情報処理学会論文誌 Vol.58 No.9 1410–1418 (Sep. 2017)

攻撃者に察知されにくい情報を用いたC&Cサーバの検知手法の提案と評価

久山 真宏[1,a)] 柿崎 淑郎[1,b)] 佐々木 良一[1,c)]

受付日 2016年11月25日, 採録日 2017年6月6日

概要：近年，標的型攻撃による被害が問題になっている．標的型攻撃では，特定の企業や組織にマルウェアを感染させた後にC&Cサーバとの間で様々な通信を行い，情報を接収する．マルウェアを感染させる手口は年々巧妙化しており，マルウェアの感染を防ぐための対策だけでは不十分である．そのため，次善の対策としてC&Cサーバの通信を検知することがあげられる．しかし，C&Cサーバを検知する従来手法では，解析する過程で攻撃者に解析していることを知られてしまう危険性がある．そこで，本研究では攻撃者に解析されていることを知られずに解析する手法を提案する．本手法では，攻撃者に知られない情報としてドメインのWHOISと検索エンジンから得られた特徴と教師あり機械学習を用いてC&Cサーバの判別を行う．提案手法に実データを適用し交差検証法でC&Cサーバの判別を行った結果，約98.9%と比較的高い検知率を得ることができ，有効性の見通しを得ることができたので報告する．

キーワード：標的型攻撃，C&Cサーバ，WHOIS，SVM，ニューラルネットワーク

Proposal and Evaluation of Method for Detecting C&C Server Using Unobserved Information by Attackers

Masahiro Kuyama[1,a)] Yoshio Kakizaki[1,b)] Ryoichi Sasaki[1,c)]

Received: November 25, 2016, Accepted: June 6, 2017

Abstract: Damages caused by targeted attacks are a serious problem. It is not enough to prevent only the initial infections, because techniques for targeted attacks have become more sophisticated every year, especially those seeking to illegally acquire confidential information. In a targeted attack, various communications are performed between the command and control server (C&C server) and the local area network (LAN), including the terminal infected with malware. It is possible to find the infected terminal by monitoring the communications with the C&C server although the attackers may notice it. In this study, we propose a method for identifying the C&C server by using the feature points obtained from WHOIS and the Google Search of C&C servers' domains, which are unobserved information by attackers, for supervised machine learning. Moreover, we conduct an experiment that applies real data, and we verify the usefulness of our method by a cross-validation method. As a result of the experiment, we could obtain a high detection rate of about 98.9%.

Keywords: targeted attack, C&C server, WHOIS, SVM, neural network

1. はじめに

近年，標的型攻撃による被害が問題になっている [1]．標的型攻撃とは，金銭や知的財産などの秘密情報の不正な取得を目的として，特定の企業や組織を標的にしたサイバー攻撃の一種である．

実際に国内の大手重工メーカや衆議院，日本年金機構といった様々な組織が標的型攻撃の被害に遭い，ニュースになるほどの重大なインシデントにつながっている．

標的型攻撃の流れを**図 1** に示す．

[1] 東京電機大学
Tokyo Denki University, Adachi, Tokyo 120–8551, Japan
[a)] kuyama@isl.im.dendai.ac.jp
[b)] kakizaki@mail.dendai.ac.jp
[c)] r.sasaki@mail.dendai.ac.jp

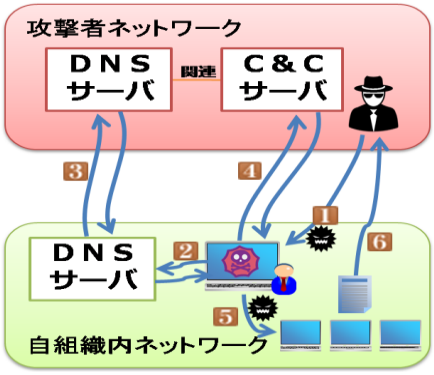

図 1 標的型攻撃の流れ
Fig. 1 Sequence of targeted attacks.

ステップ 1：標的型攻撃を行うために，LAN 内の端末にマルウェアを感染させる．
ステップ 2：マルウェアに感染した端末は，C&C サーバと接続を行うために，自組織内の DNS サーバに C&C サーバのドメインの名前解決を要求する．
ステップ 3：自組織内の DNS サーバ内に要求されたドメインに対応した IP アドレスが不明の場合，さらに上位の DNS サーバに名前解決を要求して，回答のあった IP アドレスをマルウェアに感染した端末に返す．
ステップ 4：マルウェアに感染した端末は，回答のあった IP アドレスをもとに C&C サーバと通信する．そして，目的を達成するためにより適切なマルウェアなどが端末にダウンロードされる．
ステップ 5：マルウェアは，LAN 内の他の PC やサーバに侵入範囲を拡大しようとする．
ステップ 6：重要な情報，機密情報や組織の個人情報といった目的とする情報を見つけると，攻撃者に送信される．

ステップ 1 で攻撃者から送られてくるマルウェアを検知して防衛できるのが良いが，標的型攻撃では標的となる組織ごとに検知されにくいようにカスタマイズされたマルウェアが用いられるため，検知するのは難しい．その後の一連の流れに着目すると，標的となった端末がマルウェアに感染してから攻撃者とやり取りを行うにあたり，C&C サーバは攻撃者が目的を達成するための司令塔という重要な役割を担っている．そのため，C&C サーバとの通信を検知することにより，被害を早期に発見することができる [2], [3]．

ステップ 4 の C&C サーバとの通信をファイアウォールや IPS などで遮断する手法もあるが，その場合でもステップ 3 の DNS 通信により攻撃者は解析されていることに気付き，C&C サーバを停止する危険性がある．C&C サーバの稼動期間が長ければ，長いほど解析する時間を確保することができ，より多くの情報を収集することが可能である．また，攻撃者を追跡（アトリビューション）するにも多くの情報が必要である．つまり，十分な解析を行う前に C&C サーバが停止してしまうと，その分，解析に必要な情報が得られなくなり，さらにアトリビューションを行いにくくなるといったデメリットが存在する．そのため，より深く解析を行うためには C&C サーバの生存期間が長いほうが良く，C&C サーバを解析するにも，なるべく攻撃者に解析していることに気付かれにくくすることが重要である．

そこで，本研究ではステップ 3 の通信に着目し，攻撃者が準備した C&C サーバや DNS サーバなどにアクセスすることなく収集できる情報を用いて，自組織内の DNS サーバで C&C サーバのドメインを検知し，遮断する手法を提案する．我々の研究グループでは，これまでに，メールアドレスの構造を用いる検知手法 [4] や検索エンジンから得た情報を組み合わせて検知する手法 [5] を提案してきた．本論文では，これらの方式を統合した方式を提案するとともに，新たに解析したマルウェアから得られた C&C サーバのドメインを加えた実験および評価について報告する．今回，C&C サーバにアクセスせずに得られる情報として，ドメインの WHOIS 情報と検索エンジンを用いる．これらから特徴点を抽出し，機械学習を用いて C&C サーバの検知を試みる．WHOIS と検索エンジンは，容易に情報を取得でき，さらに特定の組織が管理しているためアクセスしても攻撃者にその事実が知られにくい．そのため，本手法は攻撃者が準備したサーバ類と直接通信することなく，C&C サーバの検知が可能であり，攻撃者に解析されている事実を知られにくくできる利点がある．

2. 関連研究

2.1 類似研究

C&C サーバの特定を目的とした研究は，次の 2 種類に大別される．

(1) **C&C サーバとの通信に着目した研究**

C&C サーバとマルウェア間で行われる通信に着目し，制御通信のペイロードに含まれる文字列などの特徴を分析することで検知を行う手法 [6], [7]，テイント解析技術を応用したマルウェア解析を実施することで通信データの改ざんを検知し，C&C サーバを特定する手法 [8] などがある．

これらの手法は，実際の通信内容から検証するため，十分な検証により高い検出精度で特定することができる．しかし，ゼロデイ攻撃などの未検証な検体への対応に不十分な問題がある．

(2) **C&C サーバのドメインに着目した研究**

C&C サーバのドメインに着目し，ドメイン情報や外部リポジトリから取得した情報を併用して，RIPPER と呼ばれるデータマイニング手法を用いて検知を行う手法 [9]，WHOIS と DNS の情報から未知の悪性ドメインを推定する手法 [10]，URL の特徴や DNS，WHOIS，地理的な情報

表 1 継続調査による検出精度の変化
Table 1 Detection rates of our method over time.

モデル	検知率（年）				
	2009	2010	2011	2013	2014
2009	96.5%	85.0%	76.5%	-	-
2011	-	-	95.2%	42.5%	-
2013	-	-	-	80.3%	80.8%
2014	-	-	-	-	96.7%

から機械学習を用いて検知する手法 [11]，既知の悪性 Web サイトのコンテンツや WHOIS などの情報から検索エンジンを利用して未知の悪性ドメインを推定する手法 [12] などがある．

これらの手法は，活動中の C&C サーバに対して高い検出精度で特定することができる．しかし，C&C サーバや C&C サーバのドメインを管理する DNS サーバといった攻撃者の関与するサーバ類へリクエストが飛んでしまい，攻撃者に解析していることを検知され，攻撃者に対策されてしまう問題がある．

2.2 先行研究

当研究室では 2009 年より攻撃を受けた端末から攻撃元を追跡していき，最終的には攻撃者を特定することを目的とした多段追跡システムの研究を行っている [13]．そのなかで，数量化理論 2 類 [14] を用いてボットネットの C&C サーバを判別する手法を 2009 年に提案した．2009 年当時は 96.5%の精度でボットネットの C&C サーバを検出できていたが，継続的に調査を行ったところ，検出精度は年々下がり，2011 年には 76.5%まで検出精度が下がった [15]．これはボットネットの C&C サーバの特徴が時間経過とともに変化していることが原因である [16], [17]．そのため，一定期間ごとに最新のデータを用いて判別モデルの見直しを行ってきた．継続調査による検出精度の変化を**表 1** に，各モデルにおいて検出に用いた特徴を**表 2** に示す．

これまでの取り組みでは，ボットネットの C&C サーバを検知する研究を行っている．ボットネットの C&C サーバは 1 台の C&C サーバで複数の感染端末を同時に操作するのに対して，標的型攻撃に用いられる C&C サーバは 1 台の C&C サーバで感染端末を個別に操作する特徴がある．今回は標的型攻撃に用いられた C&C サーバの検知についてもドメインの特徴から検知可能かどうか研究する．また，従来は特徴として主に DNS 情報を用いて C&C サーバの検出を行ってきた．そのため，DNS サーバに対して通信を行う必要があり，DNS サーバを攻撃者が管理していた場合，解析していることが攻撃者に気付かれてしまう危険性があった．

攻撃者としても攻撃を成功させる必要があるため，C&C サーバを特定されて攻撃が失敗することは避けたいはずである．そのため，当該 C&C サーバが調査されていることが判明すれば，攻撃者は C&C サーバを停止し，新たに別の C&C サーバを構築することが予想される．以後は新たに構築された C&C サーバが攻撃に用いられることが予想される．これにより，C&C サーバの入れ替わりが頻繁に起こり，短期間のうちに C&C サーバの特徴が変化する要因になっていることが考えられる．また，DNS で用いる特徴は DNS サーバの設定値であるため容易に変更でき，DNS サーバの設定変更や仕様変更などによっても変化する．つまり，攻撃者に調査されていることを気付かせないことにより，C&C サーバの稼働期間を長くすることができ，さらに時間経過による特徴の変化を遅らせることができるものと考えた．そこで，DNS を用いず WHOIS に登録されている情報としてメールアドレスおよび登録期間を SVM（support vector machine）にかけて検知する手法を提案したところ，検出精度は 88.8%であった [4]．そこからさらに検索エンジンから得た情報を組み合わせてニューラルネットワークにかけて検知する手法を提案し，検出精度を 97.3%まで改善できた [5]．今回，新たに入手したマルウェアを解析して得たドメインを加えて SVM およびニューラルネットワークで実験と評価を行った．

表 2 各モデルにおける特徴の変化
Table 2 Changes in the feature used.

用いた特徴		モデル			
		2009	2011	2013	2014
DNS	逆引き	○	○	○	
	TTL				○
	minimum	○	○		○
	A レコード		○	○	
	MX レコード				
	NS レコード				○
	CNAME レコード			○	
	TXT レコード				○
WHOIS	登録期間	○	○	○	○
個数		3	4	4	5

3. 提案手法

C&C サーバなどの攻撃者が準備したサーバ類にアクセスすることなく収集できる情報を用いて C&C サーバのドメインを高精度に検出する手法を提案する．本手法を自組織内にある DNS サーバに実装することで，自組織内の DNS サーバへ名前解決のクエリからドメインを抽出し，提案手法で悪性かどうかを判別する．判別の結果，悪性と判断されたものは名前解決要求を拒否することで，その後の通信を遮断することができる．また，これにより自組織外

の DNS サーバへ通信が発生する前に通信を遮断することができ，攻撃者が関与する DNS サーバに飛ぶ前に通信を遮断することができる．

C&C サーバの判別には，ドメインの WHOIS 情報と Google の検索エンジンを用いる [5]．WHOIS とは，ドメインの登録に関する情報を管理・提供するサービスであり，RFC812 [18] および RFC3912 [19] に技術仕様や運用規則が定められている．トップレベルドメイン（TLD）のレジストラごとに特定の組織のみが運用を許可されており，WHOIS に登録されている情報は一般公開されていることから，WHOIS に登録されている情報を利用しても攻撃者は自身のドメインが WHOIS で参照されているのかどうか気付きにくい．また，WHOIS はドメインに関する連絡先や管理母体を示す情報であり，内容を変更するには運用元のレジストラに対して申請を行う必要がある．DNS よりも特徴が変化しにくく，情報を参照しても攻撃者に察知されにくいといったメリットがある．

Google も運用元が Google.inc. と特定された企業であり，提供されている情報も一般的に公開されているため，同義の理由から Google の検索エンジンから得られる情報を利用しても攻撃者は気付きにくく，さらに攻撃者の意思で特徴を変更することが困難なことから特徴が変化しにくいといったメリットがある．以上より，WHOIS と Google の検索エンジンから得られる情報を利用することとした．

WHOIS と Google の検索エンジンから得られた情報から特徴点を抽出し，機械学習を用いて C&C サーバの判定を行う．今回，悪性かどうかの 2 クラスのパターン識別として教師あり機械学習であるサポートベクタマシン（SVM）とニューラルネットワークを用いる．そのため，事前準備として，機械学習における訓練モデルを構築する．

訓練モデルの構築にあたり，まず悪性ドメインとして C&C サーバのドメイン（C&C ドメイン）と，通常の無害なドメイン（ノーマルドメイン）を準備する．そこから，各ドメインの WHOIS 情報を取得し，特徴を抽出する．抽出した特徴を機械学習で学習させ，訓練モデルを構築する．実際にアクセスする際に訓練モデルを用いてドメインの評価を行い，C&C サーバであるかどうか判別する．

3.1 評価ドメインの準備

C&C ドメインには，実際のマルウェアから抽出したドメインが最適であるため，標的型攻撃での使用率の高い Emdivi，PlugX，PoisonIvy と呼ばれる 3 種類のマルウェア [17] を収集・解析し，抽出できたドメインを利用した．

マルウェアの収集にあたっては，VirusTotal を用いて，キーワードに Emdivi，PlugX，PoisonIvy の種別名で検索を行い，2015 年 1 月～2016 年 8 月の間に投稿された計 464 件のマルウェアを収集した（**表 3**）．

収集したマルウェアは仮想環境上でマルウェアを実際に動作させて解析（動的解析）を行う Lastline で解析した．Lastline の解析結果から接続先として抽出されたドメインから，重複および WHOIS の引けないドメイン，マルウェアがインターネットに接続可能な環境かどうかを調査する目的の通信先ドメイン（たとえば Yahoo!や Google など）を排除した結果，計 89 件のドメインを得ることができたため，これを今回の実験における C&C ドメインの評価データとして用いた．

表 3 収集したマルウェアの検体数
Table 3 Collected malwares.

マルウェア種別	検体数
Emdivi	78
PlugX	311
PoisonIvy	75

ノーマルドメインには安全性の高いドメインが最適であるため，世界のアクセスランキングトップ 500 を掲載している Alexa の “The top 500 sites on the web.” に載っているドメイン 500 件を用いることとした．また，人気サイトはサイト規模が大きい傾向にあるため，特徴量に偏りが生じる可能性がある．そこで，“IR サイトランキング” に載っているドメイン 200 件と “FORTUNE” に載っている 100 件のドメインも用いた．3 サイトに載っている計 800 件のドメインの中から重複ドメインを排除して，C&C ドメインと同数となるようにランダムに 89 件のドメインを抽出し，ノーマルドメインの評価ドメインとして用いた．

3.2 特徴抽出

評価ドメインであるノーマルドメインと C&C ドメインから WHOIS と Google の検索エンジンから得られる情報を抽出する．

(1) **WHOIS からの特徴抽出**

WHOIS からは一般的に以下の情報を得ることができる．

a) 登録ドメイン名
b) レジストラ名
c) ドメインが登録されている DNS サーバ名
d) ドメインの登録年月日
e) ドメインの有効期限
f) ドメイン名登録者の連絡先
g) 技術的な連絡の担当者連絡先
h) 登録に関する連絡の担当者連絡先
i) 登録者への連絡窓口の連絡先

このなかでも，改ざんが困難なものとして a)～e) があげられる．通常のサーバであれば，長期的に運用することからドメインの登録期間は長く，逆に標的型攻撃における C&C サーバは，標的となる組織において目的が達成されればドメインを放棄するため登録期間が短い [10], [11], [12].

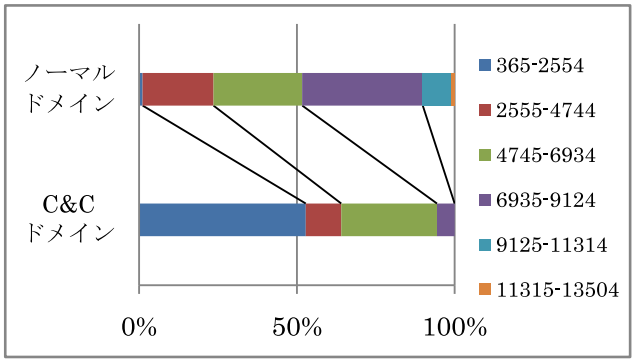

図 2 有効日数の比較
Fig. 2 Valid terms for domains.

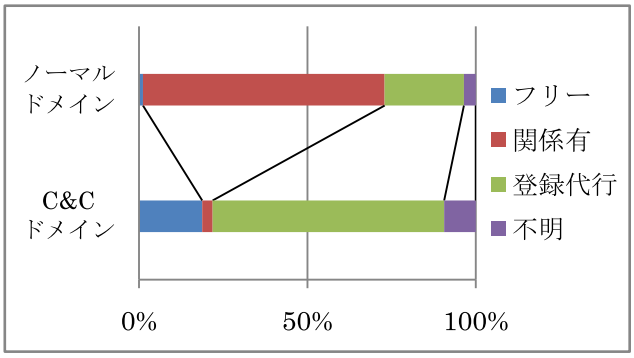

図 3 紐づくメールアドレスの比較
Fig. 3 e-mail address for domains.

このことに着目し，登録期間を割り出すため，d) の日数から e) の日数を引いた値を有効日数として用いることとした．評価ドメインの有効日数の比較を**図 2** に示す．

図 2 より，ノーマルドメインと比較して，C&C ドメインは有効日数が短いことが分かる．

他方，f)〜i) は各担当の連絡先が記載されており，以下の情報を得ることができる．

j) ID
k) 名前
l) 住所
m) 組織名
n) 郵便番号
o) 電話番号
p) 国名
q) FAX 番号
r) メールアドレス

これらは，比較的容易に秘匿や改ざんすることができる．特に C&C サーバの多くは，身元を特定されないためにドメイン登録時に WHOIS の登録を代行してくれるサービス（WHOIS 登録代行サービス）を利用して登録情報を隠蔽していたり，でたらめな情報が登録されていたりすることが多い．しかし，でたらめな情報が登録されている場合でも，r) メールアドレスは，実際に連絡を行ううえで必要なことが多いため，偽装されていない可能性が高いと考えられる．そのため，まずメールアドレスを対象に特徴点の抽出を行った [4]．比較結果を**図 3** に示す．このとき，「フリー」はメールアドレスのドメインが無料でメールサービスを提供しているドメインと一致しているものとし，「関係有」は評価ドメインとメールアドレスのドメインが同一もしくは TDL が異なるだけのもの，「登録代行」は WHOIS 登録代行サービスを提供しているドメイン提供元のドメインと一致しているもの，「不明」はそれ以外として判定した．

図 3 より，「フリー」と「登録代行」の割合は C&C ドメインが高く，「関係有」の割合はノーマルドメインが高くなる傾向にあり，ノーマルドメインと C&C ドメインにおけるメールアドレスの差異があることが判明した．

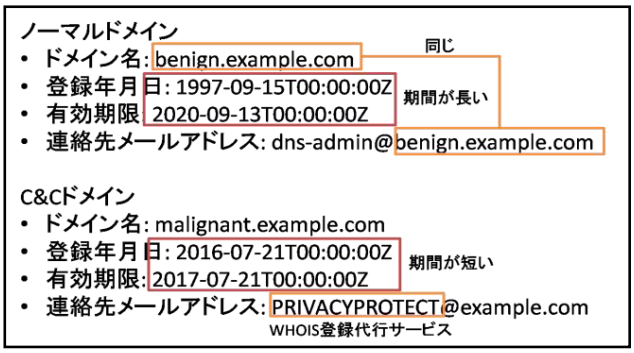

図 4 ドメインの WHOIS 例
Fig. 4 WHOIS example of domain.

以上の結果より，ノーマルドメインと C&C ドメインの WHOIS における特徴の差を**図 4** に示す．なお，倫理的な観点より実ドメインの例示は避け，特徴の差を極端に表した例を示す．図 4 より，ノーマルドメインと C&C ドメインには特徴差があるため，WHOIS よりこれらの有効日数およびメールアドレスを特徴として用いることとした．

(2) **検索エンジンからの特徴抽出**

関連研究 [12] では，既知の悪性ドメインのコンテンツなどから特徴を抽出し，抽出した結果から検索エンジンを用いて新たな悪性ドメインの発見を行っている．この研究では，ドライブ・バイ・ダウンロード攻撃における悪性ドメインの発見手法を提案しており，ドライブ・バイ・ダウンロード攻撃 [20] は Web 閲覧によって攻撃が成功する性質から集客を行うために検索エンジン最適化（SEO）を行っていることが予想される．また，閲覧させるために，正規の Web サイトを改ざんし，悪性ドメインへリダイレクトするスクリプトを仕込んでいることもある．そのため，悪性ドメインもしくは悪性ドメインへのリダイレクト元となる Web サイトは Google 検索にヒットする可能性は高いことが予想される．しかし，標的型攻撃における C&C サーバは短命であり，検索エンジンのクローラにドメインが発見される前にドメインが停止される．さらに，C&C サーバは通信を遮断されないためにも発見されないように隠れていることが予想され，検索にヒットしないことが考えら

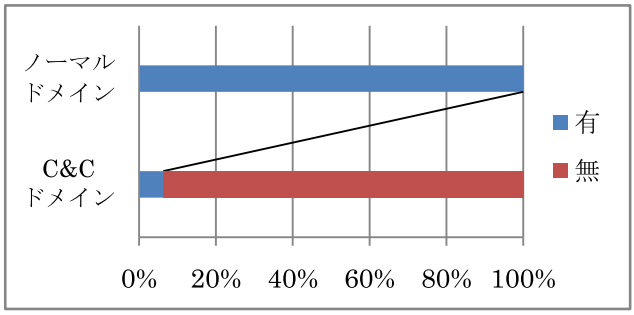

図 5 Google 検索ヒット有無
Fig. 5 Search sites finding domains.

れる．そこで，Google の検索エンジンを用いて評価ドメインが検索でヒットしたかどうか調査した [5]．Google での検索にあたっては，評価ドメインのサイト以外のサイトがヒットしないように「site:」コマンドを用いて「site:評価ドメイン」となるように検索を行い，検索結果の件数が 0 件であれば「無」，1 件以上であれば「有」とし，その結果を**図 5** に示す．

図 5 より，C&C ドメインにおいては検索にヒットしないものが多数であった．検索にヒットした C&C ドメインにおいては，改ざんもしくはサーバを乗っ取られていた可能性の高い正規のサーバであった．これは，標的型攻撃においては，正規のサーバが乗っ取られて C&C サーバ化されるよりも攻撃者が自前で用意した C&C サーバが用いられているためと考えられる．

以上の結果より，Google での検索結果を特徴として用いることとした．なお，検索のヒット有無だけで差がみられるため，特徴をより抽象化させて精度を向上させるためにヒット件数ではなくヒットの有無を用いることとした．

3.3 機械学習アルゴリズム

手法の有効性を検討するため，機械学習アルゴリズムのパラメータチューニングは行わずに実験を行うこととした．また，性能差を検討するため機械学習のアルゴリズムとして SVM（support vector machine）とニューラルネットワークの 2 種類を用いて訓練モデルの構築を行う．

SVM とは，与えられたデータからパターン認識を用いて 2 クラスの分類を行う教師あり学習の一種である [21]．関連研究 [11] において高い識別精度で判別を行っており，解析を行うデータ量が増加しても高速に識別することができる [22]．そのため，機械学習アルゴリズムの 1 つとして SVM を選択した．

ニューラルネットワークとは，脳機能にみられるいくつかの特性を数学モデル化することで，入力と出力の関係性を表現することができる教師あり学習の一種である [23]．音声や文字などの識別にも使用されており [24], [25]，誤差逆伝播法 [26] を用いることで入力と出力のあいだにどういった関係があるのかを表現することができる [27]．そのため，

表 4 機械学習への入力値
Table 4 Features of machine learning.

特徴		入力値
ラベル		ノーマル C&C
ドメイン		(文字列)
メールアドレス	ローカルパート	(文字列)
	ドメイン	(文字列)
	タイプ	フリー 関係有 登録代行 不明
有効日数		(数値)
検索エンジン		有 無

単なる数値での識別ではなく，WHOIS 情報と Google の検索エンジンからの特徴と C&C サーバとの関係を学習して識別することに期待して，機械学習アルゴリズムの 1 つとしてニューラルネットワークを選択した．

各アルゴリズムにおける訓練モデルを構築する前段階として，メールアドレスは “@” で区切って前半部分のローカルパートと後半部分にドメインに分割，評価ドメインとメールアドレスのドメインとの間での関係の有無，フリーメールアドレスの使用有無，WHOIS 登録代行サービスの使用有無を調査する．さらにドメインの有効期限年月日と登録年月日から有効日数を算出して，Google の検索エンジンを用いて評価ドメインが検索にヒットするか調査しておく（**表 4**）．これらの情報をテストデータとして各アルゴリズムに学習させて訓練モデルを構築する．

メールアドレスは “@” の前半部分はローカルパート，後半部分はドメインを表しており，それぞれで意味が異なっているため分割して入力値とした．さらに，メールアドレスの各タイプにおいて，ローカルパートおよびドメインに特徴があることが考えられる．たとえば C&C サーバで用いられる登録代行サービスやフリーメールアドレスにも攻撃者にとって都合の良い特定の業者が用いられることが考えられる．そのため，メールアドレスはタイプとあわせてローカルパートおよびドメインをセットで入力値とした．

4. 評価

今回，評価に用いるデータ量が少ないため，実際に訓練モデルを構築するための訓練データと評価に用いるテストデータを準備する評価手法では，テストデータの選び方によって精度に大きな差が生じる可能性がある．そこで，評価に用いるデータ量が少なくても比較的誤差を少なくでき

表 5 評価結果（交差検証法）
Table 5 Detection rates by cross-validation.

入力値の組み合わせ	SVM				ニューラルネットワーク			
	推定精度	TPR	TNR	処理時間	推定精度	TPR	TNR	処理時間
メール	90.4%	91.0%	89.9%	0.03 sec	88.8%	91.0%	89.9%	117.42 sec
有効日数	70.2%	76.4%	64.0%	0.02 sec	68.5%	71.9%	65.2%	26.19 sec
検索エンジン	96.6%	100%	93.3%	0.03 sec	96.6%	100%	93.3%	25.91 sec
メール + 有効日数	93.8%	94.4%	93.3%	0.03 sec	93.3%	93.3%	93.3%	108.25 sec
メール + 検索エンジン	97.8%	100%	95.5%	0.03 sec	97.8%	100%	95.5%	108.09 sec
有効日数 + 検索エンジン	96.6%	100%	93.3%	0.02 sec	96.6%	100%	93.3%	26.41 sec
メール + 有効日数 + 検索エンジン	**98.9%**	**100%**	**97.8%**	**0.02 sec**	**98.9%**	**100%**	**97.8%**	**126.33 sec**

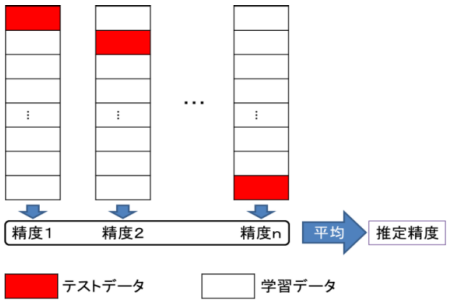

図 6 交差検証法
Fig. 6 Cross-validation method.

る手法である交差検証法を用いて評価を行う [28].

交差検証法とは，学習データとなる元のデータを一定のブロック単位に分割し，1 つのブロックをテストデータ，その他のブロックを学習データとして評価を行う．分割したブロックごとに評価を行い，各評価結果の平均を推定精度として算定する手法である（**図 6**）．この方法を用いることにより，データ量が少なくても，推定される精度の誤差を少なくすることができ，以下の数式において求めることができる．このとき，テストデータの総数は N^{ts}，正確に分類された総数は t^{ts}，n 回目の評価精度は $A^{ts}(d^n) = \frac{t^{ts}}{N^{ts}}$，求めたい推定精度は $A^{CV}(d)$ とする．

$$A^{CV}(d) = \frac{1}{n}\sum_{i}^{n} A^{ts}(d^i)$$

交差検証法において SVM およびニューラルネットワークで構築した訓練モデルを評価した結果を**表 5** に示す．今回，評価データを 10 分割し，そのうちの 1 つをテストデータ，残りを学習データとして 10 回評価を行い導き出された精度の平均を推定精度とした．また，ノーマルドメインと正しく識別された精度を TPR，C&C ドメインと正しく識別された精度を TNR として算出した．

$$\text{TPR} = \frac{\text{正しくノーマルドメインと識別された数}}{\text{ノーマルドメインの総数}}$$

$$\text{TNR} = \frac{\text{正しく C\&C ドメインと識別された数}}{\text{C\&C ドメインの総数}}$$

ノーマルドメインの選定による偏りが発生していないかどうかを調査するため，3.1 節で収集した 800 件のドメインの中から重複ドメインを排除して，ランダムに 89 件のドメインを抽出し，交差検証法を用いた評価を複数回行った．その結果，SVM，ニューラルネットワークともに推定精度に大きな変化は見受けられなかったため，本手法を採用し，ノーマルドメイン 89 件で実験を行った．

入力値はラベルとドメイン，メールアドレス，有効日数，検索エンジンの 5 種類に分類される．このなかで重みづけを行うために，種類ごとに組合せを変えて評価を行った．なお，ラベルとドメインは識別を行うのに必須であるため，どの組合せにおいても入力値として用いている．

評価結果より，SVM およびニューラルネットワークどちらにおいても大きな差異はなく比較的高い推定精度を導き出せた．これは，多段追跡システムの 2014 年モデルの検知率 96.7%を上回る結果である．また，従来は時間経過とともに検出精度は下がる傾向にあったが，今回，新たに解析したマルウェアから得られた C&C サーバのドメインを加えた結果，97.3% [5] を上回る精度であった．これは，時間経過による検出率低下への耐性に期待できる．

SVM とニューラルネットワークの推定精度に大きな差異はなかったものの，処理時間では SVM が高速に処理することができた．特に SVM は入力値が増えても当初の期待通り，比較的高速に安定した速度で処理できたといえる．他方，ニューラルネットワークはどの入力値の組合せにおいても SVM を超える推定精度は出ておらず，当初期待していた各種の特徴と C&C サーバとの関係を学習することで単なる数値の識別よりも高い精度を出せたとはいい難い．

表 5 から入力値として検索エンジンの結果が C&C サーバの識別を行うにあたり最も重要な役割を担うことが明らかになった．次いでメールアドレス，有効日数の順に有効

であることが分かった．

検索エンジンの結果に有効日数を加えても検知率に変化はないが，メールアドレスと検索エンジンの結果に有効日数を加えたところ，検知率が改善した．これは，単体での識別では検知できなかったC&Cドメインを3種類の特徴を組み合わせることで検知可能になったことを意味する．つまり，メールアドレスと有効日数，検索エンジンの結果を組み合わせることが有効であることが示された．

2009年からC&Cサーバの検知における継続調査では，有効日数は経年変化に耐性のあるとても重要な役割を持っていた．これは，C&Cサーバは基本的に短命であるため，ドメインの有効日数は短い傾向が変化していないことに由来される．C&Cサーバが短命であれば，検索エンジンにドメイン情報が収集される前段階でドメインが無効となり検索にヒットしないため，有効日数と同様に経年変化へ耐性があるものと考えられる．

WHOISに登録されているメールアドレスにも特徴があった．これは，C&Cドメインにおいて，特定のWHOIS登録代行サービスへの偏り，フリーメールアドレスが使用されていることが大きな要因として考えられる．攻撃者がC&Cサーバを準備する際により容易に安価に構築可能なサービスを用いているために特徴が出ているものと推測され，攻撃者が攻撃にかける費用や労力が増えなければ変動しにくい．

5. おわりに

本論文では，C&Cサーバなどの攻撃者が準備したサーバ類にアクセスすることなく収集できる情報であるWHOIS情報と検索エンジンから特徴を抽出し，機械学習にかけることにより，C&Cサーバの判別ができることを示した．これにより，攻撃者に解析していることを知られずにC&Cサーバを検知できたものと考える．さらに，マルウェアから抽出した実データを用いて評価した結果，従来手法より高い検知率でC&Cサーバを判別することができた．

従来はC&Cサーバを特定するために通信内容や通信先であるC&Cサーバに直接アクセスまたは接続の際に名前解決により攻撃者が管理するサーバ類へアクセスすることで，攻撃者に解析されていることに気付かれ，対策される危険性があった．本論文では，自組織内のDNSサーバに提案手法を適用することで攻撃者が管理するDNSサーバへの名前解決を行う前に通信を遮断することができる．そのため，攻撃者に知られることなくC&Cサーバを判別できる本手法は有効な手法であるといえる．

実際に標的型攻撃で用いられるマルウェアを収集して解析した結果では，正規のサーバが乗っ取られてC&Cサーバ化している例は少なかった．しかし，正規のサーバが乗っ取られてC&Cサーバ化した場合，WHOISに正規のユーザの情報が登録されており，さらに，Google検索にヒットしやすくするためSEOが行われていることが多いため，WHOIS情報やGoogleの検索結果からでは差異が出にくく，誤検知が多くなることが懸念される．そのため，正規サーバが乗っ取られたケースでは，他の手法を組み合わせて検知するといった対策が必要になると考えられる．

今後は標的型攻撃以外で用いられているC&Cサーバや，経年経過による特徴変化に対しても提案手法が有効であるか調査するとともに，正規サーバが乗っ取られたケースなどに対処できる手法との組合せについて検討する．さらに，これらを実装し，実環境での運用を通して処理時間や分析性能といった実用面からの検討を行う．

参考文献

[1] 標的型攻撃等の脅威について，入手先 ⟨http://www.nisc.go.jp/conference/suishin/ciso/dai18/pdf/2.pdf⟩（参照 2016-08-01）．

[2] 標的型攻撃対策指南書（第1版），入手先 ⟨http://www.lac.co.jp/anti-apt/guidebook/pdf/anti-apt_guidebook_ver1.pdf⟩（参照 2016-08-01）．

[3] 「高度標的型攻撃」対策に向けたシステム設計ガイド，入手先 ⟨https://www.ipa.go.jp/files/000046236.pdf⟩（参照 2016-08-01）．

[4] 久山真宏，佐々木良一：ドメインのWHOIS構造を用いた悪性ドメインの判別手法，*DICOMO2016* (2016).

[5] 久山真宏，柿崎淑郎，佐々木良一：攻撃者に察知されにくい情報を用いたC&Cサーバの判別手法，コンピュータセキュリティシンポジウム2016論文集，pp.625–631 (2016).

[6] Jang, D.I., Kim, M., Jung, H.C. and Noh, B.N.: Analysis of HTTP2P Botnet, Case Study Waledac, *2009 IEEE 9th Malaysia International Conference on Communications* (*Micc*), pp.409–412 (2009).

[7] Lu, W., Tavallaee, M. and Ghorbani, A.A.: Automatic Discovery of Botnet Communities on Large-Scale Communication Networks, *ASIACCS '09 Proc. 4th International Symposium on Information, Computer, and Communications Security* (2009).

[8] 幾世知範，青木一史，八木　毅，針生剛男：改ざんデータの出自確認に基づいたC&Cサーバ特定手法の提案，2014年電子情報通信学会ソサイエティ大会通信 (2)，pp.6–16 (2014).

[9] Tsai, M.H., Chang, K.C., Lin, C.C., Mao, C.H. and Lee, H.M.: C&C Tracer: Botnet Command and Control Behavior Tracing, *IEEE International Conference on Systems, Man and Cybernetics* (*SMC*), pp.1859–1864 (2011).

[10] Felegyhazi, M., Kreibich, C. and Paxson, V.: On the Potential of Proactive Domain Blacklisting, *USENIX Conference on Large-scale Exploits and Emergent Threats*, p.6 (2010).

[11] Ma, J., Saul, L.K., Savage, S. and Voelker, G.M.: Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1245–1254 (2009).

[12] Invernizzi, L., Benvenuti, S., Comparetti, P.M., Cova, M., Kruegel, C. and Vigna, G.: EvilSeed: A Guided Approach to Finding Malicious Web Pages, *IEEE Symposium on Security and Privacy*, pp.428–442 (2012).

[13] 三原　元，佐々木良一：数量化理論と攻撃データ

（CCCDATAset2009）を利用したボットネットの C&C サーバ特定手法の提案と評価，情報処理学会論文誌，Vol.51, No.9, pp.1579–1590 (2010).
[14] 林知己夫：数量化—理論と方法（統計ライブラリー），朝倉書店 (1993).
[15] 中村暢宏，佐々木良一：累積データを用いたボットネットの C&C サーバ特定手法の評価，コンピュータセキュリティシンポジウム 2011 論文集，pp.456–461 (2011).
[16] 岡安翔太，佐々木良一：ボットネットの C&C サーバ特定手法における数量化理論と機械学習での評価と提案，*DICOMO2015*, pp.991–917 (2015).
[17] Okayasu, S. and Sasaki, R.: Proposal and Evaluation of Methods Using the Quantification Theory and Machine Learning for Detecting C&C Server Used in a Botnet, *2015 IEEE 39th Annual Computer Software and Applications Conference* (*COMPSAC*), pp.24–29 (2015).
[18] RFC954 NICNAME/WHOIS, available from ⟨https://www.ietf.org/rfc/rfc954.txt⟩ (accessed 2016-08-01).
[19] RFC3912 WHOIS Protocol Specification, available from ⟨http://www.ietf.org/rfc/rfc3912.txt⟩ (accessed 2016-08-01).
[20] 「ウェブサイトを閲覧しただけでウイルスに感染させられる “ドライブ・バイ・ダウンロード” 攻撃に注意しましょう!」，入手先 ⟨http://www.ipa.go.jp/files/000008801.pdf⟩（参照 2016-08-01）.
[21] Vapnik, V. and Lerner, A.: Pattern recognition using generalized portrait method, *Automation and Remote Control*, Vol.24, pp.774–780 (1963).
[22] John, P.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Technical Report, MSR-TR-98-14, pp.1–21 (1998).
[23] Multilayer Perceptron, available from ⟨http://deeplearning.net/tutorial/mlp.html⟩ (accessed 2016-08-01).
[24] Toshiteru, H., Les, A. and Robert, M.: An Artificial Neural Network for Spatio-Temporal Bipolar Patters: Application to Phoneme Classification, *Advances in Neural Information Processing Systems*, pp.31–40 (1988).
[25] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D.: Backpropagation applied to handwritten zip code recognition, *Neural Computation*, Vol.1, pp.541–551 (1989).
[26] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.: Learning representations by backpropagating errors, *Nature*, Vol.323-9, pp.533–536 (1986).
[27] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, MIT Press (1986).
[28] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proc. 14th International Joint Conference on Artificial Intelligence*, Vol.2, No.12, pp.1137–1143 (1995).

久山 真宏 （正会員）

2015 年東京電機大学大学院先端科学技術研究科博士課程（後期）入学．現在，セキュリティと人工知能の研究に従事．

柿崎 淑郎 （正会員）

2003 年東海大学工学部電子工学科卒業．2008 年同大学大学院博士課程修了．博士（工学）．2008 年 4 月より，東京理科大学工学部第一部電気工学科助教．2013 年 4 月より，東京電機大学未来科学部情報メディア学科助教．電子認証技術，情報システムとしての情報セキュリティ応用等の研究に従事．電子情報通信学会，IEEE 各会員．

佐々木 良一 （正会員）

1971 年 3 月東京大学卒業．同年 4 月日立製作所入社．システム開発研究所でシステム高信頼化技術，セキュリティ技術，ネットワーク管理システム等の研究開発に従事．2001 年 4 月より東京電機大学教授，工学博士（東京大学）．2002 年に情報処理学会論文賞，2007 年および 2017 年に総務大臣表彰等を受賞．著書に，『IT リスクの考え方』（岩波新書，2008 年）等．日本セキュリティ・マネジメント学会会長，内閣官房サイバーセキュリティ補佐官等を歴任．本会フェロー．