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

久山真宏[†1]　柿崎淑郎[†1]　佐々木良一[†1]

概要：近年，標的型攻撃による被害が問題になっている．標的型攻撃における手口は年々巧妙化しており，攻撃を防ぐための対策だけでは不十分である．標的型攻撃では，マルウェアに感染した後に C&C サーバとの間で様々な通信を行う．そのため，出口対策として C&C サーバの通信を監視することにより，被害を発見することが出来る．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として，ドメインの WHOIS と検索エンジンから得られる情報から教師あり機械学習を用いて C&C サーバの判別を行う手法を提案する．提案手法に実データを適用し交差検証法にて C&C サーバの判別を行った結果，約 97.3%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

A Method for detecting C & C server using unobserved information by attackers

Masahiro Kuyama[†1]　Yoshio Kakizaki[†1]　Ryoichi Sasaki[†1]

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. Therefore, it is possible to find the infected terminal in the LAN by monitoring the communications with the C&C server. In this study, we propose a method for identifying the C&C server by using supervised machine learning and the feature points obtained from WHOIS and the Google Search of domains of C&C servers and normal domains. 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 97.3%.

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，金銭や知的財産等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．ドライブバイダウンロード攻撃や，メールなどに添付されたマルウェアに感染することによって，情報の搾取や破壊活動が行われる．

日本では，国内の大手重工メーカや衆議院，日本年金機構などにおいて，標的型攻撃の被害に遭い，実際にニュースになるほどの重大なインシデントに繋がっている．標的型攻撃の流れを図 1 に示す．

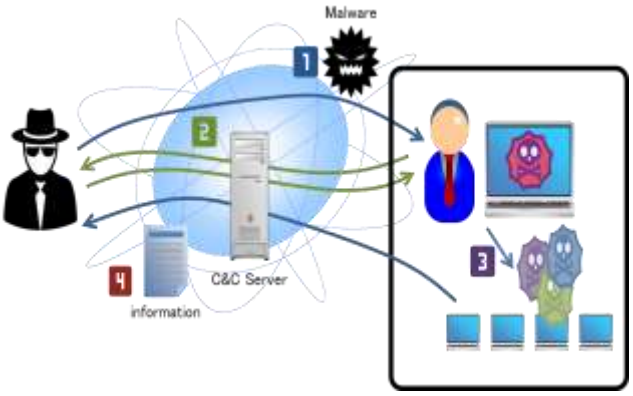

図 1　標的型攻撃の流れ

ステップ 1：標的型攻撃を行うために，LAN 内の PC 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．

†1 東京電機大学
Tokyo Denki University

ステップ 3：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 4：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．

標的型攻撃では，ステップ 1 で標的となる企業や組織ごとにカスタマイズされたマルウェアが用いられ，セキュリティ製品で検知できないことが多い．そのため，マルウェアを検知して未然に感染を防ぐ入口対策だけでは不十分であり，出口対策などの多層的・重層的な対策が求められている．

標的方攻撃の一連の流れの中でも，C&C サーバはマルウェアと攻撃者の間に位置し，攻撃者が目的を達成するための重要な役割を担っている．そのため，出口対策として C&C サーバの通信を監視することにより，被害を早期に発見することが出来る[2][3].

C&C サーバを特定するにはマルウェアおよびその通信や通信先を解析する必要があるが，調査・研究の過程で C&C サーバや C&C サーバを管理するサーバ類へアクセスを行うことがある．これでは，攻撃者が解析していることに気づき，C&C サーバが停止されてしまう危険性がある．C&C サーバの稼動期間が長ければ長いだけ解析する時間も確保することができるため，より多くの情報を収集することが可能である．また，攻撃者を追跡するにも時間が必要であり，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している．そのため，より深く多角的に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を気づかれにくくすることが求められる．

そこで，本研究では，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せずに C&C サーバの検知を試みる．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 Google の検索エンジンを用いる．特に，WHOIS 情報の中でも偽装が困難だと考えられるドメインの登録日および有効期限，メールアドレス，また Google 検索の結果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．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，攻撃者に解析されている事実を知られにくく出来る利点がある．

2. 関連研究

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[4][5]，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[6]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[7]，WHOIS と DNS の情報から未知の悪性ドメインを推定する手法[8]，URL の特徴や DNS，WHOIS，地理的な情報から機械学習を用いて検知する手法[9]，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[10]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，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．

当研究室では 2009 年より C&C サーバ，攻撃者の特定を目的とした多段追跡システムの研究を行っている[11]．その中で，ドメインから得られる情報から数量化理論 2 類[12]を用いて C&C サーバを判別する手法の研究を行っている．その後，本研究について継続的に調査を行ったところ，2009 年当時は 96.5%であった検出精度が，年々下がっていき，2011 年には 76.5%まで検出精度が下がった[13]．これは 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[14][15]．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，用いた特徴の変化を表 2 に示す．

表 1　継続調査による検出精度の変化

モデル	検知率（%）				
	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

表 2　各モデルにおける特徴の変化

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

2009 年から 2014 年の取り組みでは特徴に DNS を用いている．そのため，DNS サーバに対して通信を行ってしまうことで攻撃者に気付かれてしまう危険性がある．

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次第，C&C サーバを停止し，別の C&C サーバを構築していることが予想される．C&C サーバの入れ替わりが頻繁に起こることにより，短期間のうちに特徴が変化する要因として考えられる．そのため，攻撃者に調査されていることを気づ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，時間経過による特徴の変化を遅らせることが出来ると考える．

3. 提案手法

C&C サーバを特定するために通信内容を解析し，通信先に直接アクセスすることで，攻撃者が解析されていることに気づかれる危険性がある．そのため，攻撃者が準備したサーバ類に直接アクセスすることなく，C&C サーバを特定する検知方法が必要となる．本提案手法は，C&C サーバのドメインに着目した検知手法である．

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．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[16]および RFC3912[17]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しており，WHOIS に登録されてある情報は一般公開されていることから WHOIS に登録されてあ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づきにくい．また，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づきにくい．そのため，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

ノーマルドメインには，安全性が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.”[18]に載っているドメインを利用した．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”[19]と”FORTUNE”[20]も利用した．これらのランキングに記載されているドメインの中からランダムに 85 件を抽出し，ノーマルドメインにおける評価ドメインとした．

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[21]を用いて，キーワードに Emdivi，PlugX，PoisonIvy の種別名で検索を行い，計 163 件のマルウェアを収集（表 3）．

表 3　収集したマルウェアの検体数

マルウェア種別	検体数
Emdivi	50
PlugX	63
PoisonIvy	50

収集したマルウェアを LastLine[22]と呼ばれるマルウェアを仮想環境上にて実際に動かして解析（動的解析）を行う Sandbox を用いて解析を実施した．解析結果より，マルウェアが通信を行う接続先ドメイン 64 件を抽出し，評価データとして利用した．

3.2 特徴抽出

(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出来る．

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
j) ID
k) 名前
l) 組織名
m) 住所
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

この中でも，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[23][24][25]．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（有効日数）を用いることとした．評価ドメインの有効日数の比較を図 2 に示す．

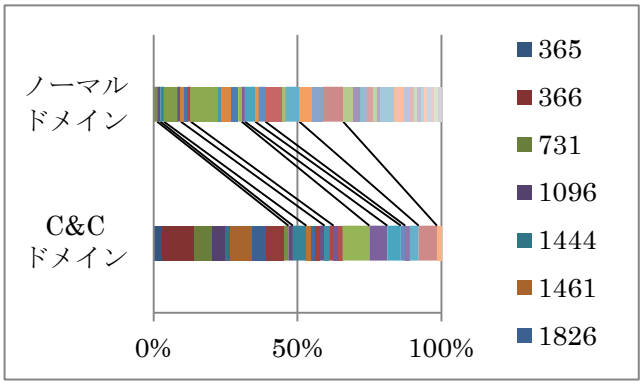

図 2　有効日数の比較

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わかる．

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．

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．まず，ノーマルドメインと C&C ドメインの WHOIS に登録されてあるメールアドレスをデータマイニングにかけて，構造の特徴を抽出した．ノーマルドメインに紐づくメールアドレスの共起ネットワークを図 3，C&C ドメインに紐づくメールアドレスの共起ネットワークを図 4 に示す．

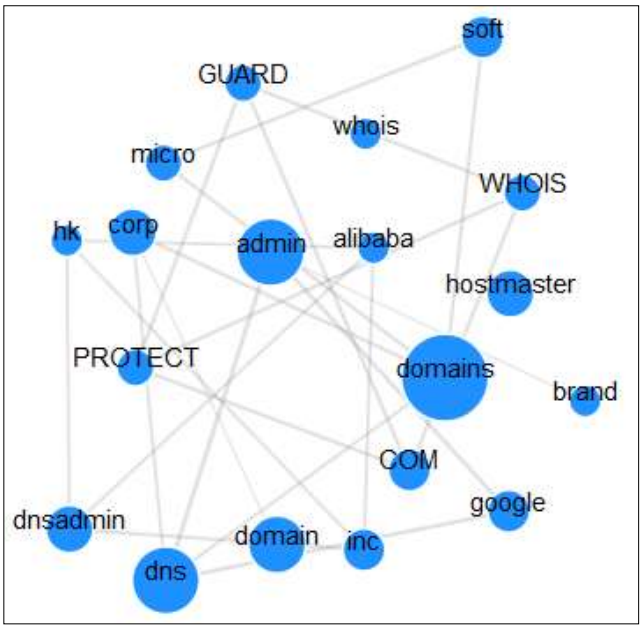

図 3　メールアドレスの共起ネットワーク
（ノーマルドメイン）

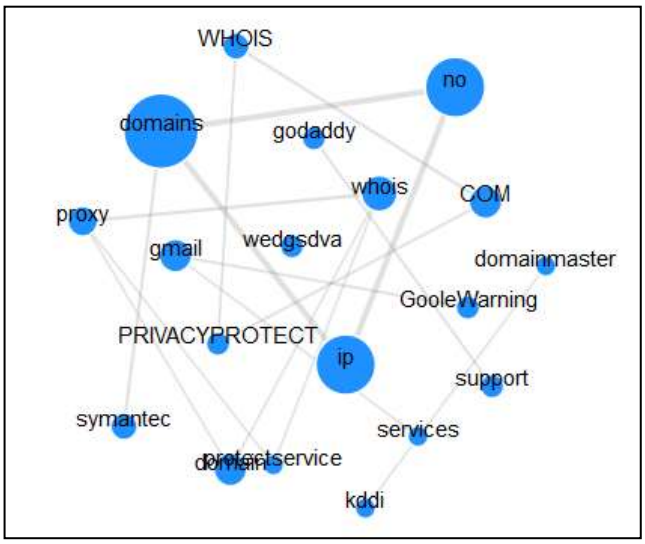

図 4　メールアドレスの共起ネットワーク（C&C ドメイン）

比較すると，ノーマルドメインの共起ネットワークは複数種類の単語が広く相互に関係を持つ大きなネットワークが存在しているのに対して，C&C ドメインの共起ネットワークは一定の単語のみが相互に関係を持つ小さなネットワークが複数存在している．C&C ドメインの小さなネットワーク一つ一つに着目すると，各かたまりの中に”no”や”PROTECT”，”proxy”といった WHOIS 登録代行サービスがよく用いる単語が含まれていた．これは，C&C ドメインでは WHOIS 登録代行サービスが特定の業者に偏っていることが原因であると考えられる．紐づくメールアドレスの比較を図 5 に示す．この時，「フリー」はメールアドレスがフリーメールアドレス，「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関係会社のドメイン，「登録代行」は WHOIS 登録代行サービスのことを指す．

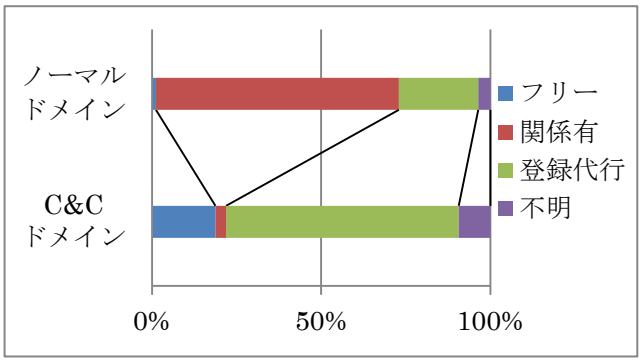

図 5　紐づくメールアドレスの比較

「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．

以上の結果より，WHOIS より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[10]では，既知の C&C サーバ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 C&C サーバ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 C&C サーバ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[26]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，C&C サーバ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，C&C サーバもしくは C&C サーバ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と予想される．しかし，標的型攻撃における C&C サーバは短命であることから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，検索にヒットしないことが考えられる．そこで，Google の検索エンジンを用いて評価ドメインを検索して，ヒットしたかどうか調査した．その結果を図 6 に示す．

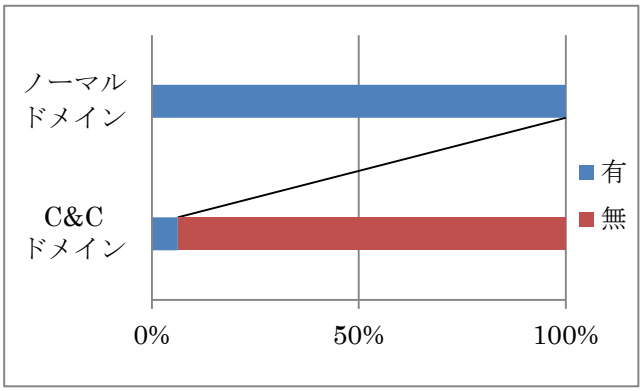

図 6　Google 検索ヒット有無

仮説どおり 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．

3.3　機械学習アルゴリズム

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[27]．関連研究[9]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[28]．そのため，機械学習アルゴリズムの一つとして 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[29]．音声や文字などの識別にも使用されており[30][31]，誤差逆電伝播法[32]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出来る[33]．そのため，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されることに期待して，機械学習アルゴリズムの一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処理として，ドメインとメールアドレスは，データマイニングを用いて構造化し，評価ドメインとメールアドレスの関係や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，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．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．

実際に構築した訓練モデルを用いて検知を行う際は，接続を試みるドメインの WHOIS より，訓練モデルを構築する際に利用した情報を抽出し，それをもとに訓練モデルを用いて判定を行う．

4. 評価

今回，用いるデータ量が少ないため，実際に評価に用いるテストデータを準備しての評価では，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．特に，標的型攻撃に用いられるドメインは，提供データが少なく，不足するため，データ量が少なくても比較的誤差を少なくできる手法である交差検証法を用いて評価を行う[34]．

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，一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 7）．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．この時，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．

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

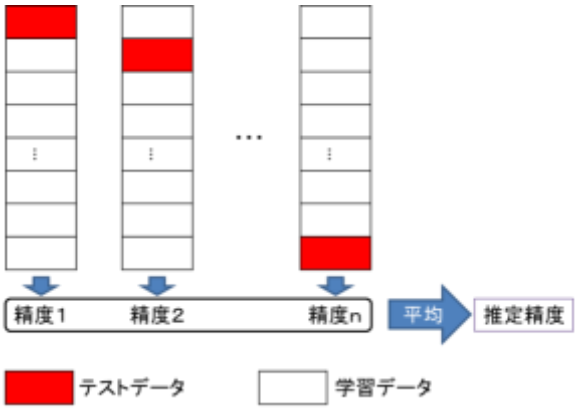

図 7　交差検証法

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 4 に示す．今回，評価データを 10 分割し，そのうちの一つをテストデータ，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．

表 4　評価結果（交差検証法）

	SVM	ニューラルネットワーク
推定精度	97.3%	97.3%

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比較的高い検知率を導き出せた．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．

5. おわりに

提案手法により，C&C ドメインに用いられるメールアドレスの特徴点を明らかにし，ドメインの有効日数や検索エンジンでの検索結果と組み合わせて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．また，抽出したメールアドレスに用いられている単語の関係を共起ネットワークで示すことにより，C&C ドメインに利用されやすい WHOIS 登録代行サービスにおいても，特徴があることを示した．

本手法は攻撃者の準備したサーバへアクセスすることなく，C&C ドメインの推定が可能である．これにより，攻撃者に解析していることを検知されにくくすることができ，時間経過による C&C ドメインの特徴変化を遅らせること

が可能になると考える．

また，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違いが出にくい．そのため，誤検知が多くなることが考えられる．しかし，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 サーバと化すことが少ないため，本手法が有効な手段であると考えられる．

今回，ドメイン名および WHOIS に登録されてあるメールアドレスやドメインの有効日数，Google の検索エンジンで得られた検索にヒットした数を入力値として，機械学習を用いて検証を行った．今後はより最適な入力値がないか調査し，高い検知精度を目指すとともに，本手法をブラウザや Proxy などへの実装・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” 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）” 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” https://www.ipa.go.jp/files/000046236.pdf, (参照 2016-08-01).

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

[5]Wei. Lu, M. Tavallaee, Ali. A. Ghorbani. Automatic Discovery of Botnet Communities on Large-Scale Communication Networks. ASIACCS '09 Proceedings of the 4th International Symposium on Information, Computer, and Communications Security(2009).

[6]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).

[7]M. H. Tsai, K. C. Chang, C. C. Lin, C. H. Mao, H. M. Lee. C&C Tracer. Botnet Command and Control Behavior Tracing, in IEEE International Conference on Systems, Man and Cybernetics (SMC), Anchorage, AK, pp.1859-1864(2011).

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

[9]J. Ma, L. K. Saul, S. Savage and G. M. Voelker. 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).

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

[11]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 51(9), pp.1579-1590(2010).

[12]林知己夫，数量化―理論と方法 (統計ライブラリー)，朝倉書店, 233p(1993).

[13]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).

[14]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).

[15]O.Shota, S.Ryoichi, 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).

[16]“RFC954 NICNAME/WHOIS” https://www.ietf.org/rfc/rfc954.txt, (参照 2016-08-01).

[17]“RFC3912WHOIS Protocol Specification” http://www.ietf.org/rfc/rfc3912.txt, (参照 2016-08-01).

[18]“Alexa Top 500 Global Sites” http://www.alexa.com/topsites, (参照 2016-08-01).

[19]“IR サイトランキング” http://www.gomez.co.jp/ranking/ir/index.html, (参照 2016-08-01).

[20]“FORTUNE” http://fortune.com/, (参照 2016-08-01).

[21]“VirusTotal” https://www.virustotal.com/, (参照 2016-08-01).

[22]“LastLine” https://www.lastline.com/, (参照 2016-08-01).

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

[24]J. Ma, L. K. Saul, S. Savage and G. M. Voelker. 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).

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

[26] “「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！」”http://www.ipa.go.jp/files/000008801.pdf, (参照 2016-08-01).

[27]V.Vapnik, A.Lerner, Pattern recognition using generalized portrait method, Automation and Remote Control.24, pp.774-780(1963).

[28]P.John, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Technical Report MSR-TR-98-14, pp.1-21(1998).

[29]“Multilayer Perceptron”http://deeplearning.net/tutorial/mlp.html, (参照 2016-08-01).

[30]H.Toshiteru, A.Les, M.Robert, An Artificial Neural Network for Spatio-Temporal Bipolar Patters: Application to Phoneme Classification, Advances in Neural Information Processing Systems 1, pp.31–40(1988).

[31]Y.LeCun, B.Boser, J.S.Denker, D.Henderson, R.E.Howard, W.Hubbard, L.D.Jackel, Backpropagation applied to handwritten zip code recognition, Neural Computation 1, pp.541-551(1989).

[32]Rumelhart,D.E., Hinton,G.E., Williams,R.J., Learning representaions by backpropagating errors, Nature Vol.323-9, pp.533-536(1986).

[33]Rumelhart,D.E., Hinton,G.E., Williams,R.J., Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, MIT Press, 570p(1986).

[34]R.Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2 (12), pp.1137-1143 (1995).