インシデント後におけるログ解析での機械学習を用いた悪性ドメインの抽出手法の提案

松岡裕和[1]　　佐々木良一[†1]

概要：近年，WEB を経由して企業や組織の情報が流出する標的型メール攻撃や Drive by Download 攻撃が問題になっている．本研究では，インシデントが発生した後のプロキシサーバのログ解析において，これらの攻撃で用いられる悪性サイトを，機械学習を用いて統合的に特定できるようにすることを目的とする．既知の悪性サイトと類似の特性を持つサイトを悪性サイトとして判別する 2 通りの識別モデルを作成し，一般サイトと悪性サイトのドメインの分類実験を行った結果，サポートベクターマシンを使う場合は 98%以上のドメインを正しく分類することができることを示した．さらに，プロキシサーバの実際のログに，9 件の悪性サイトのデータを加えたものに対し作成した識別モデルを適用した結果，すべての悪性サイトのドメインを見逃すことなく判別することができた．ログ解析処理も現実的時間でできる見通しがたったがさらなる高速化が今後の課題であることを示した．

Proposal of extraction method of malicious domain using machine learning for log analysis after incident

HIROKAZU MATSUOKA[1]　RYOICHI SASAKI[1]

1. はじめに

近年，標的型攻撃により，特定の企業や組織の情報が流出する事案が発生している．標的型攻撃では，送信元を偽装し，悪性のファイルを添付したメールを閲覧者に開かせてマルウェアに感染させたのち，C&C サーバ経由で攻撃用ツールをダウンロードする攻撃がある．一方で，Web サイトなどに不正なソフトウェアを隠しておき，閲覧者がアクセスすると自動でマルウェアをダウンロード・実行する Drive by Download（以下，DbD）攻撃などがある．DbD 攻撃で用いられる，改ざんされた正規のサイトにアクセスした場合，サイトの見た目の変化もないため，感染の事実に気づかないことがあり，組織内にマルウェアが継続して潜伏する可能性がある．

このような悪性サイトへアクセスしたログを調査する際，プロキシサーバのログ（以下プロキシログ）を調査する場合がある．プロキシログの中から悪性サイトへアクセスしたログが見つかれば，対応する送信元 IP アドレスから悪性サイトへ接続した端末が判明し，その端末を調査することで，端末の感染範囲や他の社内サーバへのアクセスの有無等，事案の解明への突破口が開けるためである．

しかし，ブラックリストに持っているドメイン名や IP アドレスから悪性サイトと判断することはできるが，それ以外にも悪性サイトは存在する．そこで，機械学習を用いて既知の悪性サイトと類似の特性を持つサイトを悪性サイトとして判別できるよう識別モデルを作成することとした．

また，大量にあるプロキシログの中から悪性サイトへアクセスしたログを特定する作業は，ログ分析に精通している者でなければ時間のかかる作業となる．例えば，C&C サーバへの定期的な通信等，ウイルスの特性に応じた分析を行うといった経験や知識がなければ，事案の解明に時間がかかってしまうことになる．

そこで，機械学習を用いた判別手法を適用することで，詳細なログ分析を行う対象を絞り込むことで，分析を効率的に行えることが期待できる．

プロキシでのフィルタリングに機械学習を使う方法は，従来著者らも行ってきたが[3][4]，今回行うのはログの解析時であり，複数の悪性サイトを扱うので，次のような対応が必要となる．

(1) インシデント後におけるログ分析をする観点から，機械学習の識別モデル作成の際，実際は悪性サイトへアクセスしたログであるが，一般サイトへアクセスしたログと誤って判断する，False Negative を少なくすることが重要となる．そこで，悪性サイトのログが残っているプロキシログを作成した識別モデルに適用させたときに，見逃すことなく判別するかを確認しておく必要がある．

(2) 不正な C&C サーバとマルウェアを配布する最終の不

1 東京電機大学
Tokyo Denki University

正な DbD 攻撃を行うサイト（以下，DbD 攻撃サイトと呼ぶ）の 2 つの確認が必要となり，また，一般にデータ量が多い場合があり現実的時間で対応できるか確認しておくとともに高速化の検討をしておく必要がある．

2. 関連研究

関連する研究として，DbD 攻撃の特定を目的とするもの，C&C サーバの特定を目的とするもの，プロキシログの分析に関するものがある．

2.1 DbD 攻撃の特定手法

文献[1]では，ドメイン情報を用いた DbD 攻撃の特定手法を提案している．DNS 情報，WHOIS サービスの登録期間等，計 13 種類の特徴量で，機械学習を用いて DbD ドメインと一般ドメインを分類している．しかし，C&C サーバへの接続については言及していない．

文献[2]では，HTML と JavaScript，Web ページの URL の情報等から計 48 種類の特徴量を抽出し，機械学習を用いて Web ページをフィルタリングする手法が提案されている．ただし，取得する特徴量が多く，取得に要する時間が多くなる可能性がある．

2.2 C&C サーバの特定手法

文献[3][4]では，ドメイン情報を用いた C&C サーバの特定手法を提案している．文献[3]では，検索エンジン，登録期間，メールアドレス情報の計 3 種類の特徴量で，機械学習を用い，文献[4]では，DNS 情報，WHOIS サービスの登録期間等，計 9 種類の特徴量を用いて，C&C サーバと一般ドメインを分類している．ただし，C&C サーバへの接続のみに焦点を当てており，DbD 攻撃サイトへの検知については言及していない．

文献[5]では，DNS クエリの挙動を観測し，不正な通信と一般の通信の識別手法を提案しているが，ISP 規模での DNS クエリの観測が必要となる．

2.3 プロキシログの分析手法

文献[6]では，プロキシログの分析において，機械学習による RAT の検知手法について提案している．指定行数のログごとの受信サイズ，リクエストの間隔，User Agent の長さ等 8 種類の特徴量を抽出し，悪性のログを識別する手法が提案されているが，事前にウイルスの被害にあった組織の悪性のログを入手する必要があることと，悪性のログとして 2 タイプの RAT の通信を学習させていることで，違うタイプの RAT の場合にも有効であるかについては言及していない．

文献[7]では，マルウェアの感染が無いと想定する組織のプロキシログをホワイトリストやマルウェアに感染したログの特徴を用いてログを削減する，いわゆるグレーリストの作成手法を提案している．この手法は，最終的に作成されたグレーリストの行数が多くなる場合，その後の分析に時間がかかる可能性がある．

3. 悪性サイトの特定手法と評価

悪性サイトの基本的特定手法の流れについて，図 1 に示す．

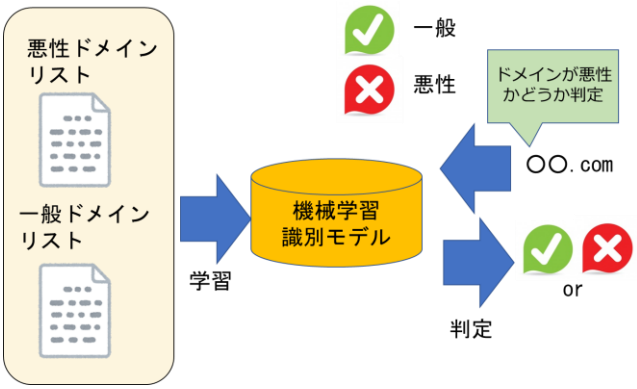

図 1　悪性サイト特定手法の流れ

まず，悪性ドメインが掲載されているブラックリストと一般ドメインリストを用いて学習し，機械学習識別モデルを作成する．次に，判別させるためのドメインを機械学習識別モデルに読み込み，悪性サイトのドメインであるかを判定する．

インシデントの調査において，DbD 攻撃サイトへのログ，C&C サーバへのログのような，悪性サイトへのログを特定できる必要性から，本論文では，DbD 攻撃サイトと C&C サーバを特定する手法を提案し，評価結果について報告する．識別モデルは SVM とニューラルネットワークで作成し，検知率をそれぞれ求めた．

3.1 SVM とニューラルネットワーク

機械学習用に用いる識別モデルは SVM とニューラルネットワークの 2 種類作成した．

SVM は，教師あり学習を用いる識別モデルの一つであり，分類や回帰問題で広く利用されている学習アルゴリズムである．比較的優れた識別能力を持ちかつ高速であることが知られているため，SVM を選択した[8]．

ニューラルネットワークは，人間の脳の仕組みから着想を得たもので，脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデルである[9]．こちらも比較的優れた識別能力を持っているのが知られているため，ニューラルネットワークを選択した．

3.2 特徴量とクラス分け

SVM とニューラルネットワークで用いる特徴量の一覧を表 1 に示す.

表 1　特徴量一覧

No.	特徴量
1	MX レコード
2	Refresh
3	Retry
4	Expire
5	Minimum
6	TXT レコード
7	NS レコード
8	A レコード
9	国名
10	ドメインの長さ
11	検索エンジン
12	登録期間
13	ReverseIP

No.1 から No.8 は対象 DNS サーバに対して問い合わせることで取得した．No.9，No.12 は WHOIS サービスを用いて取得し，No.10 は文字列の長さをカウントした．No.11 は Microsoft が提供する検索エンジンである bing を用いて取得し，No.13 は Domaintools が提供している IP アドレスを逆引きするサービス[10]を用いて取得した.

次に各項目について説明する．No.1 と No.6 から No.8 は DNS サーバから取得したドメインの情報であり，それぞれ個数を調査した．No.2 から No.5 は DNS サーバから取得したドメインの設定情報が記載されている SOA レコードであり，それぞれの項目の設定値を調査した．No.9，No.12 は WHOIS サービスを用いて，サーバを管理している国情報，ドメインがレジストリ組織に登録されている期間について調査した． No.10 は，ドメイン名の文字数を調査した．No.11 は，bing の検索欄に「site:」コマンドを用いてドメインを検索した際の検索結果の件数を調査した．No.13 は，Domaintools のサイト内の Reverse IP Lookup のサービスを用いて IP アドレスに対応するドメインの数を調査した．ここで用いた IP アドレスは，事前に DNS サーバに問い合わせて求めた.

上記で取得したそれぞれの特徴量をそのまま用いると，検索件数のような，絶対値が極端に大きな数値結果が出た場合に，絶対値が小さな値が無視されてしまう，情報落ちが発生するおそれがある．この現象を防ぐため，表 2 のようにスケーリングを行いクラス分けを行った．スケーリングにより，大きい値をとる特徴量による情報落ちを防ぐことができる．nil と記載されているところは，値が取得できなかったことを意味する.

表 2　項目のクラス分け

MXレコード(個)	値
0	1
1	2
2～	3

NSレコード(個)	値
0	1
1,2	2
3～	3

Refresh(秒)	値
nil	1
1～3600	2
3601～21600	3
21601～	4

Aレコード(個)	値
0	1
1	2
2～	3

国名	値
アメリカ	1
日本	2
その他	3

Retry(秒)	値
nil	1
1～3600	2
3601～	3

Expire(秒)	値
nil	1
1～1280000	2
1280001～	3

ドメインの長さ	値
0～10	3
11～20	2
21～	1

Minimum(秒)	値
nil	1
1～14400	2
14401～	3

検索エンジン(件)	値
0～5	1
5～1000	2
1001～	3

TXTレコード(個)	値
0	1
1	2
2～	3

登録期間(日)	値
0～500	1
501～3000	2
3001～	3

ReverseIP(個)	値
0	1
1	2
2	3
3～	4

3.3 実験用ドメイン

機械学習で学習させるデータリストとして，外部に公開されている悪性サイトと一般サイトのドメイン（以下，悪性ドメイン，一般ドメイン）のリストを用いた．悪性ドメインのデータリストとして，DbD 攻撃サイト用と C&C サーバ用の 2 種類用意した，DbD 攻撃サイトは，Squidblacklist.org[11]が提供するドメインリストを使用した．C＆C サーバは，オープンソースで脅威情報を提供する threatfees.io[12]に掲載されているドメインリストを使用した．一般ドメインのリストは，SEOBOOK が提供している Alexa ランキング[13]のドメインリストを使用した. DbD 攻撃サイトドメイン（以下，DbD ドメイン），C&C サーバドメイン（以下 C&C ドメイン），一般ドメインの件数を以下の表 3 に示す.

表 3　実験用ドメインの件数

ドメイン	件数
DbD ドメイン	540 件
C&C ドメイン	540 件
一般大規模ドメイン	180 件
一般中規模ドメイン	180 件
一般小規模ドメイン	180 件

一般大規模ドメインは，Alexa ランキング 1 位から 1000 位までのランダムなドメインを使用した．一般中規模ドメインは，Alexa ランキング 5000 位から 6000 位までのランダムなドメインを使用した．一般小規模ドメインは，Alexa ランキング 1 万位から 1 万 1000 位までのランダムなドメインを使用した．これらのドメインを識別モデルの学習用に使用した．なお，DbD ドメイン用の識別モデルを作成する際は，上記 DbD ドメイン 540 件と一般大・中・小規模ドメイン混合の 540 件を組み合わせた 1080 件を用い，C&C ドメイン用の識別モデルを作成する際は，上記 C&C ドメイン 540 件と一般大・中・小規模ドメイン混合の 540 件を組み合わせた 1080 件を用いて実験を行った．

3.4 実験概要

実験では，機械学習のアルゴリズムとして，SVM とニューラルネットワークの 2 種類を用いて比較実験を行った．また，DbD ドメイン，C&C ドメインを識別するために，それぞれ識別モデルを作成し，性能評価を行った．性能評価で用いる指標として表 4 の判定方式を用いる．一般ドメインが一般ドメインと判定された場合に TruePositive（以下，TP），悪性ドメインと判定された場合に FalsePositive（以下，FP）とする．同様に悪性ドメインが一般ドメインと判定された場合に FalseNegative（以下，FN），悪性ドメインと判定された場合に TrueNegative（以下，TN）とする．

表 4　検知判定方式

	一般ドメインと判定	悪性ドメインと判定
一般ドメイン	TP	FP
悪性ドメイン	FN	TN

上記の表を評価指標として，一般ドメインを悪性ドメインと誤って識別した精度を FPR（False　Positive Rate）として算出し，悪性ドメインを一般ドメインと誤って識別した精度を FNR（False　Negative　Rate）として算出した．

$$\mathrm{FPR}=\frac{\text{FP の数}}{\text{一般ドメインの数}}$$

$$\mathrm{FNR}=\frac{\text{FN の数}}{\text{悪性ドメインの数}}$$

また，検知率は，次の式から求めた．

$$\text{検知率}=\frac{\text{TP の数}+\text{TN の数}}{\text{一般ドメインの数}+\text{悪性ドメインの数}}$$

これらの 3 つの式を用いて評価を行った．

なお，識別モデルの評価の妥当性のため，交差検証法[14]を用いて評価を行った．交差検証法を用いることで，実験用データが少ない場合でも検知率の精度の誤差を抑えることができる．評価用データを 10 分割し，そのうち一つをテスト用データ，残りを学習用データとし，10 回評価を行った平均値を評価値として算出した．

実験を行う方法として，SVM では python のライブラリである scikit-learn[15]を用い，ニューラルネットワークでは，機械学習用ソフトウェアである Weka[16]を用いて行った．

3.5 実験結果

識別モデルの評価結果を表 5，6 に示す．

表 5　DbD と一般ドメインの識別モデル評価

	検知率	FPR	FNR
SVM	98.05%	1.1%	2.2%
ニューラルネットワーク	97.12%	2.6%	3.1%

表 6　C&C と一般ドメインの識別モデル評価

	検知率	FPR	FNR
SVM	98.70%	0.9%	1.4%
ニューラルネットワーク	97.87%	2.2%	2.0%

評価結果より，DbD ドメイン，C&C ドメインともにどちらの識別モデルでも比較的高い検知率であった．SVM とニューラルネットワークの評価結果の比較では，どの項目でも SVM の方が数値が良かった．これらの結果より，提案した方式で作成した識別モデルは有用性が期待できる．

3.6 識別モデルの生成に影響の大きかった特徴量

特徴量の中で，検知率等の精度に最も影響の大きかった特徴量は「検索エンジン」であった．DbD ドメインと一般ドメインの検索エンジンを比較したものを図 2 に示す．

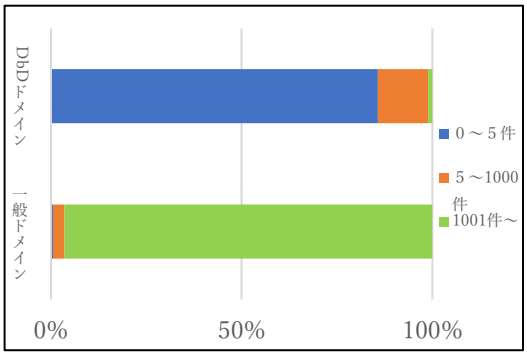

図 2　検索エンジンの比較

このように，検索エンジンの「site:」コマンドによるドメイン検索件数は，DbD ドメインの方は 0～5 件の間に多数集まっているが，一般ドメインの方はほとんどない．また，1001 件以上になると DbD ドメインの方はほとんどないが，一般ドメインの方は多数集まっている．C&C ドメインと一般ドメインの比較においても，同じような傾向を確認した．この特徴量を学習させた経緯として，悪性ドメインはドメイン検索を行っても検索結果はほとんど出てこないと予想をしたためであったが，実際に実験したところ，そのような結果となった．

4.　プロキシログでの適用と評価

3 章では，悪性ドメインと一般ドメインを識別する手法について述べた．識別の際に SVM とニューラルネットワークの識別モデルを作成し，それぞれ検知率等を実験により求めた結果，SVM の方が検知率等の性能が良かったことが判明した．

本章では，3 章の提案手法による判別方法の有効性を実際のログに対して適用することで評価を行った．評価項目では，プロキシログに入っている悪性サイトへのログを判別できるか，一般ドメインを誤って悪性ドメインと誤判別する数，処理時間についてそれぞれ行った．

4.1　プロキシログ解析の流れ

プロキシログに対する解析の流れを図 3 に示す．

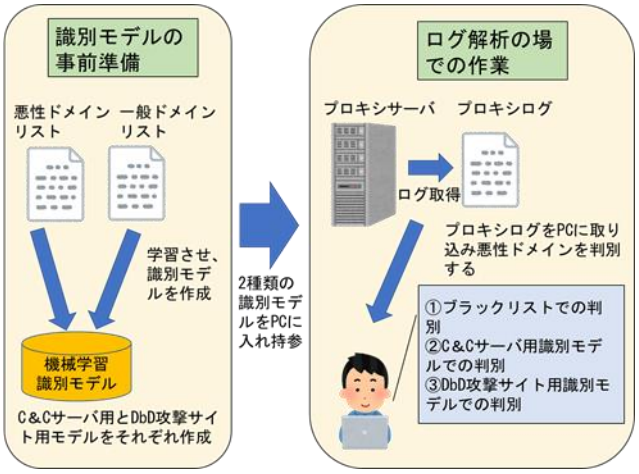

図 3　プロキシログに対する解析の流れ

まず，3 章で行ったように，悪性ドメインリストと一般ドメインリストを事前に準備し，それぞれのリストを用いて，識別モデルを作成する．その際に，悪性ドメインリストとして，C&C サーバ用と DbD 攻撃サイト用のリストをそれぞれ用いて，3 章で作成したように，C&C サーバ用と DbD 攻撃サイト用の識別モデルを作成しておく．作成した 2 種類の識別モデルをパソコンに入れて解析の準備を行う．

ログ解析の場では，プロキシサーバのログを解析するため，プロキシサーバのログを抽出する．抽出したプロキシログをパソコンに取り込み，ログの中からドメイン部分を抽出し，悪性ドメインを判別する．

判別の際に，最初に行うことは，事前に入手したブラックリストでの判別である（図 3-①）．ブラックリストには，すでに調査する組織が認知している C&C サーバと DbD 攻撃サイトのドメインリストが掲載されており，プロキシログに該当のドメインがあればパターンマッチングにより判別する．もし，ブラックリストに掲載されていない悪性ドメインがプロキシログにある場合は，この方法では判別することができないので，事前に準備した 2 種類の識別モデルにより，C&C サーバと DbD 攻撃サイトを判別させる（図 3-②，③）．C&C サーバと DbD 攻撃サイト判別の詳細方法については次節の 4.2 節で説明する．

本章では，ブラックリストに掲載されていない悪性ドメインを発見することを想定して，C&C サーバ用と DbD 攻撃サイト用の識別モデルの 2 種類についてそれぞれ評価を行う．

4.2　悪性ドメイン判別の流れ

4.1 節のログ解析の場での C&C サーバと DbD 攻撃サイトのドメイン判別の流れを図 4 に示す．

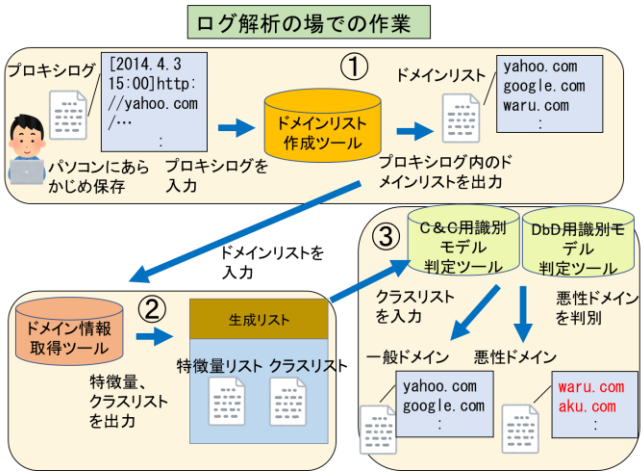

図 4　悪性ドメイン判別の流れ

まず，パソコンにあらかじめ保存されたプロキシログをドメインリスト作成ツールに入力し，プロキシログ内のドメインリストを出力する（図 4 - ①）．

次にドメインリストをドメイン情報取得ツールに入力し，DNS 情報や検索エンジン，WHOIS サイト等に問い合わせを行い，特徴量リストを作成し，特徴量リストからクラスリストを作成する（図 4 - ②）．

作成したクラスリストを C&C サーバ用と DbD 攻撃サイト用の識別モデル判定ツールに入力しそれぞれ悪性ドメインを判別する（図 4 - ③）．

ドメインリスト作成ツールでは，プロキシログの中からアクセスに失敗してコンテンツ取得に失敗しており，攻撃が成立する可能性がないログを排除し，URL 部分記載のドメインの箇所を抜き出し，ドメインの重複を排除したリストを作成している．この処理により，次のドメイン情報取得ツールにおいて，重複したドメイン情報の取得のような無駄な処理を軽減することができる．

これらを Python[21]を用いて実装した．

なお，2 つのモデルにおいて判別した結果，両方とも悪性ドメインと判別されるドメインが出てくる可能性があるが，元のプロキシログに対して該当のドメインで検索し，発生時間や発生回数などで，C&C か DbD ドメインかを判断することができる．DbD ドメインであれば，感染源なので，何度もアクセスはしないと考えられるが，C&C ドメインであれば，複数回発生している可能性が高いからである．

4.3　使用する識別モデル

識別モデルについては，C&C サーバ用と DbD 攻撃サイト用と 2 種類のモデルを作成した．学習用のデータは 3.3 節で用いた実験用のデータと同じものをそれぞれのモデル作成用に用いた．取得する特徴量も 3.1 節で提案した 13 種類をそれぞれのモデルで用いた．また，識別モデルは，3.5 節で検知率が良かった SVM を用いて作成をした．

4.4　プロキシログの準備

プロキシログは自端末で外部の Web サーバにアクセスしたログを収集して取得した．その際に，自端末からの Web アクセスを，自端末に実装したプロキシサーバ経由でアクセスすることで，プロキシログを作成した．プロキシサーバは，オープンソースソフトウェアとして公開されている Squid[17]を用いて実装した．ログの取得期間は一か月間とし，その間に取得したログを一般サイトのログとみなした．ただし，プロキシログに含まれている悪性サイトのログの判別について評価をするため，外部サイトが運営し，公開している C&C と DbD のブラックリストのドメインを，作成したプロキシログに追加することとした．

DbD ドメインは nao_sec[18]によって収集された DbD 攻撃をパケットキャプチャした pcap ファイルの中にあるドメインを使用した．pcap ファイルの内容は，2018 年下旬に確認された，Fallout エクスプロイット・キットのログである[19]．

この pcap ファイルをプロキシログ用に変換したものを疑似ログとして，先に作成したプロキシログに追加した．追加の方法として，pcap ファイルのリクエストとレスポンスの対応付けを行い，そのペアから時刻，URL の情報を抽出し，作成したプロキシログに追加した．なお，9 か所の DbD 攻撃サイトのログをプロキシログに追加した．DbD 攻撃サイト用識別モデルを評価するためのこのログを DbD 攻撃サイト用プロキシログとする．

C&C ドメインは，Feodo Tracker[20]でよって収集された C&C ドメインのリストを使用した．このドメインのリストをプロキシログ用に変換したものを疑似ログとして，作成したプロキシログに追加した．追加の方法として，掲載されていたドメイン名を含むログをプロキシログとして追加した．なお，9 か所の C&C サーバのログをプロキシログに追加した．C&C サーバ用識別モデルを評価するためのこのログを C&C サーバ用プロキシログとする．

上記のように，DbD 攻撃サイト用と C&C サーバ用の識別モデルを評価するためのプロキシログを 2 種類用意した．

4.5　性能評価

悪性ドメインの判別結果，処理速度について，次節の各項目ごとに評価結果を示す．

4.5.1 悪性ドメインの判別結果

4.3 節で作成した 2 種類の識別モデルを用いて，プロキシログの中にある悪性ドメインを判別できるか実験を行った．すなわち，DbD 攻撃サイト用の識別モデルを用いて DbD 攻撃サイト用のプロキシログの中から DbD ドメインを判別できるか，C&C サーバ用の識別モデルを用いて，C&C サーバ用のプロキシログの中から C&C ドメインを判

別できるかをそれぞれ実験した．

それぞれのプロキシログの中に悪性ドメインの数を 9 個入れた状態で実験を行ったが，DbD 攻撃サイトと C&C サーバともに，すべての悪性ドメインを判別することができた．

また，プロキシログの行数はそれぞれ 5 万 1283 行，含まれていたドメイン数は 932 個であり，そのなかから悪性ドメインと判別した数，一般を悪性ドメインと誤判別した数をそれぞれ求めた．例えば，DbD 攻撃サイト用識別モデルにおいて，プロキシログの中から悪性と判別した DbD ドメインは 45 個となり，その中に本物の悪性ドメインは 9 個という状態であった．よって，36 個は一般ドメインを悪性ドメインと誤判別したことになる．これらの関係式を以下に示す．

悪性ドメインと判別した数 ＝ 実際の悪性ドメイン数 ＋ 一般を悪性ドメインと誤判別した数

DbD 攻撃サイト用識別モデル，C&C サーバ用識別モデルごとの結果を表 7 に示す．表中の%表記は，プロキシログに含まれているドメイン数（932 個）に対する割合を記載した．例として，悪性ドメインと判別した数が 45 個の場合，

$$\frac{45}{932} \times 100 = 4.8(\%)$$

となる．

表 7　性能評価結果

	実際の悪性ドメイン	一般を悪性ドメインと誤判別	悪性ドメインと判別したものの合計
DbD 用識別モデル	9 個(すべて発見)	36 個(3.8%)	45 個(4.8%)
C&C 用識別モデル	9 個(すべて発見)	38 個(4.0%)	47 個(5.0%)

このように，大量にあるログを検査するのと比較して，調査すべき箇所を減らすことができているので，効率的に作業を行えることが期待できる．

4.5.2 処理速度

続いて，処理速度について述べる．

処理速度は以下の表 8 の環境で実施した．

表 8　実験環境

CPU	Corei7-8550U 1.80GHz
メモリ	8GB
回線速度上り	59.17Mbps
回線速度下り	51.75Mbps

なお，回線速度は BNR スピードテストにて実施した．

4.2 節の図 4 の悪性ドメイン判別の流れにおいて，保存されたプロキシログから悪性ドメインの判別までのパソコンの処理時間を計測したところ，およそ 11 分くらいであった（なお，ドメイン数は 932 であるので 1 ドメインあたりは 0.71 秒）．内訳として，ドメインリストの作成処理（図 4－①）では 2 分程度，ドメイン情報取得処理（図 4－②）では，9 分程度，悪性ドメインの判別処理（図 4－③）では，ほぼ時間がかからなかった．

ドメイン情報取得処理で時間がかかったのは，DNS 情報の取得など，外部のサイトへアクセスする処理が発生するためと考えられる．

4.5.3 学習データを変更した識別モデルの判別結果

学習用データを変更した DbD 攻撃サイト用と C&C サーバ用の識別モデルを作成し，それぞれに適用した際，悪性ドメインを判別できるか，一般ドメインを悪性ドメインと誤判別する数はどのくらいかを実験により検証した．

識別モデルはそれぞれ 4 種類ずつ作成し，学習用データは 3.3 節の一般ドメインの規模を変更したものをそれぞれ学習させた．取得する特徴量の数等，他の箇所は変更していない．学習させた一般ドメインの一覧を以下に示す．

(1) DbD 攻撃サイト識別モデル

- 大規模学習モデル：
  一般大規模ドメイン 540 件＋DbD ドメイン 540 件
- 中規模学習モデル：
  一般中規模ドメイン 540 件＋DbD ドメイン 540 件
- 小規模学習モデル：
  一般小規模ドメイン 540 件＋DbD ドメイン 540 件
- 超小規模学習モデル：
  一般超小規模（Alexa ランキング 30 万位周辺）ドメイン 540 件＋DbD ドメイン 540 件

(2) C&C サーバ識別モデル

- 大規模学習モデル：
  一般大規模ドメイン 540 件＋C&C ドメイン 540 件
- 中規模学習モデル：
  一般中規模ドメイン 540 件＋C&C ドメイン 540 件
- 小規模学習モデル：
  一般小規模ドメイン 540 件＋C&C ドメイン 540 件
- 超小規模学習モデル：
  一般超小規模（Alexa ランキング 30 万位周辺）ドメイン 540 件＋C&C ドメイン 540 件

学習用データを変更した識別モデル判別結果を表 9，10 に示す．また，4.5.1 節の結果も併せて示す．表中の%表記

は，プロキシログに含まれているドメイン数（932 個）に対する割合を記載した．

表 9　DbD 攻撃サイト用識別モデル判別結果

学習用一般ドメイン規模	実際の悪性ドメイン	一般を悪性ドメインと誤判別	悪性ドメインと判別したものの合計
大規模学習モデル	9 個(すべて発見)	66 個(7.0%)	75 個(8.0%)
中規模学習モデル	9 個(すべて発見)	66 個(7.0%)	75 個(8.0%)
小規模学習モデル	9 個(すべて発見)	47 個(5.0%)	56 個(6.0%)
超小規模学習モデル	9 個(すべて発見)	20 個(2.1%)	29 個(3.1%)
4.5.1 節作成識別モデル	9 個(すべて発見)	36 個(3.8%)	45 個(4.8%)

表 10　C&C サーバ用識別モデル判別結果

学習用一般ドメイン規模	実際の悪性ドメイン	一般を悪性ドメインと誤判別	悪性ドメインと判別したものの合計
大規模学習モデル	9 個（すべて発見)	57 個(6.1%)	66 個(7.0%)
中規模学習モデル	9 個(すべて発見)	32 個(3.4%)	41 個(4.3%)
小規模学習モデル	9 個(すべて発見)	38 個(4.0%)	47 個(5.0%)
超小規模学習モデル	9 個(すべて発見)	18 個(1.9%)	27 個(2.8%)
4.5.1 節作成識別モデル	9 個(すべて発見)	38 個(4.0%)	47 個(5.0%)

この結果より，学習用データとして，一般ドメインの規模を変更した実験では，DbD 攻撃サイト用，C&C サーバ用識別モデルともに，すべての規模でプロキシログに含まれる悪性ドメイン 9 個をすべて判別できていることが確認できた．

また，DbD 攻撃サーバ用識別モデルでは，4.5.1 節作成識別モデルと超小規模学習モデルにおいて，一般ドメインを悪性ドメインと誤判別した数が比較的少ないことが判明した．また，C&C サーバ用識別モデルでも，超小規模学習モデルを用いた場合が，一般ドメインを悪性ドメインと誤判別した数が比較的少ないことが判明した．

プロキシログの中のデータは超小規模学習モデルのデータに似ているからだと考えられる．例えば，プロキシログの中には広告サイトのような，検索エンジンの検索結果が少ないものがあり，超小規模学習モデルでも，検索結果が少ないようなドメインが多い．超小規模学習モデルではこのようなドメインにも学習過程で対応したため，悪性ドメインと誤判別した数が少なかったと想定される．

4.5.4 取得する特徴量を減らした処理速度

4.5.2 節での処理速度の結果について，プロキシログから悪性ドメインを判別する処理過程において，ドメイン情報の取得に時間がかかっていた．さらに短縮できるかを検討するため，識別モデル作成の際に必要な特徴量の数を減らし，悪性ドメインの判別数と処理時間を計測した．

今回はドメイン情報の取得時に時間が比較的かからず，かつ悪性ドメインの判別に影響力がありそうな特徴量を 3 つ選んだ．取得する特徴量は，3.1 節で記載された項目名で A レコード，ドメインの長さ，検索エンジンの 3 つとした．

取得する特徴量を 3 つにした識別モデルを DbD 攻撃サイト用と C&C サーバ用の 2 種類準備し，それぞれの判別数と処理時間を求めた．他の条件は 4.5.2 節と同様の条件で行った．

ドメイン情報取得処理で，2 分程度であったので，保存されたプロキシログから悪性ドメインの判別までのパソコンの処理時間を計測したところ，およそ 4 分くらいに短縮できた（なお，ドメイン数は 932 であるので 1 ドメインあたりは 0.26 秒）．

判別結果は以下の表 11 に示す．表中の%表記は，プロキシログに含まれているドメイン数（932 個）に対する割合を記載した．

表 11　特徴量を 3 つにした判別結果

	実際の悪性ドメイン	一般を悪性ドメインと誤判別	悪性ドメインと判別したものの合計
DbD 用識別モデル	9 個(すべて発見)	69 個(7.4%)	78 個(8.3%)
C＆C 用識別モデル	9 個(すべて発見)	47 個(5.0%)	56 個(6.0%)

判別結果において，DbD 攻撃サイト用，C&C サーバ用識別モデルともにすべての悪性ドメインを判別することができた．また，一般を悪性ドメインと誤判別した数は双方ともに，4.5.2 節よりも増加した．このように，処理時間の減少と一般を悪性ドメインと誤判別する数はトレードオフの関係であるが，調査するべき箇所を減らすことができたことも確認した．

5. おわりに

本論文では，インシデント後におけるログの調査という観点で，DbD 攻撃サイトや C&C サーバを機械学習を用いて発見する試みを行った．

識別モデルの作成は SVM とニューラルネットワークを用いて行い，それぞれの比較実験を行い，検知率等の性能評価を行った．その結果，いずれも高い検知率を示しているが，SVM で作成した識別モデルの方が優位性があった．

作成した識別モデルを用いて，実際のプロキシログに適用した際にも有用性があるかどうかを確かめる実験を行ったところ，すべての関係する悪性ドメインを判別することができた．また，DbD 攻撃サイトの場合も C&C サーバ用識別モデルでも，超小規模学習モデルを用いた場合が，一般ドメインを悪性ドメインと誤判別した数が比較的少ないことが判明した．これは，従来指摘されていなかったことである．

さらに，一般ドメインを悪性ドメインと判別する誤判別があったものの，元の大量のログを分析することと比較したら，調査するべき箇所を減らすことができたことも確認した（すべての学習モデルで誤検知の割合が 7.4%以下）．

悪性ドメイン判別の処理時間においても，5 万行以上のプロキシログに対して 11 分，取得する特徴量を 3 つに減らした場合は 4 分で処理できることを確認した．特徴量を減らした場合，一ドメインあたりの処理時間を 11 分の 4 に減らし，誤検知の割合が上昇したものの（DbD 用識別モデルでは 3.8%から 7.4%，C&C 用識別モデルでは 4.0%から 5.0%となる），調査するべき箇所を少ない状態に抑えられた．

組織のプロキシログでは，ログの量が膨大となるので，一ドメインあたりの処理時間の低減が望ましい．また，一般ドメインを悪性ドメインと誤判別する個数も増え，調査すべき箇所が増えることが予想される．よって，今後は処理時間の低減等の検討を続けていく予定である．

参考文献

[1] 木村 匡, 佐々木良一: ドメイン情報の分析による Drive by Download 攻撃の対策の提案, マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, pp.1705-1710 (2016) .

[2] Canali, D., Cova, M., Vigna, G., and Kruegel, C.: Prophiler: A fast filter for the large-scale detection of malicious webpages, In Proceedings of the International World Wide Web Conference(2011) .

[3] 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 サーバの検知手法の提案と評価, 情報処理学会論文誌, Vol.58, no.9, pp.1410-1418 (2017).

[4] 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フィルタシステムの提案と評価, 情報処理学会論文誌, Vol.58, no.1, pp.249-257 (2017).

[5] Rahbarinia, B., Perdisci, R., Antonakakis, M.: Segugio: Efficient Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks, IEEE/IFIP International Conference on Dependable Systems and Networks (2015) .

[6] 三村守, 大坪雄平, 田中英彦: プロキシのログからの機械学習による RAT の検知方式, コンピュータセキュリティシンポジウム 2015 論文集, pp.528-535 (2015).

[7] 田中功一, 堀川博史, 峰野博史, 西垣正勝: ログ解析によるマルウェア侵入検知手法の提案, マルチメディア, 散協調とモバイルシンポジウム 2014 論文集, pp.522-529 (2014).

[8] John, P.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Technical Report MSR-TR-98-14, p1-21 (1998) .

[9] Multilayer Perceptron, 入手先 <http://deeplearning.net/tutorial/mlp.html#mlp > (参照: 2019-04-16).

[10] Reverse IP Lookup, 入手先<http://reverseip.domaintools.com/> (参照: 2019-04-17).

[11] quidblacklist: SQUIDBLACKLIST.ORG, 入手先< https://www.squidblacklist.org/> (参照: 2019-04-17).

[12] threatfeeds io: Free and open-source threat intelligence feeds, 入手先<https://threatfeeds.io/> (参照: 2019-04-17).

[13] SEOBOOK : Download Alexa Top 1,000,000 Websites for Free, 入手先<http://www.seobook.com/download-alexa-top-1-000-000-websites-free> (参照:2019-04-17).

[14] Kohavi, R. : A Study of CrossValidation and Bootstrap for Accuracy Estimation and Model Selection. The International Joint Conference on Articial Intelligence, pp.1137-1143 (1995).

[15] scikit-learn:machine-learning in Python, 入手先<https://scikit-learn.org/stable/index.html> (参照: 2019-04-17).

[16] Weka: Machine Learning Software in Java, 入手先 <https://www.cs.waikato.ac.nz/~ml/weka/index.html> (参照: 2019-04-17).

[17] squid: squid-cache.org, 入手先<http://www.squid-cache.org> (参照: 2019-04-17).

[18] nao_sec: nao_sec Cyber Security Research Team, 入手先 <https://nao-sec.org/> (参照: 2019-04-16).

[19] Fallout Exploit Kit: Hello "Fallout Exploit Kit", 入手先 <https://nao-sec.org/2018/09/hello-fallout-exploit-kit.html> (参照: 2019-04-18).

[20] Feodo Tracker, https://feodotracker.abuse.ch/（最終閲覧日: 2019-5-1）

[21] python, 入手先<https://www.python.org/> (参照: 2019-04-18).