IP アドレスクラスにおけるネットワークアドレスの特徴を用いた未知の不正 Web サイト判別手法

金澤しほり[†1] 中村嘉隆[†2] 稲村浩[†2] 高橋修[†2]

概要：近年，Drive-by download攻撃やフィッシングなどWebサイトを介したサイバー攻撃が急増しており，ユーザの個人情報等が不正に取得され，経済的被害を受ける事件が増加している．このような被害を防ぐには，ユーザが不正Webサイトを閲覧する前に，アクセスを遮断するなどの対策が必要である．これまでに，Webレピュテーションや，IPS（侵入防止システム）といった技術が開発され，対策手法として用いられている．本稿では，DNSから得られるドメイン情報・IPアドレス情報や，マルウェアに感染したクライアントに関する情報を利用することで，アクセス先が未知のWebサイトである場合にも対応可能なネットワークアドレスを用いた不正Webサイト判別手法の提案を行い，評価した． 悪質な活動に多く利用されているIPアドレスクラスAの結果が高精度を示したことから，IPアドレスのネットワークアドレス部を用いた判別は，有効であることが確認できた．

A distinction method of unknown malicious web sites using IP address features of each network address class

SHIHORI KANAZAWA[†1] YOSHITAKA NAKAMURA[†2]
HIROSHI INAMURA[†2] OSAMU TAKAHASHI[†2]

1. 背景

近年，Web サイトを利用した攻撃が急増している．攻撃例として，ユーザが Web サイトを閲覧した際に，ウイルスやマルウェアなどの不正プログラムをパソコンにダウンロードさせる Drive-by download 攻撃や，ユーザを，金融機関を装った偽のサイトへ誘導するフィッシング詐欺が挙げられる．これらの攻撃法により，閲覧者のパソコンでマルウェアが活動し，保管されたデータやプログラムが破壊されることや，暗証番号やクレジットカード番号などの個人情報を不正に取得され，経済的な被害を受ける事件が増加している．2011 年から 2015 年までの警察庁広報資料「インターネットバンキングに係る不正送金事犯発生状況」によると，個人情報が不正に取得されて被害に遭った 2012 年までの発生件数が 50 件程であったのに対し，2015 年には 1400 件近くまで急増しており，現在も増加傾向にある[1]．このような被害を防ぐために，ユーザが危険な不正 Web サイトを閲覧して被害に遭う前に，Web アクセス時に注意を促す対策を Web ブラウザ側で講じている．例えば，Web アクセスの際に，アクセス先が不正 Web サイトである場合，ユーザに警告を促し，ユーザ自身でアクセスを止める方法が Web ブラウザ側で実施されている．しかし，警告を促すことができるのは既知の不正 Web サイトに対してであり，未知の不正 Web サイトに関しては対策できない．そのため，今後は未知の不正 Web サイトを含めた対策が重要となる．

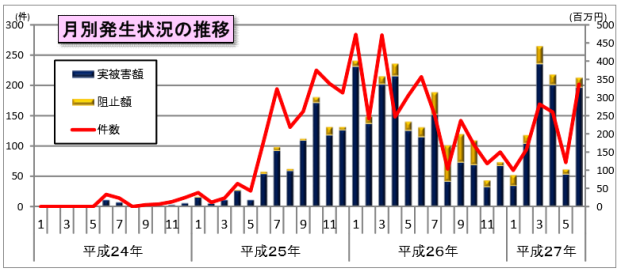


図 1 インターネットバンキングに係る不正送金事犯の発生件数

これまでに，不正 Web サイトへのアクセスを遮断するため，Web レピュテーションなどのソフトウェアや，IPS（侵入防止システム）といった技術が開発され，対策手法として用いられている．

1.1 Web レピュテーション

Web レピュテーション技術[2]は，不正 Web サイトブロック機能を持つソフトウェアである．ユーザが Web サイトにアクセスするなどの通信が発生する際に，接続先のドメインや Web サイトが不正な場合にはアクセス自体をブロックすることによって不正プログラムによる感染，フィッシングによる被害を防止している．しかし，不正 Web サイトとしてブロックされる条件は，ウイルス配信，フィッシ

†1 公立はこだて未来大学大学院 システム情報科学研究科
†2 公立はこだて未来大学 システム情報科学部

ング詐欺など，不正行為を行ったことが確認された Web サイトのみである．

1.2 IPS(侵入防止システム)

Intrusion Prevention System(IPS) [3]は，ファイアウォールやアンチウイルスだけでは防御が困難とされていた DoS 攻撃やボットなど巧妙かつ高度なセキュリティの脅威に対応しており，通信パケットの内容や振る舞いを検査し，不正な通信を検出した後，遮断を行う．このとき，Web サイトへアクセスが行なわれた際に，通信パケットに含まれる不正な通信を検出して，遮断する仕組みになっている．しかし，IPS は，不正 Web サイトに含まれる既知の不審パケットのみ検出している．

1.3 既存技術の問題点

前述した 2 つの既存技術の共通の利点として，ブラックリストや不正 Web サイトに含まれる既知の不審パケットなどの既知の情報を用いた検出手法は，既知の不正 Web サイトの検出率が高い．しかし，欠点として，未知の不正 Web サイトに対応した検出ができず，また，未知の Web サイトに対して正規・不正 Web サイトの判別ができない．

このような問題点を解決するためには，未知の不正 Web サイトを含めた検出ができる検出条件を考慮し，また，検出された未知の Web サイトに対して正規 Web サイトか不正 Web サイトを判別する必要がある．

2. 関連研究

既存技術の問題点に対して，既知の不正 Web サイトの特徴を用いて未知の不正 Web サイトを検出する手法が提案されている[4,6]．Web アクセス時の通信パケットやドメイン名，IP アドレス，TTL 値(Time To Live)などから不正 Web サイトの特徴を抽出して，不正の疑惑がある Web サイトと照合することで，同様の特徴を持った未知の不正 Web サイトを検出している． Domain Name System(DNS)は，あらゆる Web サイトのアクセス元とアクセス先を把握しており，ドメイン名や IP アドレスといった各 Web サイトの情報を一元管理しているため，未知の不正 Web サイトに対しても有効な情報を保持していると考えられる．そこで，DNS から得られる情報(DNS 情報)に着目し問題点を解決する．

DNS 情報を用いて不正 Web サイトを発見する方法は，大きく 2 つに分類される．1 つは，未知の Web サイトに対して既知の情報と照合し，一致したものを不正 Web サイトとして検出する方法である．もう 1 つは，既知の情報と類似した特徴を持つ Web サイトを正規と不正のどちらかに判別する方法である．

2.1 検出手法

劉ら[4]は，DNSから得られる情報のうち，不正Webサイトに見られる3つの特徴を検出条件に設定して，検出を行っている．不正Webサイトのドメイン名は，英数字と数字が混在するものが多い傾向があるため，1つ目の検出条件として英数字が混在するドメイン名を利用している．不正Webサイトのドメイン名は，ボットに感染したパソコン群(ボットネット)を利用してフィッシングやウイルス配布などを行うFast-Flux[5]などの攻撃手法を用いて自動生成されることが多いため，人間にとって扱いにくい10文字以上の長い文字列で構成されるものが多い．このため，2つ目の検出条件として10文字以上で構成されるドメイン名であることを利用している．また，DNSキャッシュに設定されているパケットの有効期間を表す生存時間(TTL値)として設定された時間を超えた場合に，該当するDNSキャッシュのデータを保持しているネームサーバはDNSキャッシュを破棄することで，データベースの整合性を維持している．生存時間を短くすると，キャッシュが即座に無効になり，最新のデータを頻繁に問い合わせることになるが，不正Webサイトでは，TTL値を小さく設定することで，ドメイン名を捕捉されにくくしている傾向があるため，TTL値が300秒以下のドメインであるかどうかを3つ目の検出条件としている．これら3つの特徴のいずれかに該当する不正Webサイトを検出している．

田中ら[6]は，マルウェアが通信を行う際の特徴を利用して，DNS 通信の観測を通じた新たな不正 Web サイトの検出を行っている．通信を行うマルウェアに感染しているクライアントは複数の不正 Web サイトにアクセスを行う傾向にある．不正 Web サイトにアクセスを行ったクライアントは，他の不正 Web サイトにもアクセスを行っている可能性が高い．そのため，DNS 通信において既知の悪性ドメインにアクセスを行っていたクライアントから名前解決要求のあるドメインは，マルウェアとの関連が深いドメインであると考えられるため，新たな不正 Web サイトとして検出している．しかし，これらの検出手法は，ユーザが Web サイトにアクセスする際に，各特徴を利用した検出条件を満たす Web サイトのみを検出しているため，検出できる不正 Web サイトが限定される．

2.2 判別手法

千葉ら[7]は，IPアドレス，FQDN文字列，ドメイン名の登録日の3つの情報から特徴を抽出し，分類器によって悪性Webサイトを判別する手法を提案している．悪質な活動に利用されるIPアドレスが一部のネットワークに密集する傾向にあることから，IPアドレスの構造的な特徴を元に抽出している．FQDN文字列は，ホスト名とドメイン名を省略せず繋げて記述した文字列のことであり，このFQDN文字列が，ランダムに文字を組み合わせて構成されている場合

は悪性Webサイトである傾向にあることから，文字列の構成を元に特徴を抽出している．また，新しいドメイン登録日を持つWebサイトの方が，悪性度が高い傾向があることから，ドメインの鮮度を特徴として抽出している．しかし，3つの特徴のうち，IPアドレスを用いた判別が，FQDN文字列やドメイン名の登録日よりも有効であることは判明しているが，各IPアドレス全てに有効な判別ではない．実際にアクセスが行なわれたWebサイトから取得したIPアドレスからビットごとに特徴を抽出して，分類器に適用した実験が行われているが，実環境のWebサイトは，Classless Inter-Domain Routing(CIDR)を用いてIPアドレスの上位24ビットをネットワークアドレスとして利用するものが多く，上位8ビット~上位32ビットそれぞれ判別したときの精度を評価した結果でも，上位24ビットをネットワークアドレスとして用いた時の精度が良いことが判明している[8]．しかし，実環境ではCIDR の上位24ビットのネットワークアドレス以外のネットワークアドレスを用いたWebサイトも存在するため，これを考慮した判別に有効な手法が必要である．

3. 提案方式

3.1 アプローチ

未知の不正 Web サイトに対応するために，「不正 Web サイトの判定」を「検出」と「判別」の二段階で行う手法を提案する．まず「検出」は，各関連研究のドメイン名を用いた検出条件を組み合わせることで，既知の不正 Web サイトの集合であるブラックリストに存在しない不正 Web サイトに対しても検出範囲の拡張を可能にする．また，「判別」では，悪質な活動に利用される IP アドレスは一部のネットワークアドレスに密集する傾向にあることを利用して，各 IP アドレスのネットワークアドレス部のみを用いた判別を行い，全体の精度向上を図る．

3.2 システム構成

システム構成図を図 2 に示す．本研究ではクライアントが Web アクセスを行う際に通信する DNS サーバを用いる．DNS サーバの内部には不正 Web サイトに関するブラックリストと，未知の不正 Web サイトに対応する検出部と判別部が存在する．DNS サーバに対してクライアントが Web アクセスを行った際に，ブラックリストを用いて不正 Web サイトの疑惑のある Web サイトを検出部で求め，その情報を判別部で利用する．判別部は，分類器で疑惑 Web サイトを判別してクライアントに通知する．本研究ではアクセス先が不正 Web サイトであることを通知し，警告を促すことによって不正 Web サイトのアクセスを防止する．これらのシステムの処理を 3.3 節で説明する．

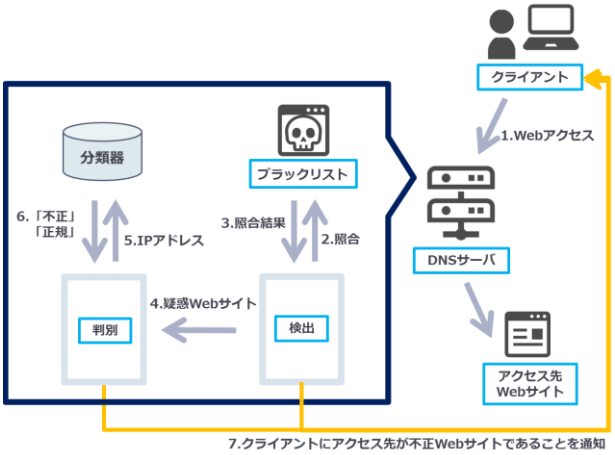


図 2 システム構成図

3.3 提案方式の概要

提案方式のシステムの処理を図 3 に示す．クライアントから DNS サーバに Web アクセスが行なわれた際に(1), DNS サーバから，ブラックリストを参照してアクセス先の Web サイトと照合する(2)．アクセス先がブラックリストに載っている既知の不正 Web サイトであれば，クライアントにアクセス先が不正 Web サイトであることを通知する(8). ブラックリストに載っている Web サイトを除外し(3), 不正 Web サイトのドメインの集合であるブラックリストと一致しなかった Web サイトからさらに，ドメイン名に基づいた検出条件と，マルウェアに感染された複数のクライアントからアクセスされている Web サイトのドメイン名を調べ(4), 条件を満たす Web サイトは，未知の不正 Web サイトである可能性が高いとみて検出する(5). (5)で検出されたアクセス先の Web サイトが，正規 Web サイトか不正 Web サイトのどちらであるかを判別する(6). アクセス先が不正 Web サイトであると判別されたときは(7)，クライアントにアクセス先が不正 Web サイトであることを伝える(8). 検出された新たな不正 Web サイトは，ブラックリストを更新するために提供することで(9)，ブラックリストを常に最新の状態に保つ．検出に関する詳細は 3.4，3.5 節で，判別に関する詳細は 3.6 節で説明する．

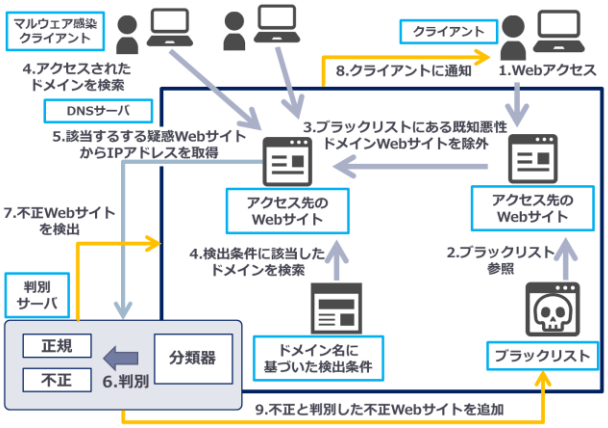


図 3 システム処理の流れ

3.4 マルウェア感染クライアントの検出

疑惑のある Web サイトを選定するために，マルウェアに感染されているクライアント(以下マルウェア感染クライアント)を利用する．マルウェア感染クライアントの検出方法を図 4 に示す．一般にマルウェアは，感染を拡大させるために多数の不正 Web サイトへアクセスを試みる．そのため，不正 Web サイトは，同時に複数のマルウェア感染クライアントからアクセスが行われている可能性が高い．そこで，DNS サーバに Web アクセスしているクライアントの中から(1)，既知の不正 Web サイトのドメイン(悪性ドメイン)にアクセスしているクライアントを探索し(2)，マルウェア感染クライアントとして検出する．このときに検出されたマルウェア感染クライアントを，未知の不正 Web サイトを検出する際に利用する．

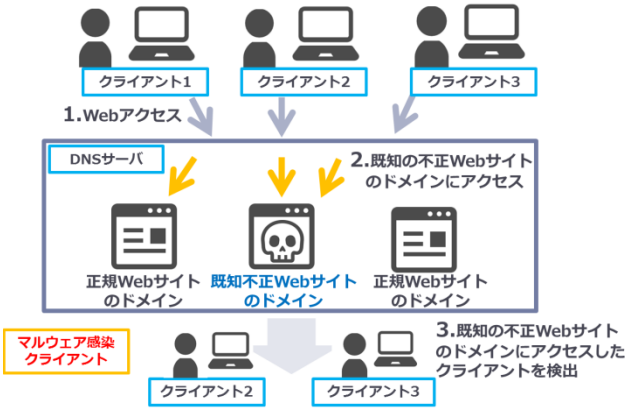


図 4 マルウェア感染クライアント検出

3.5 未知の不正 Web サイトの検出

未知の不正 Web サイトの検出方法を図 5 に示す．クライアントから Web アクセスが行われた際に(1)，アクセスされた Web サイトのドメインと不正 Web サイトのドメインの集合であるブラックリストを照合する(2)．既知の悪性ドメインにアクセスされた場合は，クライアントにアクセス先が不正 Web サイトであることを通知するが(7)，不正 Web サイトのドメインの集合であるブラックリストと一致せず残った Web サイトのドメインは，文字数が 10 文字以上，または英数字が混在したドメイン名を抽出する(3)．さらに複数のマルウェア感染クライアントからアクセスがあるドメインも抽出し(3)，抽出したドメインの IP アドレスを取得する．取得した IP アドレスは，正規 Web サイトか不正 Web サイトのどちらであるか判別するために利用する(4)．このとき，IP アドレスの良性・悪性は，正規 Web サイトに利用される IP アドレスを良性 IP アドレス，不正 Web サイトに利用される IP アドレスを悪性 IP アドレスとし，これらの IP アドレスを用いて特徴ベクトルを構成する．特徴ベクトルについては，3.6 節で詳細に説明する．

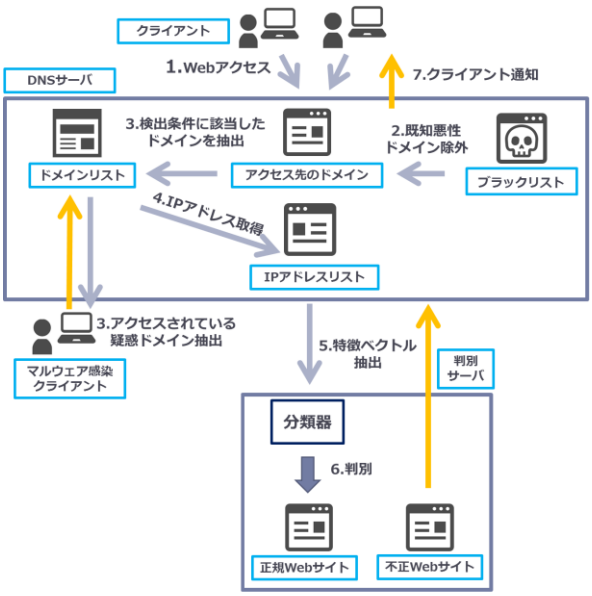


図 5 未知の不正 Web サイトの検出

3.6 IP アドレスを用いた Web サイトの判別方法

未知の疑惑 Web サイトに対する判別方法を図 6 に示す．不正 Web サイトの疑惑がある Web サイトを正規 Web サイトと不正 Web サイトに判別するために，疑惑 Web サイトから IP アドレスを取得し，IP アドレスクラスごとに分けて，それぞれの IP アドレスクラス専用の分類器に適用して判別を行う．このとき，それぞれの分類器には，各クラスの良性 IP アドレスと悪性 IP アドレスを教師データとして用いて分類器を構成する．悪質な活動に利用される IP アドレスは一部のネットワークアドレスに密集する傾向にあることから[8]，ネットワークアドレス部分とホストアドレス部分を合わせた IP アドレス全体を用いて Web サイトの判別を行う[8]より，各 IP アドレスクラスのネットワークアドレス部を用いて判別する手法の方が，様々なネットワークアドレスを利用する Web サイトも考慮した，より正確な判別ができるのではないかと考えられる．判別には，教師あり機械学習法の一つである Support Vector Machine(SVM)を用いる．悪性 IP アドレスが一部のネットワークアドレスに密集する特徴を元に，既知の不正 Web サイトが密集しているネットワークアドレスを教師データとして用いることで，悪性 IP アドレスの近傍に存在する Web サイト，つまり，悪性 IP アドレスの特徴に類似した特徴をもつ Web サイトを，未知の不正 Web サイトとみなして判別する．

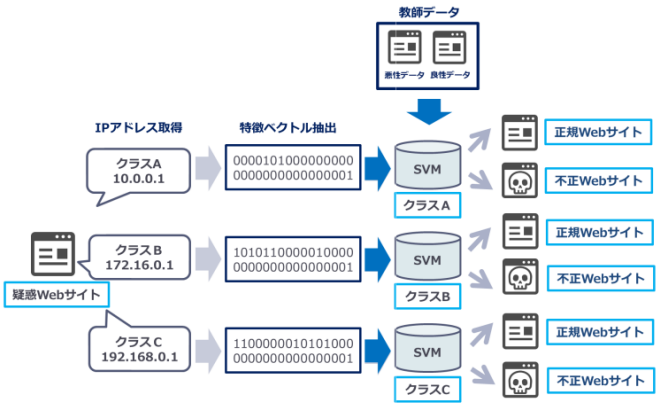
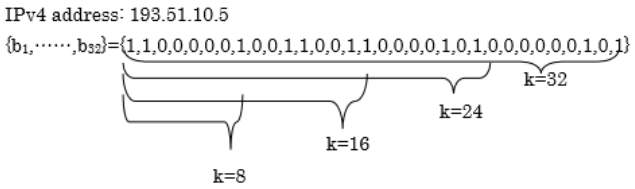


図 6 判別方法

SVM は，分類器の一種であり，新しく入力された未知の IP アドレス(テストデータ)を良性クラスと悪性クラスに分類する．そのため，予め良性・悪性が判明している IP アドレス(教師データ)を用いて分類器を構築する．SVM では，教師データの様々な特徴を元にクラスを構築し，良性クラスと悪性クラスの境目となる決定境界を決める．このとき，クラスを構築するために用いる特徴を特徴ベクトルと呼ぶ．特徴ベクトルの個数を次元数と呼ぶ．本研究では，文献[8]における特徴ベクトル生成手法を参考にし，図 7 のような形で IP アドレスから特徴ベクトルを生成する．IP アドレスをバイナリビット列に変換して構成する．IP アドレスの各ビットを{ b_1,……,b_k}で表す．アドレスクラス A を k=1~8 の 8 次元，アドレスクラス B を k =1~16 の 16 次元，アドレスクラス C を k =1~24 の 24 次元，最後にホストアドレスまで含めた k =1~32 の 32 次元でクラスの決定境界を構成する．



k	特徴ベクトル
8	b_k=1(k=1,2,8) b_k=0(その他)
16	b_k=1(k=1,2,8,11,12,15,16) b_k=0(その他)
24	b_k=1(k=1,2,8,11,12,15,16,21,23) b_k=0(その他)
32	b_k=1(k=1,2,8,11,12,15,16,21,23,30,32) b_k=0(その他)

図 7 特徴ベクトル抽出

教師データとして用いる IP アドレスには，クラスごとに悪性を 1，良性を 0 と設定し，データセットを構成する．データセットの例を表 1 に示す．

表 1 教師データセットの例

IP アドレス	特徴ベクトル	ラベル
193.51.10.5	1,1,0,0,0,0,0,1,0,0,1,1,0,0,1,1	1
10.10.10.10	0,0,0,0,1,0,1,0,0,0,0,0,0,1,0,1	1
203.4.12.89	1,1,0,0,1,0,1,1,0,0,0,0,0,1,0,0	0
…	…	…

悪性の教師データ，良性の教師データセットを表 1 のように作成し，SVM を構築して，図 6 の手法で抽出した疑惑 Web サイトの IP アドレスをテストデータとして用いて分類する．

4. 評価実験

本章では，IP アドレスのネットワークアドレス部を用いた Web サイトの判別の有効性を示す実験について述べる．

4.1 実験データ

本実験では，良性 IP アドレスデータと悪性 IP アドレスデータの 2 種類のデータを標本として使用する．良性 IP アドレスデータは Alexa の公表する Alexa Top Global Sites[10]から良性 IP アドレスを取り出す．悪性 IP アドレスデータは，CCC DATAset[11]に含まれる 2008 年から 2011 年の 4 年分の通信データから，悪性 IP アドレスデータを抽出してリストを構成する．CCC DATAset は，マルウェア検体を収録したボット観測データ群であり，CCC 運営連絡会が運用するサイバークリーンセンターハニーポットで収集したマルウェア検体とウイルス対策ソフト 6 製品での検知名をリスト化したデータである．CCC DATAset には，マルウェア検体，攻撃通信データ，攻撃元データの 3 つから構成されたボット観測データ群が含まれている．

4.2 実験概要

ネットワークアドレスによる分類によって，より正確に判別を図ることができるかについて，SVM を用いて基礎評価実験を行った． IP アドレスは，ネットワーク部とホスト部から成り立っており，ネットワーク部は，その IP アドレスが属しているネットワークを識別するための部分を指す．ホスト部は，ネットワーク内のコンピュータを識別するための部分を指す．IP アドレスはクラス A からクラス E の 5 つのアドレスクラスにわけられている．須藤ら[12]による CCC DATAset の解析によると，CCC DATAset はアドレスクラス A の IP アドレスが頻繁に悪質な活動に利用されている．本実験では，アドレスクラス A のネットワークアドレスを持つデータを用い，良性データ数 9251 個，悪性データ数 46258 個で実験を行い評価した．

4.3 評価方法

精度，適合率，再現率の 3 項目によって評価を行う．評価は，図 7 で定義した k=8，k=16，k=24，k=32 の 4 つで行う．悪性 IP アドレスを正しく悪性 IP アドレスと判定した数を真陽性(TP)，良性 IP アドレスを誤って悪性 IP アドレスと判定した数を偽陽性(FP)，良性 IP アドレスを正しく良性 IP アドレスと判定した数を真陰性(TN)，悪性 IP アドレスを誤って良性 IP アドレスと判定した数を偽陰性(FN)としたとき，精度，適合率，再現率をそれぞれ下記の計算式で評価する．

精度は，テストデータを良性・悪性 2 つのクラスにそれぞれ正確に判別できた割合である．

$$精度 = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

適合率は，悪性と判定した IP アドレスのうち，実際に悪性 IP アドレスであった割合である．

$$適合率 = \frac{TP}{TP + FP} \quad (2)$$

再現率は，全ての悪性 IP アドレスのうち，悪性と判定した IP アドレスの割合である．

$$再現率 = \frac{TP}{TP + FN} \quad (3)$$

4.4 実験結果

アドレスクラス A の評価は，k=8，k=16，k=24，k=32 の 4 つを表 2 にまとめる．アドレスクラス A のネットワークアドレスである k=8 で再現率が一番高い数値を示した．しかし，精度と適合率は k=32 で一番高い数値を示した．

表 2 アドレスクラス A の精度評価

	精度	適合率	再現率
k=8	89.83949	93.13721	94.79225
k=16	89.81787	93.1318	94.77063
k=24	89.81787	93.1318	94.77063
k=32	90.19618	93.54914	94.77063

5. 考察

アドレスクラス A のネットワークアドレスである k=8 で再現率が一番高い数値を示したことから，アドレスクラス A では，悪性の特徴が強く表れており，ネットワークアドレスを用いて悪性 IP アドレスを正確に判別することに適していると考えられる．一方，k=32 で精度と適合率が一番高い数値を示したことから，一見 k=32 の判別が適していると思われるが，k=16，k=24，k=32 の TP は一定であり，悪性 IP アドレスを一番正確に判別した数が多いのは k=8 である．しかし，k=32 で TN が増加した影響で FN が減少したため，適合率の数値が高くなり，全体の精度も向上したと考えられる．従って，悪性 IP アドレスとして多く利用されているアドレスクラス A では，ネットワークアドレスを用いた悪性 IP アドレスの判別が有効であると考えられる．

6. 追加実験

提案方法がアドレスクラス A 以外の Web サイトに対しても有効であるかどうかを評価するため，追加実験を行った．アドレスクラス B は，良性データ数 647 個，悪性データ数 3238 個で実験を行った．アドレスクラス B の評価は，k=8，k=16，k=24，k=32 の 4 つを表 3 にまとめる．アドレスクラス B のネットワークアドレスである k=16 で精度，再現率が一番高い数値を示した．

表 3 アドレスクラス B の精度評価

	精度	適合率	再現率
k=8	79.41328	87.57707	87.73935
k=16	84.01956	89.11809	92.063
k=24	83.45342	89.04604	91.38357
k=32	83.94236	89.22569	91.81594

しかし，他のアドレスクラスに比べ，悪性 IP アドレス数が圧倒的に少ないため，悪性 IP アドレスの教師データ数が不足している．そのため，今後アドレスクラス B の悪性 IP アドレスの利用頻度が変化する可能性があり，新たな特徴に応じた判別方法が必要である．

アドレスクラス C は，良性データ数 14271 個，悪性データ数 75000 個で実験を行った．アドレスクラス C の評価は，k=8，k=16，k=24，k=32 の 4 つを表 4 にまとめる．精度，適合率は k=16 で一番高い数値を示し，再現率はすべて一定の数値を示した．

表 4 アドレスクラス C の精度評価

	精度	適合率	再現率
k=8	67.36454	91.05442	67.81733
k=16	67.43287	91.15396	67.81733
k=24	67.40711	91.11641	67.81733
k=32	67.42839	91.14743	67.81733

このような結果が得られた要因として，アドレスクラス C は他のアドレスクラスより IP アドレスの範囲が狭く，尚且つ，全体的に悪質な活動に多く利用されているため，悪性

IP アドレスの特徴が表れにくかったと考えられる．また，k=8 で判定を間違った IP アドレスは，ビット数が k=16，k=24，k=32 のすべてにおいて判定を誤っていた．これらから，教師データを利用して悪性，良性のクラスを構築する際に，IP アドレスの k=8 の特徴のみでクラスを判定している可能性があるため，判定結果を間違った IP アドレスを間引きする，あるいは，教師データのビットの順序を変更してクラスを構築し，精度向上を図る必要がある．

7. まとめ

本稿では，ユーザが Web サイトにアクセスする際に，アクセス先が不正 Web サイトであるか正確に検出し，未知の Web サイトに対して正規・不正 Web サイトの判別を目的とした．この目的を達成するために，DNS サーバから得られる情報のうち，ドメイン情報と IP アドレス情報，マルウェア感染クライアントを利用して未知の不正 Web サイトへのアクセスにも対応した手法を提案し，IP アドレスクラスのネットワークアドレスを用いて疑惑 Web サイトの「判別」を評価した．悪質な活動に多く利用されているアドレスクラス A の結果が高精度を示したことから，IP アドレスのネットワークアドレス部を用いた判別は，有効であることが確認できた．今後は， アドレスクラス C の精度が全体的に芳しくなかったため，教師データの調整を行い，改善を図る必要がある．

参考文献

[1] 警察庁広報資料: 平成26年中のインターネットバンキングに係る不正送金事犯の発生状況等について, <https://www.npa.go.jp/cyber/pdf/H270212_banking.pdf> (参照 2016-05-10).
[2] TREND MICRO: Web レピュテーション, <http://www.trendmicro.co.jp/why-trendmicro/spn/features/web/index.html> (参照 2016-05-10).
[3] CISCO: テクノロジー解説, <http://www.cisco.com/web/JP/news/cisco_news_letter/tech/index.html> (参照 2016-05-10).
[4] 劉亦晨: DNS 情報による悪意のあるサイトの検出法, 2012 年度 早稲田大学大学院 基幹理工学研究科 情報理工学専攻 修士論文 (2012).
[5] 日立ソリューションズ: 情報セキュリティブログ, <http://securityblog.jp/words/2898.html> (参照 2016-05-10).
[6] 田中晃太郎, 長尾篤, 森井昌克: DNS ログからの不正 Web サイト抽出についてー解析手法とその匿名化ー, コンピュータセキュリティシンポジウム 2013 論文集, Vol.2013, No.4, pp.132-138 (2013).
[7] 千葉大紀,森達哉, 後藤滋樹: 悪性 Web サイト探索のための優先巡回順序の選定法, コンピュータセキュリティシンポジウム 2012 論文集, Vol.2012, No.3, pp.805-812 (2012).
[8] Daiki Chiba, Kazuhiro Tobe, Tatsuya Mori, Shigeki Goto: Detecting Malicious Websites by Learning IP Address Features, Proc. 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet(SAINT2012), pp.29-39 (2012).
[9] 平井有三: はじめてのパターン認識, 森北出版 (2012).
[10] Alexa: The top 500 sites on the web, <http://www.alexa.com/topsites> (参照 2016-05-10).
[11] 秋山満昭, 神薗雅紀, 松木隆宏, 畑田光弘: マルウェア対策のための研究用データセット～MWS Datasets 2014～, 情報処理学会研究報告, Vol. 2014-CSEC-66, No. 19, pp. 1-7, 2014.
[12] 須藤年章: CCC Dataset 2010 によるマルウェア配布元 IP アドレス評価に関する一考察, 情報処理学会シンポジウム論文集, Vol.2010, No.9 pp.19-24 (2010).
[13] 金澤しほり, 中村嘉隆, 高橋修: DNS 情報を用いた不正 Web サイト検知システムの提案, 情報処理学会第 78 回全国大会講演論文集, Vol.2016, pp.3_563-3_564, 2016.