情報処理学会論文誌 Vol.59 No.9 1610–1623 (Sep. 2018)

統合型マルウェア検査サービス Virus Total を用いた悪性ドメイン検知手法

田辺 瑠偉[1,a)] 森 博志[1] 原田 耕也[1] 吉岡 克成[2] 松本 勉[2]

受付日 2017年11月28日, 採録日 2018年6月8日

概要：近年，攻撃者が用意したサーバと通信を行うことで不正活動を行うマルウェアが増加している．これらのサーバには固有のドメインが割り当てられている場合があるため，これらのドメインをブラックリスト化することで，マルウェア感染ホストを検出する手法が広く利用されている．しかし，昨今マルウェアの数が爆発的に増加しており，ブラックリストの迅速な更新が求められている．本稿では，統合型マルウェア検査サービスの一種である Virus Total を用いた悪性ドメイン検知手法を提案する．提案手法は，マルウェアの解析結果が蓄積されている Virus Total から検査対象ドメインを名前解決する検体の情報を取得することで，検査対象ドメインが悪性ドメインであるか判定する．評価実験では，26 種類の既知の悪性ドメイン群に対して提案手法を適用し，期待どおり悪性ドメインとして検知できることを確かめた．また，正規の Web，Mail，NTP，DNS サービスに対応する 100 種類の正規ドメイン群に対して提案手法を適用し，正規ドメインを誤検知する可能性を確かめた．そして，セキュリティベンダのマルウェア解析レポートから抽出した 16 種類の未知のドメイン群に対して提案手法を適用し，既知のブラックリストに記載されるよりも平均 26 日早く悪性ドメインを検知できることを確かめた．最後に，提案手法を実現したシステムを連携先組織の 1 日分の DNS サーバのトラヒックに適用し，その速度性能を確かめた．マルウェアが名前解決するドメインの中には正規のドメインも含まれるが，提案手法はその中から悪性ドメインを検知できる点で有用であり，ブラックリストの拡張に役立てることができる．

キーワード：悪性ドメイン，Virus Total，ブラックリスト

Detecting Malicious Domains Using Virus Total an Integrated Malware Analysis Service

Rui Tanabe[1,a)] Hiroshi Mori[1] Kouya Harada[1] Katsunari Yoshioka[2] Tsutomu Matsumoto[2]

Received: November 28, 2017, Accepted: June 8, 2018

Abstract: In recent years, malware that communicate with compromised servers are increasing. Some of the servers even have characteristic domains that, domain blacklists of these servers are widely used to detect malware infected hosts. However, the portion of malware is increasing and blacklist expansions are required. In this study, we propose a method to detect malicious domains using Virus Total, which is an integrated malware analysis services. The proposal method decides if the target domain is a malicious domain by analyzing details of malware analysis results gathered from Virus Total. At the experiment, we evaluate the proposal method using 26 malicious domains and show that the proposal method can detect malicious domains. Similarly, we evaluate the proposal method using 100 benign domains that were gathered from Web, Mail, NTP, and DNS services and investigate the possibility of benign domains detected as malicious domains. Finally, using 16 domains that were collected from malware analysis reports of security vendors, we show that the proposal method detected malicious domains in average of 26 days faster than known blacklists. Furthermore, we tested the proposal method implemented system against one day DNS server traffic of cooperated organization and evaluated the process speed. Domains that malware query are not all malicious and so we propose a method to detect malicious domains that were queried from malware. In practical, the proposal method can be used to expand blacklists.

Keywords: malicious domain, Virus Total, blacklist

1. はじめに

近年，攻撃者が用意したサーバと通信を行うことで不正活動を行うマルウェアが増加している．これらのサーバには固有のドメインが割り当てられており，マルウェアはこのドメインを名前解決して接続を試みる場合が多い．このため，マルウェアが不正活動に利用するドメイン（以降では，悪性ドメインと呼ぶこととする）をブラックリスト化することでマルウェア感染ホストを検知する手法が広く利用されている．

たとえば，企業や官公庁などといった組織では，ネットワークの管理者を中心にネットワークトラヒックの分析が行われている．ネットワーク管理者は，収集したトラヒックをブラックリストとマッチングすることでマルウェア感染ホストを特定する．このため，悪性ドメインに関する研究開発が活発に行われており，多数のブラックリストが公開されている [19], [20], [21], [22], [23], [24], [25], [26]．しかし，昨今マルウェアの数が爆発的に増加しており，ブラックリストの拡充が求められている．また，過去にマルウェアが名前解決したドメインが現在も不正活動に利用されているとは限らないため，ブラックリストの迅速な更新が必要である．

本稿では，統合型マルウェア検査サービスの一種である Virus Total [13] に蓄積されているマルウェアの解析結果を用いて，検査対象ドメインの良悪性判定を行う手法を提案する．Virus Total とは，ユーザから投稿された検体を多数のウィルス対策エンジン群やサンドボックスで解析するオンラインサービスであり，毎日約百万近くのファイルが検体として投稿されている．投稿された検体の解析結果は自動的に蓄積され，ユーザは過去に投稿された検体の解析結果を検索することができる．本稿では Virus Total が蓄積する解析結果のうち，ウィルス対策エンジン群による検知結果と Virus Total によるサンドボックス解析時に各検体が名前解決したドメインに着目する．ウィルス対策エンジンにマルウェアであると判定された検体が名前解決したドメインは悪性ドメインである可能性があるが，マルウェアの中には正規のドメインを名前解決するものも多く注意が必要である．そこで提案手法では，検査対象ドメインを名前解決した検体のうち，(1) 1 つ以上のウィルス対策ソフトで悪性であると判定された検体の数，(2) どのウィルス対策ソフトでも悪性と判定されなかった検体の数，(3) 1 つ以上のウィルス対策ソフトで悪性であると判定された検体の最終投稿日時などから検査対象ドメインが悪性ドメインであるか判定する．提案手法はドメインの良悪性判定を目的としており，ネットワーク管理者などがブラックリストに新たな悪性ドメインを追加することで，より多くのマルウェア感染ホストの検知を目指す．このため，提案手法の運用形態の 1 つとして，連携先組織の 1 日分の DNS サーバのトラヒックを入力として，Virus Total を用いて検査対象ドメイン群の良悪性判定を行い，その結果をもとにブラックリストを 1 日単位で更新するシステムを実現した．

評価実験では，マルウェア長期動的解析やセキュリティベンダの解析レポートから特定した 26 種類の悪性ドメイン群に対して提案手法を適用し，提案手法が悪性ドメインを検知できることを確かめた．次に，正規の Web，Mail，NTP，DNS サービスが運用されている 100 種類の正規ドメイン群に対して提案手法を適用し，提案手法が正規ドメインを誤検知する可能性を確かめた．そして，これらのドメイン群をデータセットとして機械学習を行い，検査対象ドメインの良悪性判定を行うためのパラメータの組合せを決定した．マルウェアが名前解決するドメインは様々であり，正規のドメインを名前解決するものも存在する．そこで，Symantec セキュリティレスポンス [14] が公開しているマルウェア解析レポートから抽出した 16 種類のドメインを未知のドメイン群として提案手法に適用した．実験の結果，10 種類のドメインを悪性ドメインと判定した．これらのドメインのうち 7 種類のドメインはブラックリストに記載されるよりも平均で 26 日以上早く検知することができた．また，悪性と判定したドメインのうち 2 種類のドメインは観測期間内にブラックリストに記載されなかった．このように，提案手法はマルウェアが名前解決するドメインの中から悪性ドメインを検知できる点で有用であり，ブラックリストの拡張に役立てることができる．最後に，2018 年 3 月 1 日に連携先組織の DNS サーバで取得した 1 日分の DNS トラヒックを用いて提案手法の速度性能を評価した．実験の結果，上記の DNS サーバで収集した 63,623 ドメインを Virus Total でドメイン検索し，その結果が得られるまでの時間は 1 ドメインあたり平均 1.005 秒であった．また，ドメイン検索の結果を用いてそのドメインの良悪性判定に必要な時間は 1 ドメインあたり平均 0.002209 秒であった．

以降の構成は次のとおりである．2 章で統合型マルウェア検査サービス Virus Total を説明し，3 章で提案手法について述べる．そして，4 章で評価実験を説明し，5 章で考察を行う．最後に，6 章で関連研究を紹介し，7 章でまとめと今後の課題を述べる．

2. 統合型マルウェア検査サービス

統合型マルウェア検査サービスとは，任意のユーザか

[1] 横浜国立大学
Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
[2] 横浜国立大学大学院環境情報研究院/横浜国立大学先端科学高等研究院
Graduate School of Environment and Information Sciences and Institute of Advanced Sciences, Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
[a)] tanabe-rui-nv@ynu.jp

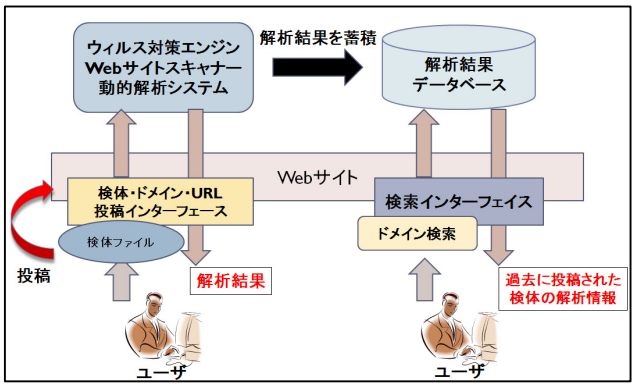

図 1 Virus Total の仕組み
Fig. 1 Mechanism of Virus Total.

ら投稿されたファイルなどを分析エンジンにより解析した結果を提示するサービスである．代表的なものとして Jotti's malware scan [15], Malwr [16], Virus Total [13] といったサービスが存在する．本章では，統合型マルウェア検査サービスの実態について，Virus Total を例にあげて説明する．

2.1 Virus Total

Virus Total とはユーザから投稿された検体，ドメイン，URL をウィルス対策ソフト，Web サイトスキャナ，動的解析システムなどを用いて解析した結果を返すオンラインサービスである．**図 1** に Virus Total の仕組みをまとめる．ユーザは，Virus Total が提供するインタフェースを通して様々なサービスを受けられる．たとえば，ユーザは投稿インタフェースを通して検体などの投稿を行い，その解析結果を得ることができる．また，投稿された検体の解析結果は自動的に蓄積されるため，ユーザは検索インタフェースを通して過去に投稿された検体の解析結果を得ることができる．これらのサービスは任意のユーザが利用できるが，投稿された検体をダウンロードする機能など一部有料なサービスも存在する．Virus Total には毎日約百万近くのファイルが検体として投稿されており，世界中から集められた情報が集約されている．このため，Virus Total をはじめとする統合型マルウェア検査サービスに蓄積された解析結果から標的型攻撃などの高度な攻撃に利用されたマルウェアを特定する手法や [27]，統合型マルウェア検査サービスの実態調査を行った研究 [28] が存在する．

2.2 Virus Total の検索機能

Virus Total の検索インタフェースでは，ハッシュ値，URL，ドメイン，IP アドレスを入力することで，過去に投稿された検体の解析結果から，入力データと関連のある解析結果を検索することができる．たとえば，ドメイン名を検索した場合，次の 10 種類の情報が得られる．ただし，Virus Total の無料サービスでは単位時間あたりに検索可能なドメインの数は限られており，有料サービスで制限を緩和することができる．

ドメイン検索で得られる情報

① 検索ドメインをこれまでに名前解決したときに割り当てられていた IP アドレスのリスト
② 検索ドメインの WHOIS 情報
③ 検索ドメインのサブドメインのリスト
④ 検索ドメインを含む URL のうち，URL スキャナや悪性 URL リストで悪性と判定された URL のリスト
⑤ 検索ドメインからダウンロードされた検体のうち，1 つ以上のウィルス対策ソフトで悪性と判定された検体のリスト
⑥ 検索ドメインからダウンロードされた検体のうち，ウィルス対策ソフトで悪性と判定されなかった検体のリスト
⑦ 解析環境内で実行したときに検索ドメインを名前解決した検体のうち，1 つ以上のウィルス対策ソフトで悪性と判定された検体のリスト
⑧ 解析環境内で実行したときに検索ドメインを名前解決した検体のうち，ウィルス対策ソフトで悪性と判定されなかった検体のリスト
⑨ 検索ドメインが埋め込まれている検体のうち，1 つ以上のウィルス対策ソフトで悪性と判定された検体のリスト
⑩ 検索ドメインが埋め込まれている検体のうち，ウィルス対策ソフトで悪性と判定されなかった検体のリスト

なお，Virus Total に蓄積されているデータは過去にマルウェア検体を解析した結果であるが，Virus Total は過去に投稿された検体を再解析することで解析結果を更新するサービスを提供している．

3. 統合型マルウェア検査サービス Virus Total を用いた悪性ドメイン検知手法

本章では，統合型マルウェア検査サービス Virus Total を用いて悪性ドメインを検知する手法を提案する．3.1 節で提案手法の基本アイディアを説明し，3.2 節で提案手法の流れを説明する．そして，3.3 節で Virus Total を用いた悪性ドメイン検知システムを説明する．

3.1 基本アイディア

提案手法は，Virus Total に蓄積されている解析結果を用いて悪性ドメインを検知する．以下では，実際にあるマルウェアが名前解決する悪性ドメイン "e.ppift.com" [17]（MD5 ハッシュ値：071eb0abdcb09e64a621fbab707afe7e, Symantec Norton による名称：W32.Morto.B）を Virus Total で検索した際に出力される結果を用いて提案手法の基本アイディアを説明する．

Symantec のマルウェア解析レポートから，W32.Morto.B

表 1 Virus Total で “e.ppift.com” [17] を検索した際に出力される当該ドメインを名前解決した検体のリスト
Table 1 List of samples that query “e.ppift.com” [17], obtained from Virus Total search engine.

検知結果	解析日時	名前解決した検体のMD5ハッシュ値
53/56	2016/3/18 13:55	3c17cd4e52434b6e867d0720cb85fa4b
48/55	2016/2/17 20:32	5ff31717c0caac4c61cd6a416eef046a
49/56	2016/1/24 8:06	d7709f2bb343389560938ff460ecacae
45/53	2015/12/28 9:50	aa80128ba9ff8bd79d8397ddc3fceb0b
47/56	2015/12/1 9:48	42115192b50b400e680329590620ed28
…	…	…
41/46	2013/8/19 0:51	8f1f3844701b8b176b4733b02aeef102
41/45	2013/8/19 0:50	9a2f3e18a0cfa19f63065c676622d413
43/46	2013/8/19 0:47	b04df22a2a4fbec864f3b6fcaff12800

表 2 Virus Total で “e.ppift.com” [17] を検索した際に出力される結果の例
Table 2 Example of results obtained from Virus Total when searching “e.ppift.com” [17].

①検索ドメインをこれまでに名前解決したときに割り当てられていたIPアドレス	6種類
②検索ドメインのWHOIS情報の有無	有り
③検索ドメインのサブドメイン	1種類
④検索ドメインを含むURLのうち，悪性と判定されたURL	1種類
⑤検索ドメインからダウンロードされた検体のうち，悪性と判定された検体	無し
⑥検索ドメインからダウンロードされた検体のうち，悪性と判定されなかった検体	無し
⑦解析環境内で実行したときに検索ドメインを名前解決した検体のうち，悪性と判定された検体	100種類
⑧解析環境内で実行したときに検索ドメインを名前解決した検体のうち，悪性と判定されなかった検体	無し
⑨検索ドメインが埋め込まれている検体のうち，悪性と判定された検体	無し
⑩検索ドメインが埋め込まれている検体のうち，悪性と判定されなかった検体	無し

は “e.ppift.com” をはじめとするドメインを用いて，悪質なファイルをダウンロードすることが知られている [17]．**表 1** に 2017 年 11 月時点で当該ドメインを検索した際に出力される，⑦検索ドメインを名前解決した検体のリストをまとめる．なお，表 1 における検知結果とは，悪性と判定されたウィルス対策ソフト数/悪性判定に用いたウィルス対策ソフトの数である．また，**表 2** に当該ドメインを検索した際に出力される結果の例をまとめる．ここで，④検索ドメインを含む URL のうち悪性 URL の数を DU 値（Detected URLs），⑦検索ドメインを名前解決した悪性検体の数を DFC 値（Detected Files that Communicate with this domain）と定義する．また，⑥⑧⑩ウィルス対策ソフトで悪性と判定されなかった検体の合計数を UF（Undetected Files）値と定義する．たとえば，当該ドメインは DU 値 1，DFC 値 100，UF 値 0 となる．DFC 値の結果から，Virus Total には当該ドメインを名前解決するマルウェア検体が多数投稿されていることが分かる．一方，UF 値の結果から，当該ドメインを名前解決する検体はすべて，Virus Total に登録されているウィルス対策ソフトのいずれかで悪性と判定されたことが分かる．以上の結果から，当該ドメインはマルウェアが不正活動に利用しており，正規のファイルが当該ドメインを名前解決する可能性は低いことが予想される．このように，Virus Total に蓄積されているマルウェアや正規ファイルの情報は，ドメインの良悪性判定に有効である．そこで，Virus Total の解析結果を用いて悪性ドメインを検知する手法を提案する．

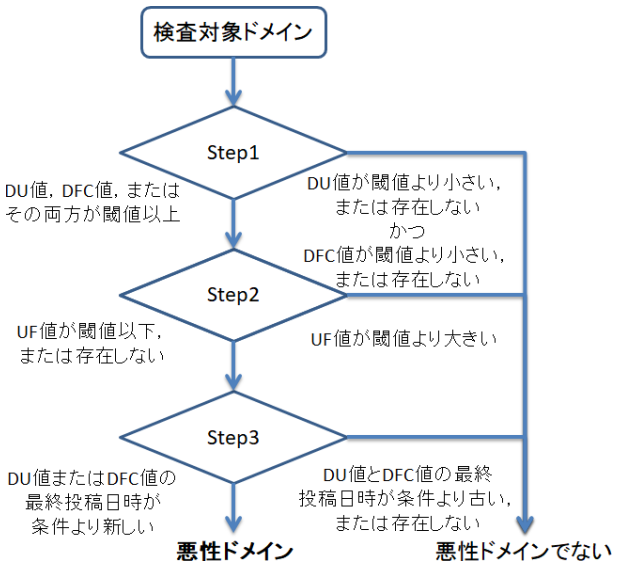

図 2 提案手法の流れ
Fig. 2 Flow of proposal method.

3.2 統合型マルウェア検査サービス Virus Total を用いた悪性ドメイン検知手法

本節では，提案手法の流れについて説明する．**図 2** に提案手法の全体像をまとめる．提案手法は，検査対象ドメインを Virus Total でドメイン検索し，3 つの step により検査対象ドメインの良悪性判定を行う手法である．

検査対象ドメインがマルウェアの不正活動に利用されていた場合，Virus Total に当該ドメインを名前解決するマルウェア検体が投稿されている場合がある．特に，不特定多数のマシンに感染するマルウェアは Virus Total に解析結果が存在する可能性は高い．そこで，まず初めに Step1 で DU 値と DFC 値を取得して，検査対象ドメインがマルウェアに関連していることを確かめる．ただし，マルウェアの中には正規サービスと通信を行うものが存在する．このため，マルウェアが名前解決するドメインがすべて悪性であるとは限らない．そこで，次に Step2 で UF 値を取得して，検査対象ドメインが正規ファイルに関連していないことを確かめる．一方，過去にマルウェアの不正活動に利用されていたドメインが現在も有効であるとは限らない．実際に，攻撃者はブラックリストによる検知を防ぐため，悪性ドメインを短期間でのみ利用する場合がある．そのため，最後に Step3 で検索ドメインを名前解決した検体のうち，1 つ以上のウィルス対策ソフトで悪性と判定された検体の最終投稿日時を取得して，検査対象ドメインが不正活動に利用されていた期間を確かめる．提案手法は，Step1 から順にドメインの判定を行い，すべての Step で悪性であると判定された場合にのみ，検査対象ドメインを悪性ドメインであると判定する．以下では，各 Step について説

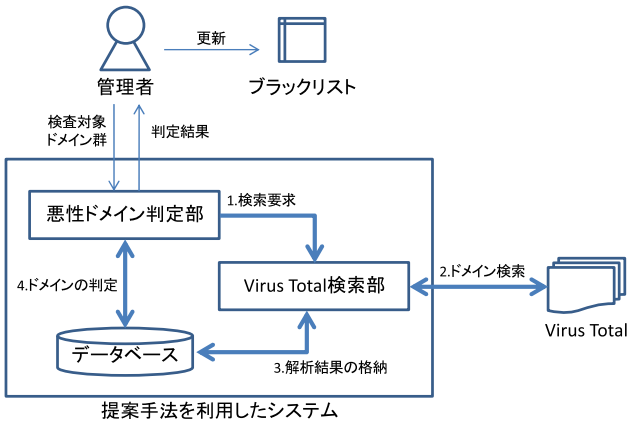

図 3 提案手法の利用の流れ

Fig. 3 Flow of system that use proposal method.

明する．

STEP1：ウィルス対策ソフトで悪性と判定される検体/URL

検査対象ドメインを Virus Total で検索し，DU 値，DFC 値を取得する．DU 値と DFC 値がどちらも閾値以上であれば Step2 へ移行する．また，DU 値と DFC 値のどちらか一方が閾値以上である場合にも Step2 へ移行する．しかし，DU 値と DFC 値のどちらも閾値よりも小さい場合には悪性ドメインでないと判定する．また，DU 値と DFC 値がどちらも存在しない場合には悪性ドメインでないと判定する．

STEP2：ウィルス対策ソフトで悪性と判定されない検体

検査対象ドメインを Virus Total で検索し，UF 値を取得する．UF 値が閾値以下，あるいは存在しない場合には Step3 へ移行する．一方，UF 値が閾値よりも大きければ悪性ドメインでないと判定する．

STEP3：検体の最終投稿日時 検査対象ドメインを Virus Total で検索し，検索対象ドメインを名前解決した検体のうち，1 つ以上のウィルス対策ソフトで悪性と判定された検体の最終投稿日時が条件よりも新しければ悪性ドメインと判定する．あるいは，検索対象ドメインとこれまでに対応関係があった URL のうち，URL スキャナや悪性 URL リストで悪性と判定された URL の最終投稿日時が条件よりも新しければ悪性ドメインと判定する．一方，どちらも条件よりも古い，あるいは存在しない場合には悪性ドメインでないと判定する．

3.3 統合型マルウェア検査サービス Virus Total を用いた悪性ドメイン検知システム

本節では，Virus Total を用いて検査対象ドメインの良悪性判定を行うシステム（以降では，システムと呼ぶこととする）について説明する．**図 3** にシステムの全体像をまとめる．

システムの運用形態として，組織のネットワーク管理者が提案手法を利用して，(1) セキュリティベンダの解析レポートなどから収集したドメインの良悪性判定を行い，組織内で利用されているブラックリストに新たな悪性ドメインを追加する場合，(2) ブラックリストに含まれるドメインの良悪性判定を行い，良性であると判定された場合にはブラックリストから除外する場合，(3) 組織内の DNS サーバで収集した DNS トラヒックの良悪性判定を行い，ネットワークベースでのマルウェア感染ホストの検知や，ブラックリストへ悪性ドメインを追加する場合，(4) 組織内に存在するユーザマシン上でドメインの良悪性判定を行い，ホストベースでのマルウェア感染検知やブラックリストへ悪性ドメインを追加する場合が想定される．

システムの運用形態の 1 つとして，4 章の評価実験では，ネットワーク管理者が実際にブラックリストに新たなドメインを追加する状況を仮定する．インターネット上には，多数のマルウェア解析レポート公開されている．たとえば，Symantec セキュリティレスポンス [14] は，マルウェアが名前解決するドメインとその目的（C&C，検体ダウンロード，ネットワーク接続の確認，など）が記載されているレポートを公開している．しかし，解析レポートのドメインすべてがマルウェアの不正活動に利用されているとは限らない．また，一般にネットワーク管理者はネットワーク内のトラヒックを収集することができる．しかし，組織内の DNS サーバなどで収集したドメインには多数の正規ドメインが存在する．このため，これらのドメインをブラックリストに含める場合には誤検知について検討する必要がある．そのため，ネットワーク管理者は当該システムを用いてドメインの良悪性判定を行い，事前に収集したドメイン群から悪性ドメインを抽出することでブラックリストに新たなドメインを追加するものとする．以下では，システムの流れや構成を説明する．なお，その他の運用形態については 5 章で考察する．

当該システムは，Virus Total に対してドメイン検索を行う Virus Total 検索部，得られた解析結果を蓄積するデータベース，検査対象ドメインの良悪性判定を行う悪性ドメイン判定部から構成される．管理者は検査対象ドメイン群を入力することで，各ドメインの判定結果を得ることができる．また，その結果を利用してブラックリストの更新を行う．検査対象ドメインの良悪性判定はパラメータの組合せ（DU 値，DFC 値，UF 値，最終投稿日時の閾値や条件）に応じて結果が変化する．しかし，大規模なネットワークなどで誤検知を少なくしたい場合や，その一方で，基幹システムなどで見逃しを少なくしたい場合など，管理者のニーズは様々である．このため，4 章の評価実験では機械学習を用いてパラメータの組合せを決定する．なお，Virus Total ではドメイン検索を行った際，Virus Total のデータベースに蓄積されている解析結果が 100 件を超えていた場合，検体解析日時の新しいものから順に，最大で 100 件の

解析結果が出力される仕組みとなっている．このため，DU 値，DFC 値，UF 値の上限は 100 である．

悪性ドメイン判定部：Virus Total 検索部に検査対象ドメインの検査を要求する（図 3 の 1）．その後，データベースにドメイン名と検索日を入力して検査対象ドメインの解析結果を取得し，3.2 節で提案した手法を用いて検査対象ドメインの良悪性判定を行う（図 3 の 4）．最後に，管理者に判定結果を返す．

Virus Total 検索部：悪性ドメイン判定部からの要求により，検査対象ドメインを Virus Total で検索して解析結果を取得する（図 3 の 2）．

データベース：Virus Total 検索部で取得した結果を蓄積する（図 3 の 3）．Virus Total の解析結果は時間とともに追加されることが予想される．このため，データベースにはドメイン名，ドメイン検索を行った日付，解析結果の 3 種類の情報が蓄積される．

4. 評価実験

提案手法は，Virus Total に蓄積されている解析結果を用いて検査対象ドメインが悪性であるか判定する．このため，まず初めに 4.1 節でマルウェア動的解析により得られた悪性ドメイン群に対して提案手法を適用し，悪性ドメインを検知できることを確かめる．次に，4.2 節で正規のドメイン群に対して提案手法を適用し，正規ドメインを誤検知しないことを確かめる．そして，4.3 節で機械学習を用いて提案手法に用いるパラメータの組合せを決定し，セキュリティベンダの解析レポートに記載されていた 16 種類のドメイン群を未知のドメイン群として，提案手法が未知のドメイン群から悪性ドメインを検知できることを確かめる．また，既存のブラックリストと比較することで，提案手法が悪性ドメインをどの程度早く検知できるか確かめる．最後に，4.4 節で提案手法を実現したシステムを連携先組織の 1 日分の DNS サーバのトラヒックに適用し，その速度性能を確かめた．

4.1 悪性ドメイン群を用いた検知精度評価（実験 1）

実験方法：2012 年 5 月から 2017 年 5 月までの間に連携先組織や低対話型ハニーポットを用いて収集した実マルウェア検体の中から，インターネット上のホストと通信を行う検体を 9 種類抽出した．そして，それらの検体群が名前解決するドメインのうち，長期動的解析やセキュリティベンダの解析レポートを用いて特定した 26 種類の悪性ドメイン群に対して提案手法を適用し，検知精度を評価した．**表 3** に評価に用いたマルウェア検体と名前解決した悪性ドメインをまとめる．なお，検体の分類には 2017 年 6 月時点での Symantec Norton の検知結果を用いた．

実験結果：提案手法の検知能力はその設定によって変化する．そこで，Step1，Step2 における閾値を個別に評価し

表 3 提案手法の検知精度評価に用いたマルウェア検体とその検体が名前解決する悪性ドメイン（実験 1）

Table 3 Malware samples and malicious domains used for evaluating proposal method (experiment1).

Symantec検知名	md5ハッシュ値	悪性ドメイン
Ransom.Wannacry	7339a0efc768310a86b6d4f61d88b910	www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com
Trojan Horse	65dc0682604e08c4bb2201ea67204181	mail.loaadss.pl
		mx2.finansgroups.com
		mx3.finansgroups.com
		mx4.finansgroups.com
		mx5.finansgroups.com
		mail7.digitalwaves.co.nz
		mxs.mail.ru
W32.Gobot.A	4681d09d953a3952208b9e55aefccffb	fucko.servebeer.com
		fucko1.servebeer.com
		fucko2.servebeer.com
W32.IRCbot	742cc19b2e31090a0a657381b8ced269	proxim.ircgalaxy.pl
W32.IRCbot	742cc19b2e31090a0a657381b8ced269	kreten.banjalucke-ljepotice.ru
		prcolina.prichaonica.com
		sombrero.balkan-hosting.net
W32.Korgo.S	80132461323e6c274c4eab9e518598a0	proxim.ntkrnlpa.info
W32.Morto.B	071eb0abdcb09e64a621fbab707afe7e	e.ppift.com
		e.ppift.in
		e.ppift.net
W32.Ramnit.B!inf	0e471a19871bff0b76946fb582365696	rterybrstutnrsbberve.com
		erwbtkidthetcwerc.com
		rvbwtbeitwjeitv.com
WS.Reputation.1	c3cecaa020a72f1302f470dd50f7a84e	jebena.ananikolic.su
		juice.losmibracala.org
		peer.pickeklosarske.ru
		teske.pornicarke.com

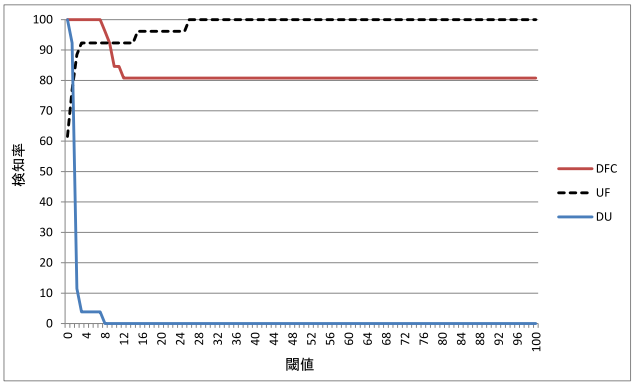

図 4 提案手法の Step1，2 における閾値と悪性ドメインの検知率の関係（実験 1）

Fig. 4 Detection result of proposal method step1 and step2, against malicious domains (experiment1).

たときの検知結果を調査した．**図 4** に 2017 年 6 月 7 日時点における提案手法の検知結果をまとめる．また，検体最終投稿日時によって検知結果がどのように変わるか調査した．図内の赤実線が Step1 の DFC 値を閾値以上に設定したときの検知結果であり，閾値を 7 以下に設定するとすべてのドメインを検知することができる．続いて，各ドメインの具体的な値を調査すると，5 種類ドメイン（約 20%）は DFC 値が 7 以上 11 以下であった．検体最終投稿日時については，4 種類ドメインは 2013 年から 2014 年であり，過去に流行したマルウェアが利用していたことが予想される．残りの 1 種類のドメインは検体最終投稿日時が 2017 年 5 月であったが，攻撃者はすでに新たなドメインを用いてマルウェアを制御していることが報告されている [29]．一

方，21 種類のドメイン（約 80%）は DFC 値が 100 であった．検体最終投稿日時を調査したところ，2016 年 11 月から 2017 年 6 月であり，比較的最近まで検体投稿が行われていたことが分かる．これらの結果から，マルウェアを網羅的に検知したい場合には，DFC 値を低く，検体最終投稿日時の条件を古い検体が含まれる条件に設定する必要がある．一方，Virus Total において比較的最近まで投稿が確認されていたマルウェアを検知したい場合には，DFC 値を高く，検体最終投稿日時の条件を古い検体が含まれない条件に設定する必要がある．続いて，図内の青実線が Step1 の DU 値を閾値以上に設定したときの検知結果であり，21 種類のドメイン（約 80%）は DU 値が 1 となった．また，DU 値の最大は 7 であった．今回の実験では，悪性ドメインとこれまでに対応関係があった URL は少なかった．しかし，マルウェアの中には HTTP プロトコルを介して攻撃者と通信を行うマルウェアも存在する [30]．このため，5 章で考察する．一方，図内の黒点線が Step2 の UF 値を閾値以下に設定したときの検知結果であり，閾値を 26 以上に設定するとすべてのドメインを検知することができる．続いて，各ドメインの具体的な値を調査すると，11 種類のドメイン（約 42%）は UF 値が存在しなかった．また，13 種類のドメイン（50%）は UF 値が 1 以上 3 以下であった．この結果から，悪性ドメインを名前解決するファイルの多くは，1 つ以上のウィルス対策ソフトで検知されていたことが分かる．一方，UF 値が 15，26 となるドメインが存在した．このため，見逃しを少なくするためには，UF 値を高く設定する必要がある．なお，実際には Step1，2，3 の組合せによりドメインの判定を行う．

4.2 正規ドメイン群を用いた検知精度評価（実験 2）

実験方法：2016 年 12 月に収集した ALEXA [19] の Global top 73 サイトのドメイン 73 種類，google や yahoo などのサービスのメールサーバのドメイン 8 種類，プール NTP サーバのドメイン 6 種類，DNS ルートサーバのドメイン 13 種類，を含む合計 100 種類のドメインを正規のドメインと仮定し，これらのドメイン群に対して提案手法を適用してその検知精度を評価した．

実験結果：初めに，Step1，Step2 における閾値を個別に評価したときの検知結果を調査した．**図 5** に 2017 年 6 月 7 日時点における提案手法の検知結果をまとめる．また，検体最終投稿日時によって検知結果がどのように変わるか調査した．図内の赤実線が Step1 の DFC 値を閾値以上に設定したときの検知結果であり，閾値を高くすることで検知率（誤検知）は低くなる．閾値を 80 に設定した場合の検知率は 50%であり，閾値を 100 に設定した場合の検知率は 45%である．また，20 種類のドメインは DFC 値が 0 であった．一方，検体最終投稿日時が 2017 年以降のものも複数存在した．続いて，図内の青実線が Step1 の DU 値を

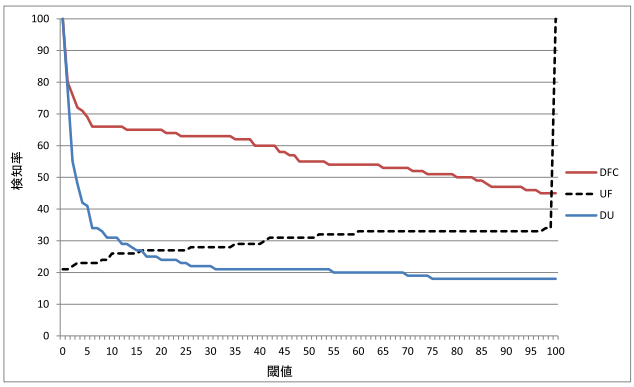

図 5 提案手法の Step1，2 における閾値と正規ドメインの検知率の関係（実験 2）

Fig. 5 Detection result of proposal method step1 and step2, against benign domains (experiment2).

閾値以上に設定したときの検知結果であり，DFC 値同様に閾値を高くすることで検知率（誤検知）は低くなる．閾値を 3 に設定した場合の検知率は 48%であり，閾値を 100 に設定した場合の検知率は 18%である．また，22 種類のドメインは DFC 値が 0 であった．これらの結果から，正規ドメインの多くは DFC 値や DU 値が低くなることが分かる．このため，正規ドメインを誤検知しないためには DFC 値と DU 値を高く設定する必要がある．一方，一部のドメインで DFC 値や DU 値が高くなった理由として，マルウェアの中に正規のサービスを利用して不正活動を行うものが存在することがあげられる [18]．このようなマルウェアはブラックリストを用いて検知することは難しく，5 章で考察する．最後に，図内の黒点線が Step2 の UF 値を閾値以下に設定したときの検知結果であり，閾値を高くすることで検知率（誤検知）は高くなる．閾値を 0 に設定した場合の検知率は 21%であり，閾値を 100 に設定した場合の検知率は 100%である．この結果から，正規ドメインの多くは UF 値が高くなることが分かる．このため，正規ドメインを誤検知しないためには UF 値を低く設定する必要がある．なお，実際には Step1，2，3 の組合せによりドメインの判定を行う．

4.3 未知のドメイン群を用いた検知精度評価（実験 3）

実験 1 では悪性ドメイン群を用いた提案手法の検知精度を，実験 2 では正規ドメイン群を用いた提案手法の検知精度を評価した．本節では，実験 1 に用いた 26 種類の悪性ドメインと，実験 2 で用いた 100 種類の正規ドメイン（以降ではこれらのドメインを良性ドメインと呼ぶこととする），合計 126 種類のドメインを用いて進化計算アルゴリズムの一種である差分進化アルゴリズムにより，悪性ドメイン判定部のパラメータの組合せ（DFC 値，DU 値，UF 値，最終投稿日時の閾値や条件）を決定した．そして，提案手法を実現したシステムに決定したパラメータを設定すること

で，提案手法が未知のドメイン群から悪性ドメインを検知できることを確かめた．また，既存のブラックリストと比較することで，提案手法が悪性ドメインをどの程度早く検知できるか確かめた．

パラメータの決定方法：126 種類のドメイン群を用いて 5 個のデータセットを作成し，クロスバリデーションを行った．また，悪性ドメイン数と良性ドメイン数の割合の違いによる判定精度への影響を避けるため，作成した 5 個のデータセットにはそれぞれ同数の悪性ドメインと良性ドメインが含まれるようにした．具体的には，26 種類の悪性ドメインからランダムに 20 種類のドメインを選択し，100 種類の良性ドメインからランダムに 20 種類のドメインを選択し，計 40 種類のドメイン群を 1 つのデータセットとして 5 個のデータセットを作成した．このため，データセット間でのドメインの重複は許容する．次に，作成したデータセットから 10 種類の悪性ドメインと 10 種類の良性ドメインを訓練データとし，python の scipy モジュールの differential_evolution 関数を用いてパラメータの組合せを決定した．そして，残りの 10 種類の悪性ドメインと 10 種類の良性ドメインをテストデータとして，パラメータの組合せの評価を行った．なお，学習に使用する評価関数には，False Positive，False Negative のどちらの場合でも判定に失敗したドメインごとにスコアが 1 加算される関数を使用した．実験の結果，検知精度の平均値は適合率が 0.98，再現率が 1.0 となった．また，データセットごとに決定した合計 5 種類のパラメータの組合せで 126 種類のドメイン群の良悪性判定を行ったところ，検知精度の平均値は適合率が 0.85，再現率が 0.99 となった．最も精度の良い結果となったパラメータの組合せは DU = 15，DFC = 6，UF = 15，最終投稿日時 1,590 日であり，当該パラメータの組合せで 126 種類のドメイン群の良悪性判定の結果は適合率が 0.93，再現率が 1.0 であった．以降の実験では，当該パラメータの組合せを用いる．

ブラックリストの登録日と提案手法の検知日について：各ドメインの登録日が記載されているブラックリストについては正確な登録日を用いた．一方，各ドメインの登録日が記載されていないブラックリストについては 2 日に 1 度当該リストを取得し，各ドメインの登録状況を確認することで登録日を確認した．提案手法によるドメインの判定は 2 日に 1 度行った．

実験方法：提案手法の運用形態は様々であるが，実験 3 ではネットワーク管理者が実際にブラックリストに新たなドメインを追加する状況を仮定する．そこで，2015 年から 2016 年までの間に Symantec セキュリティレスポンス [14] が公開した，インターネット上で感染が確認されている 6 種類のマルウェアが名前解決する 16 種類のドメインを未知のドメイン群と仮定し，提案手法を用いて悪性ドメインの検知を試みた．Symantec セキュリティレスポンスの解析レポートはインターネット上で公開されており，ブラックリスト作成サービスも収集することができるため，これらのドメイン群が既知のブラックリストに記載される可能性がある．そこで，提案手法で検知したドメインが既知のブラックリストに記載されている場合，検知日にどの程度差異が存在する調査した．このように，提案手法と既知のブラックリストを比較することで提案手法の検知精度を評価した．**表 4** に実験に用いたブラックリストを，**表 5** に実験に用いた検体の内訳をまとめる．なお，解析レポートの収集やブラックリストに記載されているドメインの調査は 2016 年 4 月から 2016 年 6 月までの間に 2 日に 1 度行った．

実験結果：2016 年 4 月から 6 月までの 3 カ月の間に，16 種類のドメインに対して提案手法を適用した．**表 6** にその結果をまとめる．以降では，提案手法で検知したドメインをマルウェアごとに説明する．

Trojan.Huntpos [32], [33]：2016 年 4 月 4 日，当該マルウェアが C&C サーバとの接続に用いる 4 種類のドメインに対して提案手法を適用した．その結果，いずれのドメインについても検知することができなかった．このようになった理由として，当該マルウェアは POS (Point Of Sale) システムを狙っており，Virus Total にはこれらのドメインに関連するファイルがあまり投稿されていないなどの理由があげられる．実際に，実験に用いたブラックリストにおいても 4 種類のドメインのうち，2 種類のドメインのみが記載されていた．提案手法の Step1：DU 値を 4 以下に設定することで 1 種類のドメインを検知できるが，POS システムを狙うマルウェアを検知するためにはさらなる調査が必要である．

Trojan.Zlob.Q [34]：2016 年 4 月 4 日，当該マルウェアがマルウェアのダウンロードに用いるドメインに対して提

表 4 評価に用いたブラックリスト（実験 3）
Table 4 Blacklist used for evaluation (experiment3).

Fortinet's Web Filter[20]	Phish Tank[24]
MDL[21]	Spamhaus DBL[25]
DNS-BH[22]	SURBL[26]
Open Phish[23]	

表 5 Symantec セキュリティレスポンス [14] で 2015 年から 2016 年までに発見された解析レポート（実験 3）
Table 5 Malware analysis reports found from Symantec Security response [14] during 2015 to 2016 (experiment3).

Symantec検知名	解析レポート作成日
Ransom.TeslaCrypt	2015/2/17
Trojan.Zlob.Q	2016/2/3
Trojan.Huntpos	2016/4/1
Trojan.Phytob	2016/4/21
Ransom.Locky!g6	2016/5/10
Ransom.ODCODC	2016/6/5

表 6 Symantec の解析レポート [14] から抽出したドメインに対して提案手法を適用した結果（実験 3）

Table 6 Detection result against Symantec analysis report using proposal method (experiment3).

Symantec検知名	名前解決するドメイン	DFC値	DU値	UF値	最終投稿日時	提案手法適用日	提案手法による検知	ブラックリスト掲載日
Trojan.Huntpos	3sipiojt.com	1	0	0	2015/9/7	2016/4/4	-	2016/4/28 (DNS-BH)
							-	2016/5/14 (SURBL)
	millionjam.eu	0	1	0	2016/4/1		-	-
	cortykopl.com	0	0	0	-		-	-
	friltopyes.com	0	4	0	2016/4/15		-	2016/5/14 (SURBL)
Trojan.Zlob.Q	deris.info	3	100	0	2016/1/22	2016/4/4	o	2016/5/2 (SURBL)
Ransom.Locky!g6	drlarrybenovitz.com	100	2	1	2016/4/30	2016/5/12	o	-
	holishit.in	100	7	1	2016/4/30		o	2016/6/7 (SURBL)
	grosirkecantikan.com	100	8	2	2016/4/30		o	2016/3/29 (MDL)
Ransom.TeslaCrypt	naturstein-schubert.de	100	1	1	2016/5/7	2016/5/12	o	-
	csskol.org	30	8	3	2016/5/2		o	2016/6/7 (SURBL)
	kknk-shop.dev.onnetdigital.com	100	0	1	2016/4/23		o	2016/6/7 (SURBL)
	casasembargada.com	100	5	2	2016/5/10		o	2016/6/7 (SURBL)
	forms.net.in	100	1	1	2016/5/7		o	2016/6/7 (SURBL)
	mahmutersan.com.tr	100	1	1	2016/5/7		o	2016/6/7 (SURBL)
Trojan.Phytob	tracking.huijang.com	2	3	0	2016/4/26	2016/5/12	-	-
Ransom.ODCODC	inststats.com	0	1	0	2016/5/20	2016/6/10	-	-

案手法を適用した．その結果，当該ドメインを検知することができた．Symantec の解析レポートが公開されたのが 2 月 3 日であり，実際にあるブラックリストに記載されたのはその 88 日後の 5 月 2 日であった．提案手法は 4 月 4 日に適用したため，ブラックリストよりも最大で 28 日間早く検知できた．

Ransom.Locky!g6 [35], [36]：2016 年 5 月 12 日，当該マルウェア感染後に発生する攻撃者のサーバとの通信に用いる 3 種類のドメインに対して提案手法を適用した．その結果，いずれのドメインについても検知することができた．Symantec の解析レポートが公開されたのが 5 月 10 日であり，“holishit.in” については実際にブラックリストに記載されたのはその 28 日後の 6 月 7 日であった．提案手法は 5 月 12 日に適用したため，ブラックリストよりも最大で 26 日間早く検知できた．また，“drlarrybenovitz.com” については観測期間中にブラックリストに記載されなかった．攻撃者は，当ドメインを先のドメインと同じ目的で利用していることから，ブラックリストに記載すべきである．一方，“grosirkecantikan.com” については Symantec の解析レポートが公開されるよりも早くブラックリストに記載されていた．このため，検査対象ドメインの収集方法について 5 章で考察する．

Ransom.TeslaCrypt [37], [38]：2016 年 5 月 12 日，当該マルウェアが C&C サーバとの通信に用いる 6 種類のドメインに対して提案手法を適用した．その結果，いずれのドメインについても検知することができた．Symantec の解析レポートが公開されたのが 2015 年 2 月 17 日であり，5 種類のドメインが実際にブラックリストに記載されたのは 1 年以上経過した 2016 年 6 月 7 日であった．提案手法は 5 月 12 日に適用したため，ブラックリストよりも最大で 26 日間早く検知できた．また，“naturstein-schubert.de” については観測期間中にブラックリストに記載されなかった．攻撃者は，当ドメインを先のドメインと同じ目的で利用していることから，ブラックリストに記載すべきである．

Trojan.Phytob [39], [40]：2016 年 5 月 12 日，当該マルウェアがマルウェアのダウンロードに用いるドメインに対して提案手法を適用した．その結果，当該ドメインについて検知することができなかった．また，観測期間中にブラックリストにも記載されなかった．このようになった理由として，当該マルウェアは Python で書かれていることから，Virus Total にこのドメインに関連するファイルがあまり投稿されていないなどの理由があげられる．

Ransom.ODCODC [41]：2016 年 6 月 10 日，当該マルウェアが C&C サーバとの通信に用いるドメインに対して提案手法を適用した．その結果，当該ドメインについて検知することができなかった．また，観測期間中にブラックリストにも記載されなかった．このようになった理由として，攻撃者は他の種類のランサムウェアを用いてサイバー攻撃を行っており，Virus Total にこのドメインに関連するファイルがあまり投稿されていないなどの理由があげられる．

これらの結果から，提案手法は一部のドメインについて既存のブラックリストよりも早く悪性ドメインを検知することができた．提案手法は，良悪性判定を行った 16 種類のドメインのうち 10 種類のドメインを悪性ドメインと判定した．提案手法で悪性ドメインと判定したがいずれのブラックリストにも記載されていなかったドメインは 2 種類である．これらのドメインについては，今後ブラックリストに記載される可能があるためさらなる調査が必要である．提案手法で悪性ドメインと判定した時点ですでにブラックリストに掲載されていたドメインは 1 種類である．また，提案手法で悪性ドメインと判定インした後にブラックリス

トに記載されたドメインは7種類であり，ブラックリストに掲載されるよりも平均で26日以上早く検知することができた．このように，提案手法は悪性ドメインを迅速に，かつ正確に検知できる可能性があり，ブラックリストの拡張や更新に役立てることができる．

4.4 組織のDNSサーバで収集したドメイン群を用いた速度性能評価（実験4）

実験方法：提案手法の運用形態は様々であるが，実験4ではネットワーク管理者が組織内のユーザが名前解決したドメインの良悪性判定を行うことで，ブラックリストに新たなドメインを追加する状況を仮定する．そこで，連携先組織のDNSサーバで収集した1日分のDNSトラヒックを入力として，検査対象ドメイン群の良悪性判定結果を出力するシステムを実装した．そして，2018年3月1日に連携先組織のDNSサーバで取得したDNSトラヒックを用いて，Virus Totalのドメイン検索にかかる時間や検査対象ドメインの良悪性判定にかかる時間を調査した．なお，連携先組織は数千人規模の組織であり，入力データには組織内のユーザの一部が名前解決した2,652,664ドメイン（63,623ドメインはユニーク）が含まれている．

実験結果：2018年3月1日に連携先組織のDNSサーバで取得した全63,623ドメインに対し，2018年3月16日に提案手法を適用してその速度性能を評価した．Virus Totalに対して前述のユニークな63,623ドメインのうち1万件のドメインを検索したところ，Virus Total検索部でVirus Totalからドメイン検索結果が得られるまでの時間は1ドメインあたり平均1.005秒であった．また，悪性ドメイン判定部で検査対象ドメインの判定に必要な時間は1ドメインあたり平均0.002209秒であった．Virus Totalへのドメイン検索処理は並列化が可能であるため，仮にこの処理を10並列で行うとすると上記の連携先組織の1日分のDNSトラヒックにおける全63,623ドメインを約107分で判定できる試算となる．前日24時間分のDNSトラヒックを分析し，その結果をもとにブラックリストを1日単位で更新する運用形態を想定した場合，提案手法は十分に適用可能である．

5. 考察

実験1では，動的解析により得られた悪性ドメイン群に対して提案手法を適用し，提案手法が悪性ドメインを検知できることを確かめた．実験2では，正規のサービスが運用されている正規ドメイン群に対して提案手法を適用し，提案手法が正規ドメインを誤検知する可能性を確かめた．実験3では，実験1，2で提案手法の評価に用いたドメイン群から機械学習を用いて提案手法のパラメータの組合せを決定し，セキュリティベンダの解析レポートから収集した未知のドメイン群に対して提案手法を適用することで，提案手法が一部のブラックリストよりも早く悪性ドメインを検知できることを確かめた．実験4では，提案手法を実現したシステムを連携先組織の1日分のDNSサーバのトラヒックに適用し，その結果をもとにブラックリストを1日単位で更新する運用形態を想定した場合に十分に適用可能であることを確かめた．しかし，提案手法の実用化にはさらなる検討が必要である．そこで，以下では提案手法の課題について考察する．

提案手法の閾値/条件の設定方法：提案手法は3つのステップから構成され，それぞれで閾値/条件を設定している．Step1の閾値は，値が大きくなると検知できるドメイン数は減少し，見逃しが増えてしまう．反対に，値が小さくなると検知できるドメイン数は増加し，見逃しが少なくなる．Step2の閾値は，値が大きくなると検知できるドメイン数は増加するが，誤検知が増えてしまう．反対に，値が小さくなると検知できるドメイン数は減少するが，誤検知を減らすことができる．同様に，Step3の検体最終投稿日時は，日付の古い検体も含まれる条件に設定した場合，過去にマルウェアが利用していたドメインも検知できる可能性がある．一方，日付の古い検体が含まれない条件に設定した場合，現在マルウェアに利用されているドメインを検知できる可能性が高くなる．このため，閾値/条件の設定は適用先に応じて，見逃しと誤検知の関係を考慮しながら決定すべきである．たとえば，大規模感染しているマルウェアを検知したい場合には，Virus Totalにも多くの検体が投稿されていることが予想されるため，誤検知を少なくする設定で検知できる可能性がある．一方，特定の地域を狙ったマルウェアを検知したい場合には，Virus Totalに投稿されている検体は限られていることが予想されるため，見逃しを少なくする設定で検知できる可能性がある．

提案手法の検知精度の向上：一般に，マルウェアが不正活動に利用するドメインは正規のファイルが名前解決する可能性は低い．また，感染が拡大しているマルウェアであれば，統合型マルウェア検査サービスに投稿されている可能性は高い．提案手法はこれらの特徴に注目して悪性ドメインを検知している点で特徴的である．しかし，提案手法はVirus Totalの検索結果の一部しか利用していない．たとえば，DFC値やDU値は1つ以上のウィルス対策ソフトで検知された検体/URLの数であるが，検知されたウィルス対策ソフトの種類や種類数を利用することで検知精度を向上させることができる可能性がある．同様に，Step2のUF値はウィルス対策ソフトで検知されなかった検体の数（3.1節における⑥⑧⑩の合計）であるが，それぞれの検索結果に対して閾値を設定することで検知精度を向上させることができる可能性がある．また，提案手法では悪性ドメインはDU値やDFC値が大きくなる傾向になることを予想していたが，実験1の悪性ドメインではDU値が7を超えるものがなく，反対に，実験2の正規ドメインではDU

値が上限の 100 になるものが複数存在した．このため，DU 値やドメイン検索で得られる情報についてさらに検討することで検知精度を向上させられる可能性がある．

提案手法の拡張：評価実験では，ネットワーク管理者がブラックリストに新たなドメインを追加する状況を仮定し，Symantec セキュリティレスポンス [14] を入力として悪性ドメインの検知を行った．しかし，インターネット上には多くの解析レポートが存在する．このため，解析レポートを自動的に収集する仕組みを構築することが今後の課題の 1 つである．同様に，評価実験では，組織内の DNS サーバのトラヒックを入力として検査対象ドメインの良悪性判定を行ったが，より長期間のデータを用いて実験を行うことが今後の課題である．一方，提案手法の運用形態は様々である．たとえば，ブラックリストに含まれるドメインの良悪性判定を行い，良性であると判定された場合にはブラックリストから除外することでブラックリストの更新を行うことができる．組織内の DNS サーバのトラヒックを用いてマルウェア感染ホストをネットワークベースで検知することができる．また，論文 [31] で提案されている手法と組み合わせることで，大規模なネットワークへの適用やマルウェア感染をホストベースで検知することができる．提案手法は悪性ドメインを検知する手法であるが，同じ原理で悪性 URL を検知できる可能性がある．そのため，今後は提案手法の拡張を目指す．

統合型マルウェア検査サービスの種類：提案手法は統合型マルウェア検査サービスの一種である Virus Total を用いて悪性ドメインの検知を行うが，その他のサービスを用いて悪性ドメインを検知できる可能性がある．また，複数のサービスを組み合わせて悪性判定を行うこともできる．このため，今後はより多くのサービスで提案手法を実現することを検討する．

提案手法は統合型マルウェア検査サービス Virus Total を用いてブラックリストの拡張や更新を行うことを目的としている．以下では，提案手法の限界について考察する．

Virus Total に蓄積されている解析結果：提案手法は Virus Total を用いて悪性ドメインの検知を行っているが，つねに Virus Total に解析結果が存在するとは限らない．たとえば，標的型攻撃に用いられるマルウェアの場合，検体収集が困難な場合がある．また，提案手法をマルウェアが流行する前に適用した場合，Virus Total に十分な検体数が投稿されておらず，悪性ドメインとして検知できない可能性がある．しかし，ブラックリストは既知の脅威に対して有効に働く手法である．このため，提案手法は Virus Total に十分な解析結果が蓄積されている既知のマルウェアの検知を前提としている．ただし，4.3 節の実験 3 の結果から，最新のマルウェアが名前解決するドメインであっても提案手法は既知のブラックリストよりも早く悪性ドメインを検知できる可能性がある．提案手法と既知のブラックリストに対する，同時期に収集したドメインを用いた悪性ドメインの検知結果の比較についてはさらなる調査が必要であり，今後の課題とする．

提案手法の誤検知や見逃し：Virus Total に解析結果が存在する場合でも正規ドメインを誤検知する可能性がある．たとえば，ある正規ドメインを名前解決する正規ファイルが投稿されておらず，正規ドメインを名前解決するマルウェア検体のみが投稿されていた場合，当該ドメインを悪性ドメインとして検知してしまう．実際に，実験 2 に用いた正規ドメイン群のうち “north-america.pool.ntp.org” は，DFC 値 5，UF 値は存在しないという結果になった．一方，正規の目的で利用されていたサービスが攻撃者に乗っ取られ，現在はマルウェアの不正活動に利用されている場合，提案手法では当該サービスに対応するドメインを検知できないことが予想される．提案手法で検知できるドメインには限りがあるため，ホワイトリストなどにより正規ドメインを除外する方法や，Step3 の条件をある期間に設定する方法，あるいは他のブラックリストと組み合わせることでこのような攻撃を防ぐことが望ましい．

提案手法による検知の回避：攻撃者は提案手法による検知を回避するために，悪性ドメインを名前解決するが不正活動を行わず，ウィルス対策ソフトで検知されない検体を統合型マルウェア検査サービスに投稿することができる．この場合，Step2 の検査対象ドメインを名前解決する正規ファイル数が閾値を上回ることで提案手法による検知を回避することができる．また，正規サービスをプロキシとして利用することで不正活動を行うマルウェアのドメインを検知することは難しい [18]．同様に，DGA などにより大量のドメインを名前解決するマルウェアが Virus Total に多数投稿されていた場合，ブラックリストに記載されるドメインの増加につながってしまう．しかし，これらのマルウェアはその特徴から別の方法で検知される可能性がある．このため，ブラックリストによる検知だけでなく，その他のセキュリティ対策技術も重要である．

6. 関連研究

DNS サーバのトラヒックには，正規ユーザが名前解決したドメインのほかに，マルウェア感染ホストが名前解決した悪性ドメインが含まれる場合がある．このため，実 DNS サーバのトラヒックから悪性ドメインを検知する研究が活発に行われている．たとえば，論文 [1], [2] では，複数のリカーシブ DNS サーバのトラヒックから悪性ドメインに共通して見られる特徴を用いて悪性ドメインを検知する手法が提案されている．また，論文 [3] では，権威 DNS サーバの DNS トラヒックから悪性ドメインが名前解決されるパターンを用いて悪性ドメインを検知する手法が提案されている．同様に，DNS キャッシュサーバのトラヒックから悪性ドメインを検知する手法が多数提案されている．論

文 [9] では，悪性ドメインが名前解決される前後で名前解決されるドメインやその順番を用いて悪性ドメインを検知する手法が提案されている．論文 [10] では，ベイズ推定法を用いて大量メール拡散型ワームに感染したホストを検知する手法が提案されている．論文 [11] では，ボットに感染したホストは同一のドメインを名前解決するという特徴を用いて悪性ドメインを検知する手法が提案されている．論文 [12] では，ボットに感染したホストは協調して動作を行うという特徴を用いて悪性ドメインを検知する手法が提案されている．DNS サーバのトラヒックを用いて作成したブラックリストはマルウェア感染ホストの検知に有効であるが，DNS サーバのトラヒックの収集にはコストがかかる．また，提案手法の適用時期，あるいは国や地域に応じて名前解決されるドメインに違いが存在することが予想される．このため，ブラックリストを更新する必要がある．

一方，ブラックリストを拡張する研究が行われている．論文 [4] では，DNS キャッシュサーバのトラヒックと既知の悪性ドメインの共起関係を用いて，ブラックリストを拡張する手法が提案されている．論文 [5] では，DNS キャッシュサーバのトラヒックと既知の悪性ドメインから，DNS クエリグラフを作成することでブラックリストの精度を向上させる手法が提案されている．論文 [6] では，既知の悪性ドメインの登録者情報やネームサーバ情報を用いてブラックリストを拡張する手法が提案されている．先行研究は，ブラックリストに記載されているマルウェアと類似したマルウェアを検知するのに有効であるが，ブラックリストに掲載されていない未知のマルウェアに対しては有効に働くとは限らない．このため，さらなる対策が必要である．

未知のマルウェアが名前解決する悪性ドメインを検知する方法の 1 つに，動的解析結果を用いる方法がある．論文 [7] では，ウィルス対策ソフトによる検知名が同一である検体群を動的解析し，共通して名前解決するドメインを悪性ドメインとして検知する手法が提案されている．論文 [8] では，擬似 DNS サーバ，IRC サーバを用いてマルウェア検体を動的解析した際に得られるドメインと，静的解析で得られるドメインの近接関係から悪性ドメインを検知する手法が提案されている．しかし，マルウェアの数は爆発的に増加しており，すべてのマルウェア検体を動的解析することは難しい．このため，迅速に悪性ドメインを検知する手法が求められている．

本研究では上記関連研究の状況をふまえ，動的解析の結果が蓄積されている統合型マルウェア解析サービス Virus Total を用いて悪性ドメインを検知する手法を提案した．

7. まとめと今後の課題

統合型マルウェア検査サービス Virus Total を用いて，マルウェアが不正活動に利用するドメインを検知する手法を提案した．評価実験の結果，閾値/条件を適切な設定にすることで悪性ドメインを検知できることを確かめた．また，既知のブラックリストに記載されるよりも平均 26 日早く悪性ドメインを検知することができた．提案手法は，マルウェアが名前解決するドメインの中から悪性ドメインを検知できる点で有用であり，ブラックリストの拡張に役立てることができる．今後は，Symantec などのセキュリティベンダの解析レポートを参考に，マルウェア検体が名前解決するドメインを自動的に収集することで，より多くのドメインに対して提案手法を適用することを目指す．

謝辞 本研究の一部は，文部科学省国立大学改革強化推進事業の支援を受けて行われた．加えて，本研究の一部は，国立研究開発法人情報通信研究機構（NICT）の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」によって行われた．

参考文献

[1] Antonakakis, M., Perdisci, R., Dagon, D., Lee, W. and Feamster, N.: Building a Dynamic Reputation System for DNS, *Proc. 19th USENIX Security Symposium* (*USENIX Security '10*) (2010).
[2] Bilge, L., Kirda, E., Kruegel, C. and Balduzzi, M.: Exposure: Finding malicious domains using passive dns analysis, *Proc. NDSS* (*NDSS '11*) (2011).
[3] Antonakakis, M., Perdisci, R., Lee, W., Vasiloglou, N. and Dagon, D.: Detectin Malware Domains at the Upper DNS Hierarchy, *Proc. 21th USENIX Security Symposium* (*USENIX Security '12*) (2012).
[4] Sato, K., Ishibashi, K., Toyono, T., Hasegawa, H. and Yoshino, H.: Extending Black Domain Name List by Using Co-occurrence Relation between DNS Queries, *IEICE Trans. Commun.*, Vol.E95-B, No.3, pp.794–802 (2012).
[5] Ishibashi, K., Toyono, T. and Iwamura, M.: Improving accuracy of black domain list by using DNS query graph, *Proc. 1st Internet Workshop on Information Network Design*, Fukuoka, Japan (Dec. 2008).
[6] Felegyhazi, M., Kreibich, C. and Paxson, V.: On the Potential of Proactive Domain Blacklisting, *Proc. 3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats* (2010).
[7] Stalmans, E. and Irwian, B.: A Framework for DNS Based Detection and Mitigation of Malware Infections on a Network, *Information Security South Africa Conference* (2011).
[8] 朝長秀誠，田中英彦：Botnet の命令サーバドメインネームを用いた Bot 感染検出手法，情報処理学会研究報告（CSEC2006），pp.13–18 (2006).
[9] Lee, J., Kwon, J., Shin, H.J. and Lee, H.: Tracking multiple c&c botnets by analyzing dns traffic, *The 6th IEEE Workshop on Secure Network Protocols* (*NPSec*) (2010).
[10] Ishibashi, K., Toyono, T., Toyama, K., Ishino, M., Ohshima, H. and Mizukoshi, I.: Detecting Mass-Mailing Worm Infected Hosts by Mining DNS Traffic Data, *Proc. 2005 ACM SIGCOMM Workshop* (2005).
[11] Choi, H., Lee, H. and Kim, H.: BotGAD: Detecting botnets by capturing group activities in network traffic, *Proc. 4th International Conference on Communication System Software and Middleware* (*COMSWARE 2009*) (2009).

[12] Choi, H., Lee, H. and Kim, H.: Bot detection by monitoring group activities in DNS traffic, *Computer and Information Technology* (2007).
[13] Virus Total, available from ⟨https://www.virustotal.com/⟩.
[14] Symantec：セキュリティレスポンス，入手先 ⟨https://www.symantec.com/ja/jp/security_response/⟩.
[15] Jotti's Malware Scan, available from ⟨https://virusscan.jotti.org/⟩.
[16] Malwr, available from ⟨https://malwr.com/⟩.
[17] W32.Morto.B, available from ⟨https://www.symantec.com/ja/jp/security_response/writeup.jsp?docid=2012-071013-3812-99⟩.
[18] Backdoor.Makadocs, available from ⟨https://www.symantec.com/ja/jp/security_response/writeup.jsp?docid=2012-111609-4148-99&tabid=2⟩.
[19] Amazon Alexa, available from ⟨http://alexa.amazon.com/spa/index.html⟩.
[20] Fortinet's Web Filter, available from ⟨http://fortiguard.com/webfilter⟩.
[21] Malwre Domain List, available from ⟨https://www.malwaredomainlist.com/⟩.
[22] DNS-BH, available from ⟨http://www.malwaredomains.com/⟩.
[23] Open pish, available from ⟨https://openphish.com⟩.
[24] Phish Tank, available from ⟨https://www.phishtank.com/⟩.
[25] Spamhaus DBL, available from ⟨https://www.spamhaus.org/⟩.
[26] SURBL: Lookup, available from ⟨http://www.surbl.org/surbl-analysis⟩.
[27] Mariano, G., Davide, C., Leyla, B., Andrea, L. and Davide, B.: Needles in a Haystack: Mining Information from Public Dynamic Analysis, Sandboxes for Malware Intelligence, *Proc. 24th USENIX Security Symposium* (*USENIX Security '15*) (2015).
[28] Yokoyama, A., Ishii, K., Tanabe, R., Papa, Y.M., Yoshioka, K., Matsumoto, T., Kasama, T., Inoue, D., Brengel, M., Backes, M. and Rossow, C.: SANDPRINT: Fingerprinting Malware Sandboxes to Provide Intelligence for Sandbox Evasion, *Proc. Research in Attacks, Intrusions and Defenses* (*RAID '16*), Lecture Notes in Computer Science (2016).
[29] WannacCry：情報まとめ, 入手先 ⟨https://blog.kaspersky.co.jp/wannacry-faq-what-you-need-to-know-today/15594/⟩.
[30] WORM_KELIHOS.SM, available from ⟨https://www.trendmicro.com/vinfo/jp/threat-encyclopedia/print/malware/WORM_KELIHOS.SM⟩.
[31] 森島周太，中野弘樹，藤原礼征，吉岡克成，松本 勉：多数のユーザの Web アクセスログから効率的に悪性サイトを抽出する手法，情報処理学会コンピュータセキュリティシンポジウム 2017 (2017).
[32] Trojan.Huntpos, available from ⟨https://www.symantec.com/security_response//writeup.jsp?docid=2016-040112-5251-99⟩.
[33] TREASUREHUNT: A Custom POS Malware Tool, available from ⟨https://www.fireeye.com/blog/threat-research/2016/03/treasurehunt_a_cust.html⟩.
[34] Trojan.Zlob.Q, available from ⟨https://www.symantec.com/security_response/writeup.jsp?docid=2016-020300-4629-99⟩.
[35] Ransom.Locky, available from ⟨https://www.symantec.com/security_response/writeup.jsp?docid=2016-021706-1402-99⟩.
[36] Malware-Traffic-Analysis-Net, available from ⟨http://www.malware-traffic-analysis.net/2016/03/29/index.html⟩.
[37] Ransom.TeslaCrypt, available from ⟨https://www.symantec.com/security_response/writeup.jsp?docid=2015-030201-5710-99&tabid=2⟩.
[38] A Close Look at TeslaCrypt 3.0 Ransomware, available from ⟨https://blog.threattrack.com/close-look-teslacrypt-3-0-ransomware/⟩.
[39] Trojan.Phytob, available from ⟨https://www.symantec.com/security_response//writeup.jsp?docid=2016-042121-3315-99⟩.
[40] Python-Based PWOBot Targets European Organizations, available from ⟨https://researchcenter.paloaltonetworks.com/2016/04/unit42-python-based-pwobot-targets-european-organizations/⟩.
[41] Ransom.ODCODC, available from ⟨https://www.symantec.com/security_response/writeup.jsp?docid=2016-030408-0817-99&tabid=2⟩.

田辺 瑠偉 （正会員）

2017 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期修了．博士（情報学）．同年 4 月より横浜国立大学大学院環境情報研究院で産学官連携研究員として勤務．2018 年 4 月より横浜国立大学先端科学高等研究院特任教員（助教）．情報セキュリティ，特にネットワークセキュリティの研究に従事．2017 年情報処理学会山下記念研究賞受賞．

森 博志 （学生会員）

2013 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了．修士（工学）．同年 4 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期に進学．ネットワークセキュリティに関する研究に従事．

原田 耕也

2017 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了．修士（工学）．同年 4 月富士通株式会社入社．在学中，情報セキュリティに関する研究に従事．

吉岡 克成 （正会員）

2005 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期修了．博士（工学）．同年 4 月独立行政法人情報通信研究機構で研究員として勤務．2007 年 12 月より横浜国立大学学際プロジェクト研究センター特任教員（助教）．2011 年 4 月横浜国立大学大学院環境情報研究院准教授．マルウェア解析やネットワーク攻撃観測・検知等のネットワークセキュリティの研究に従事．2009 年文部科学大臣表彰・科学技術賞（研究部門），2016 年産学官連携功労者表彰総務大臣賞，2017 年情報セキュリティ文化賞をそれぞれ受賞．

松本 勉 （正会員）

1986 年 3 月東京大学大学院工学系研究科電子工学専攻博士課程修了．博士（工学）．同年 4 月横浜国立大学講師．2001 年 4 月より同大学大学院環境情報研究員教授．2014 年 12 月より同大学先端科学高等研究員（IAS-YNU）情報物理セキュリティ研究ユニットリーダーを兼務．ネットワーク・ソフトウェア・ハードウェアセキュリティ，暗号，耐タンパー技術，生体認証，人工物メトリクス等の「情報・物理セキュリティ」の研究教育に 1981 年より従事．1982 年にオープンな学術的暗号研究を目指した「明るい暗号研究会」を 4 名で創設．2005～2010 年国際暗号学会 IACR 理事．1994 年第 32 回電子情報通信学会業績賞，2006 年第 5 回ドコモ・モバイル・サイエンス賞，2008 年第 4 回情報セキュリティ文化賞，2010 年文部科学大臣表彰・科学技術賞（研究部門）各受賞．