぞれ特徴ベクトルを算出する．

③ 遷移状態の特徴算出

遷移状態ごとに振り分けられたメールの集合から，遷移状態の特徴ベクトルを算出する．

④ メール生成モデル生成

振り分けた状態ごとに，メールの内容から，メール生成モデルを生成する．

以下，各処理について，詳細を説明する．

4.4.1 メールの振り分け

本処理では，学習データとして収集したメールを各遷移状態に振り分ける．本処理により，この後の処理である遷移状態の学習が行えるようになる．

まず，収集したメールを，やり取り毎に分けていく．これは，例えばあるメールを始点として，何度かのやり取りを行った後に終点となるような一連のやり取り毎に分けていく．

図 3 に，一連のやり取り毎に分けられた後の，ある一つのやり取りを振り分ける様子を示す．振り分けのルールは，訓練対象の組織側ではなく，外部から送信されているメールを始点とする．この理由は，問合せは外部が起点となるためである．一連のやり取りの起点となる送信者を外部側，問合せに対応する方を組織側として振り分ける．続いて，外部側に対して状態を与える．ここで状態を振り分けるル

ールは，メール開始前の状態を開始状態，やり取りが終了した状態を終了状態，添付ファイル，或いは本文中に URL がないメールを問合せ状態，添付ファイル，或いは本文中に URL があるメールを添付状態，自分から連続して送信しているメールを催促状態として，外部側の全てのメールに対して状態を与える．続いて，外部側のメールがどのように遷移しているかで，外部側から送信されるメールと，組織側から送信されるメールの両方に，遷移状態を与える．

4.4.2 特徴ベクトル算出

本処理では，メールから特徴ベクトルを抽出する．手法としては，既存手法に mail2vec という手法がある[14]．これは，単語を特徴ベクトルに変換する技術である word2vec と，予め学習したデータセットの情報を元に，メールを特徴ベクトルへ変換する手法である．

特徴量への変換手法により，メールは T 次元のベクトルへと変換される．メールから特徴ベクトルを抽出するイメージを図 4 に示す．

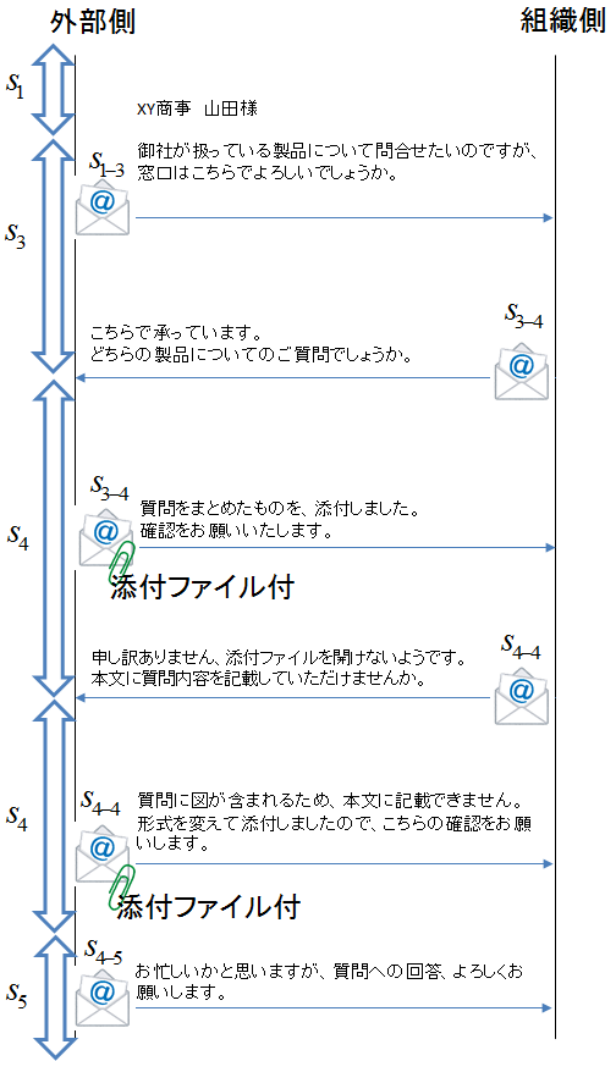

図 3　やり取りされるメールの例

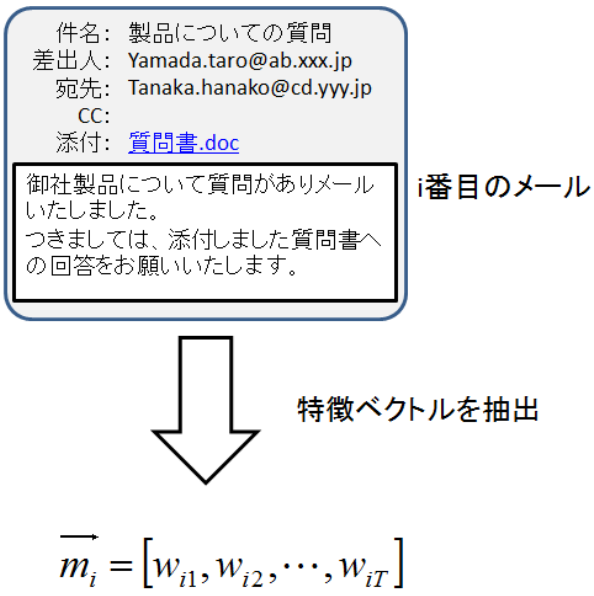

図 4　メールからの特徴ベクトル抽出

4.4.3 遷移状態の特徴算出

本処理では，4.4.1 で振り分けたメールを利用して，各遷移状態の特徴ベクトルを算出する．ここで，遷移状態は，メールの振り分けによって，メール集合が形成されている．そのメール集合中の各メールに対して，特徴ベクトルを算出することができるため，遷移状態は複数の特徴ベクトルの集合で表現されている．

ここで，遷移状態の特徴ベクトルは，遷移状態が有する特徴ベクトル集合の平均とすることができる．

以下に，遷移状態の特徴ベクトルの算出式を示す．

$$\vec{l} = \frac{1}{\left|\sum \vec{v}_i\right|} \sum \vec{v}_i \qquad (1)$$

ただし，$\vec{l}$ は遷移状態の特徴ベクトルで，$\vec{v}_i$ は遷移状態に含まれる i 番目の特徴ベクトルで，変域は $(0 \le i \le L)$ で，i は整数で，L は集合の要素数である．

4.4.4 メール生成モデル生成

本処理では，メールを生成するためのメール生成モデルを生成する．本手法では，マルコフモデルによって生成モデルを表現する．

まず，メール生成モデル生成部は，前処理を行うことで，学習データの抽象度を上げる．図 5 は前処理の様子を示している．このように，メールの送信相手企業名や苗字を，[訓練者企業名]や，[訓練者苗字]といった属性情報のタグと同一名称の記号で置き換える．具体的には，MeCab[15]のような既存の技術を用いて，形態素解析を行い，名詞の組織名や人名を参照することで，どの記号に置き換えるかを特定する．この時，置き換える記号の主体が，訓練者か，攻

撃元かの判断は，送信元アドレスを用いたり，敬称の有無で判断したりすることで行う．

続いて，前処理済みの文章を入力として，この文章を形態素解析し，形態素解析済みのデータを入力としてマルコフモデルを生成する．マルコフモデルを生成する様子を図 6 に示す．

4.5 訓練フェーズ

訓練フェーズでは，登録フェーズで入力した訓練対象の情報を元に，訓練対象に対してメールのやり取りを行う．図 7 に，訓練対象者とのメールやり取りの流れを示す．

まず，本手法における訓練システムは，一通目のメールを生成し，対象に送信し，対象からのメール返信を待ち受ける．

メールの返信があるか，一定時間が経過後，状態遷移処理を行い，現在の状態から次の状態へと遷移する．

状態遷移処理について，詳細に説明する．図 8 に，状態遷移処理の流れを示す．

図 9 に，メールから算出した特徴ベクトルと，遷移状態の特徴ベクトルの模式図を示す．図 9 は，現状態が偵察状態 S3 である場合に，メールから算出した特徴ベクトル $\vec{m}_i$ と，偵察 S3 から偵察 S3 に遷移する遷移状態 S3-3 の特徴ベクトルと，偵察 S3 から攻撃 S4 に遷移する遷移状態 S3-3 の特徴ベクトルを T 次元空間上で示したものである． ここで，各特徴ベクトルは T 次元ベクトルである．

この時，S3 から S4 に遷移するような S3-4 の遷移状態が選択される場合は，次の二式を同時に満たす場合である．

$$\left|\vec{m}_i - \vec{s}_{3-4}\right| \le \delta \quad (2)$$

$$\left|\vec{m}_i - \vec{s}_{3-4}\right| \le \left|\vec{m}_i - \vec{s}_{3-3}\right| \quad (3)$$

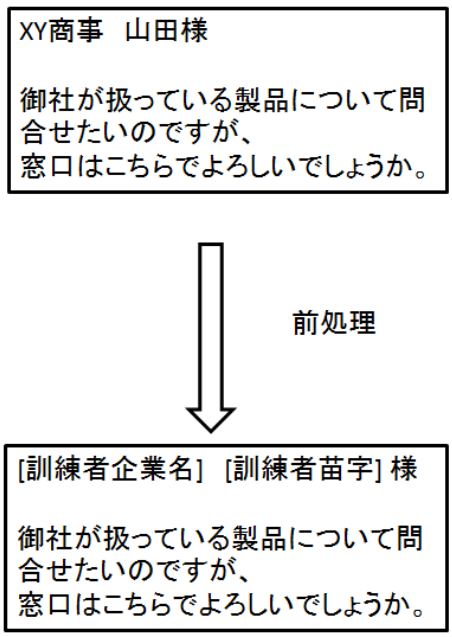

図 5　前処理が行われた例

状態遷移先が終了の場合には処理を終了する．それ以外の場合には，前状態と現状態から分かる遷移状態を元に，利用するメール生成モデルを決定する．メール生成モデルに従って，メールの文章を生成し，対象にメールを返信する．

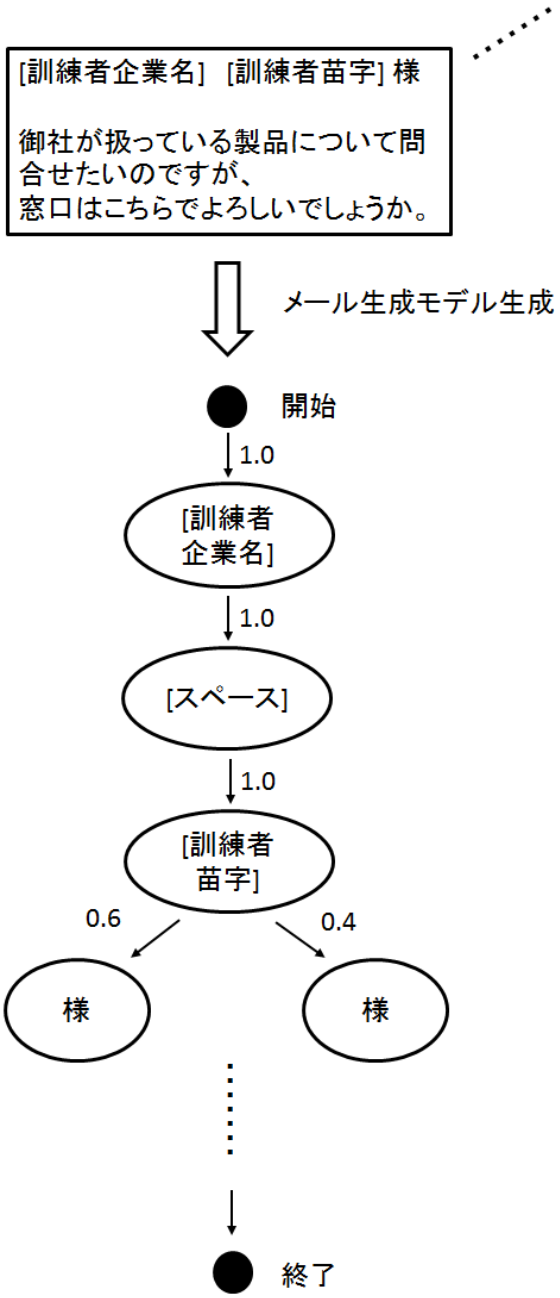

図 6　マルコフモデルによるメール生成モデル

Vol.2017-DPS-170 No.33
Vol.2017-CSEC-76 No.33
2017/3/3

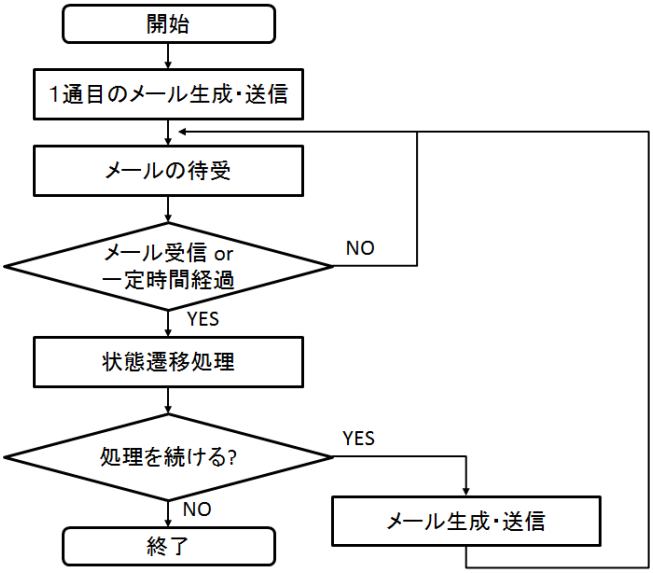

図 7　訓練対象とのメールやり取りの流れ

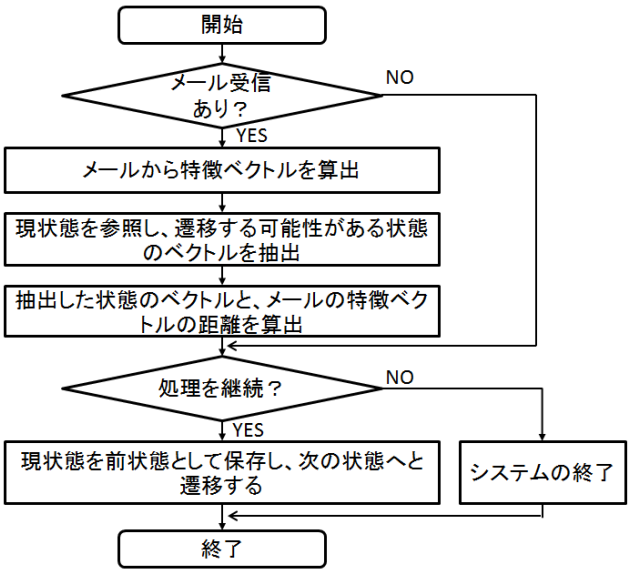

図 8　状態遷移処理の流れ

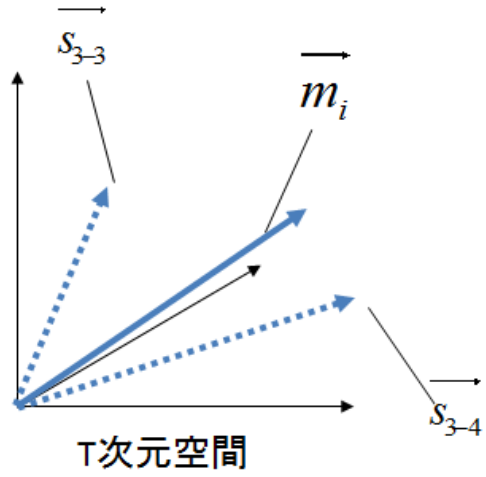

図 9　メールベクトルと遷移状態との距離算出

5. 考察

5.1 メールの収集について

自然言語処理によって文章を自動で生成できるようにするには，膨大な量の学習データであるメールが必要となる．しかし，攻撃で利用されるメールは数が少なく，学習に十分な量を集めることは困難である．そこで，攻撃の状態遷移を通常の問い合わせとのアナロジーを考える．具体的には，偵察は問合せ，攻撃は添付ファイルや本文中に URL を参照しているような状態，催促はそのまま催促と同一視することができる．この同一視により，通常の問い合わせにおけるメールのやり取りを学習データとして用いることを可能とする．これは，訓練対象組織に協力を依頼し，訓練で実施するようなメールのデータを大量に収集することが考えられる．

5.2 文章の生成について

提案手法での文章の生成では，マルコフモデルによってメール文章を生成する．しかし，学習データとして用意するメールを，5.1 で述べたような正常なやり取りをベースにした場合，やり取り型攻撃で見られる，執拗に添付ファイルを開かせたり，URL をクリックさせたりさせるような「言い訳」は通常行わない．そのため，執拗な攻撃の再現が難しい．

解決策としては，言い訳用の学習データを実際の攻撃事例から収集することや，言い訳専用のテンプレートを用意しておくことが考えられる．

5.3 実際の実施について

実際に提案手法を実現するには，クラウド上でシステムを構築することが考えられる．クラウド上に構築した本システムから，訓練対象にメールを送信し，訓練対象が訓練メールの添付ファイルや URL を開いた際に，誰が開いたかを本システムに通知する仕組みを追加しておくことで効率的に評価を行うことができる．このような，誰が開封したかを通知する仕組みは，既存の技術も備えている[2-11].

6. むすび

本稿では，標的とやり取りを行った後にマルウェアを送るやり取り型攻撃を自動でシミュレートすることができるシステムを提案した．本手法により，やり取り型攻撃を自動でシミュレートでき，訓練を実施し，組織の標的型メールに対する耐性を高めることができるようになる．

今後は，本手法を実装し，訓練を実施することができるだけのメール文章が生成されるか実験する．訓練手法の研究と合わせて，やり取り型攻撃を防ぐ手法の検討も進める．

参考文献

1) トレンドマイクロ：COMBATING MALICIOUS EMAIL AND SOCIAL ENGINEERING ATTACK METHODS, available from <https://www.trendmicro.com/cloud-content/us/pdfs/business/datasheets/ds_social-engineering-attack-protection.pdf>.
2) 縁マーケティング研究所：標的型攻撃メール対応訓練実施キット, available from <https://kit.happyexcelproject.com/>.
3) 富士通マーケティング : FUJITSU セキュリティソリューション AZSECURITY 標的型メール攻撃訓練, available from <http://www.fujitsu.com/jp/group/fjm/solutions/business-technology/security/azsecurity/targeted-attack/>.
4) NTT ソフトウェア：標的型攻撃メール訓練サービス, available from <https://www.ntts.co.jp/products/apttraining/>.
5) 大塚紹介：標的型メール訓練サービス, available from <https://www.otsuka-shokai.co.jp/products/security/consulting-education/aptmail-training-service/>.
6) IIJ：標的型メール攻撃訓練ソリューション, available from <http://www.iij.ad.jp/biz/sec-sol/targeted-m.html>.
7) LAC：IT セキュリティ予防接種, available from <http://www.lac.co.jp/education/inoculation/>.
8) STNet：標的型メール訓練サービス, available from <http://www.stnet.co.jp/security/attackmailtraining/ >.
9) GSX：標的型メール訓練サービス, available from <http://www.gsx.co.jp/informationsecurity/attackmailtraining.html>.
10) 三菱スペースソフトウェア：メルトレ, available from <https://www.mailtore.jp/>.
11) Wombat: Anti-Phishing Training Suite, available from <https://www.wombatsecurity.com/suggested-programs/anti-phishing>.
12) 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 ～, available from <https://www.ipa.go.jp/security/topics/alert20141121.html>.
13) John Seymour, Philip Tully: Weaponizing Data Science for Social Engineering, BlackHat Asia 2016.
14) mail2vec, available from <https://devpost.com/software/mail2vec>.
15) MeCab, available from < http://taku910.github.io/mecab/ >.