情報処理学会論文誌 Vol.62 No.3 829–837 (Mar. 2021)

辞書に基づくDGAマルウェアに起因した悪性ドメインの判別

佐藤 彰洋[1,a)] 福田 豊[1] 井上 純一[1] 中村 豊[1]

受付日 2020年6月12日, 採録日 2020年12月1日

概要：高度なDGAマルウェアの出現により，コンピュータネットワークは深刻な脅威に直面している．このDGAマルウェアは，それ自体が有する辞書の単語を連結することで，人為的に生成したものと判別が困難な悪性ドメインを機械的に生成する．本稿では，ドメインの文字列を構成する単語間の関係性を考慮することで，辞書に基づくDGAマルウェアにより生成された悪性ドメインの判別を試みる．また実験を通じて，提案手法が0.9977の再現率と0.9869の適合率で悪性ドメインを判別可能であることを確認した．この結果から，ネットワークに内在する多様なマルウェアへの迅速な対処が可能となるため，ネットワークの運用において安全性の向上が期待できる．

キーワード：マルウェア，辞書に基づくドメイン生成，ドメイン名，ネットワークセキュリティ

An Approach for Identifying Malicious Domain Names Caused by Dictionary-based DGA Malware

Akihiro Satoh[1,a)] Yutaka Fukuda[1] Jun'ichi Inoue[1] Yutaka Nakamura[1]

Received: June 12, 2020, Accepted: December 1, 2020

Abstract: Computer networks are facing serious threats by the emergence of sophisticated new DGA malware. This type of malware family has its own dictionary, from which it concatenates words to dynamically generate malicious domain names that are difficult to distinguish from human-generated domain names. In this paper, we propose an approach that focuses on relations among the words that constitute the character string of each domain name for identifying malicious domain names generated by Dictionary-based DGA malware. Our evaluation demonstrates that this approach has high identification ability, with a recall of 0.9977 and a precision of 0.9869. By enabling to address various malware encroachments, our approach contributes to dramatically improving network security.

Keywords: malware, dictionary-based domain generation, domain name, network security

1. はじめに

マルウェアはインターネットにおける重大な脅威の1つである．サイバー犯罪者は，C&C（Command-and-Control Server）を介してマルウェアに感染した端末を操作することで，機密情報窃取，フィッシング詐欺，標的型攻撃などの悪意ある活動を試みる．米McAfee社の報告によると，亜種を含めて約30万のマルウェアが日々誕生しており，それによる世界の総損失額は年間6,000億ドルを超える[1]．そのため，マルウェアに抗する技術の確立が急務である．

マルウェアによる被害の抑止のため，管理者は自身のネットワークに内在する感染端末を迅速に排除することが求められる．一方，多くのマルウェアには，検出を回避するための機能としてDGA（Domain Generation Algorithm）が実装されている[2]．DGAとは，C&Cのドメインを頻繁に変更することで，マルウェアからC&Cへ向けた通信であるコールバックを隠蔽するための仕組みである．具体

[1] 九州工業大学
Kyushu Institute of Technology, Kitakyushu, Fukuoka 804–8550, Japan
[a)] satoh@isc.kyutech.ac.jp

的には，マルウェアはDGAに基づいて多数のドメインを機械的に生成した後，それらドメインに対して名前解決を試みる．その結果，正しい応答を返したドメインをC&Cのものと見なし，そのドメインとの間で通信を確立する．

いくつかの研究では，良性と悪性のドメイン文字列の差異からDGAマルウェアのコールバック通信を検出している [3], [4], [5]．これは，登録済みのドメインとの衝突を避けるため，悪性ドメインが無意味な文字列からなることに起因する．それに対して，これまでの検出を無効化する高度なDGAマルウェアが出現している．このマルウェアは，それ自身が有する辞書の単語を連結することで，人為的なものと判別が困難なドメインを機械的に生成する [6]．

本稿では，DNS（Domain Name System）に対する膨大な数の名前解決要求から，辞書に基づくDGAマルウェアにより生成されたドメインの判別を試みる．DNSに着目した理由は，マルウェアによる通信に先んじて必ず名前解決が生じること，暗号化による通信内容の隠蔽が困難であることに起因する．我々は，良性ドメインと悪性ドメインでは文字列で頻出する単語や共起する単語に明確な差異が現れるという仮定をふまえ，ドメインの文字列を構成する単語の関係性に基づく悪性ドメイン判別手法を提案する．これにより，文字列のみからドメインの良性と悪性の判別が可能となる．また実験を通じて，提案手法が0.9977の再現率と0.9869の適合率で悪性ドメインを判別可能であることを確認した．すなわち，悪性ドメインの名前解決を予兆として辞書に基づくDGAマルウェアを高精度で検出できることを示した．この結果から，ネットワークに内在する多様なマルウェアへの迅速な対処が可能となるため，ネットワークの運用において安全性の向上が期待できる．

本稿の構成は次のとおりである．まず，2章で既存研究とその問題点を整理する．3章でドメイン文字列を構成する単語の関係性に基づく悪性ドメイン判別手法を提案した後，4章で提案手法の有効性を議論する．最後に5章で本研究の貢献と課題をまとめる．

2. 関連技術

本章では，マルウェアを中心とした関連技術について述べる．2.1節でDGAマルウェアの詳細について説明した後，2.2節で既存研究とその問題点を整理する．

2.1 DGAマルウェア

ConfickerやGameOver Zeus，Torpigなど，世界で深刻な被害を齎したマルウェアには，その機能の一部としてDGAが実装されている．また，それらマルウェアを広範囲に拡散するため，ウェブページやウェブ広告への不正コードの埋め込みが観測されている [7], [8]．

図1にDGAマルウェアによるコールバックの概要を示す．ここで，図中の**Q1**と**Q2**で示す通信はマルウェアからRDNS（Recursive DNS Server）[*1]に対する名前解決を，**R1**と**R2**はその応答を意味する．また，C&Cのものとして，DGAにより生成されたドメインが事前に登録されているものとする．まず，マルウェアはDGAに基づいて複数のドメインを機械的に生成し，それらドメインを自身の属するネットワーク内のRDNSに問い合わせる．RDNSは，ドメインが登録済みであった場合，そのドメインに対応付けられたアドレスを，ドメインが未登録であった場合，エラーメッセージとしてNXDOMAIN（Non-Existing Domain）をそれぞれ応答する．最終的に，マルウェアは正しい応答があったドメインをC&Cのものと見なし，そのドメインに対してコールバックを試みる．

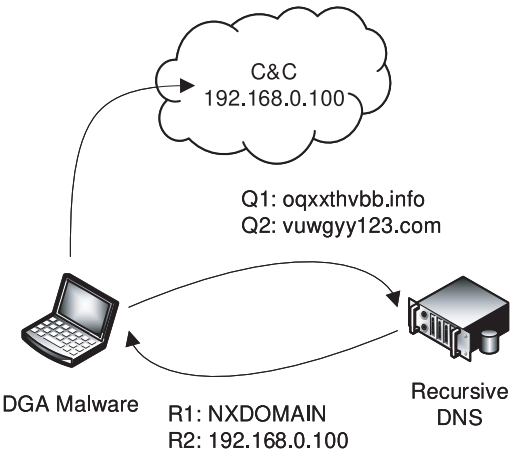


図1 DGAマルウェアとC&Cとのコールバック通信
Fig. 1 Callback communication between DGA malware and C&C.

良性ドメインと比較して，一般的なDGAマルウェアによるドメインは文字列に明確な差異が見てとれる．たとえば，`1ygx14u1vnf8hb1twhv8619h8ygr.net`や`cipu0wdgsnq9u8st8m1lym0hq.com`など，GameOver Zeusのドメインは文字長が15から30までのランダムな英数字からなる [9]．これは，すでに登録済みのドメインとの衝突を避けることが狙いである．一方，辞書に基づくDGAマルウェアは，自身が有する辞書の単語を連結することで，`accelerateaccountant.in.net`や`accelerateactress.in.net`など，人為的に生成したものと判別が困難なドメインを機械的に生成する [6]．

DGAの目的は，マルウェアとC&Cの間に可用性の高い通信経路を確立することにある．具体的には，C&Cのドメインを変更することで，ブラックリストに基づく通信の遮断を容易に回避することが可能となる．加えて，ネットワーク内から外へ向けた通信は宛先が多岐にわたるためコールバックの発見が困難となること，アドレス変換やファイアウォールにより通信を制限されないことがあげら

[*1] RFC8499で定義されるRecursive Resolverを指す．Full-Service Resolverとは，キャッシュ機能の有無により区別される．

れる．ここで留意すべきは，マルウェアと C&C とで同一 DGA を用いることで，ドメインの変更にいっさいの情報交換を必要としない点である．

2.2 既存研究と問題点

ブラックリストの高度化については継続的な研究が行われており，現在もネットワークにおける脅威防御戦略の中核を成している．Soldo らは，複数の参加者から提供される過去の攻撃ログに基づいて，ブラックリストを更新する方法を提案している [10]．また，Freudiger らは，P2P の技術を応用することで，機密性を担保した攻撃ログの共有を実現している [11]．それに対して，DGA マルウェアは，C&C のドメインを頻繁に変更することにより，ブラックリストに基づく通信の遮断を回避する機能を有している．

マルウェアと C&C との通信を補足するために，パケットのペイロードを参照する DPI（Deep Packet Inspection）が用いられてきた．Gu らは，DPI に基づく受動的なネットワーク監視システムとして BotHunter を実装した [12]．BotHunter は，マルウェアの一般的な挙動をモデル化して，それと関連の強い通信を感染の根拠とする．また，DPI の性能改善に向けた取り組みなどが報告されている [13]．一方，2017 年の時点でインターネットにおける暗号化通信の割合は 50%を超えること，それとあわせて約 70%のマルウェアが通信を暗号化することが確認されている [14]．この暗号化の普及に反比例して，DPI を適用可能な通信は極わずかなもののみとなっている．ゆえに，マルウェアの検出のための情報源として，暗号化の影響を受けない DNS の名前解決が注目されている．

Rahbarinia らは，DNS の名前解決において既知の悪性ドメインと高確率で共起するドメインから未知の悪性ドメインを発見する Segugio を開発した [15]．Segugio は次の直感的知見，(1) 同一マルウェアファミリに感染した端末は，同一悪性ドメイン群と通信する傾向にあること，(2) 未感染の端末は，悪性ドメインと通信することがないことに基づいている．一方，DGA マルウェアにおいては，コールバック通信に生存時間が極端に短い一時的な悪性ドメインを用いるため，その一時的な悪性ドメインと共起するドメインは存在しない．ゆえに，このシステムは DGA マルウェアの通信に対して効果をなし得ない．

Berger らは，名前解決におけるアドレスとドメインの関係の変化を継続的に学習する DNSMap を構築した [16]．DNSMap は，C&C のアドレスが複数のドメインに，そのドメインが複数のアドレスに対応付けられること，それらの対応関係が時間経過にともない急速な変化を示すことに着眼している．一方，Wang らは，名前解決の挙動と分布特性に基づいて DGA マルウェアの検出する DBod を実装した [17]．その検出は，同一 DGA により生成された候補ドメイン群に対して接続を試みるため，同一マルウェアファミリに感染した端末による名前解決が特定の期間中に高い類似性を示すことに基づいている．これらのシステムは広範囲にわたる DNS トラフィックの観測を必要とするため，その適用は ISP（Internet Service Provider）などの大規模なネットワークに限定される．

これまでに，DGA マルウェアが生成するドメインについての報告がなされている [18]．それをふまえ，文字列の特徴のみを用いたドメインの判別が試みられてきた．Truong らは，ドメイン文字列のみから良性・悪性を判別する手法を提案した [4]．この手法は，教師あり機械学習とバイグラムモデリングによりドメインにおける頻出文字パターンを学習する．Anderson らは，深層学習を用いた文字レベルのモデリングにより，その手法を拡張した [5]．加えて，Vinayakumar らにより，多様な機械学習と深層学習を用いた判別精度の比較が示されている [19]．これらの手法は，悪性ドメインを生成するためのルールに識別可能な偏りが存在することに基づいている．しかしながら，単語を考慮しない文字レベルのモデリングでは，辞書に基づく DGA マルウェアの検出において十分な精度が期待できない．

本稿と同様に，辞書に基づく DGA マルウェアに焦点を当てたものとして Pereira らの取組みがある [20]．この手法は，DGA マルウェアの有する辞書を，それが生成したドメインから再構築することに主眼を置いている．一方，良性と悪性の判別は非常に単純で，その辞書にドメイン文字列をなす単語が閾値以上含まれるか否かに基づいている．ゆえに，多岐にわたる DGA に対して精度を維持するためには，その判別の仕組みの高度化が必須である．

3. 提案

本稿では，DNS に対する膨大な数の名前解決から，辞書に基づく DGA マルウェアにより生成されたドメインの判別を試みる．**表 1** に，辞書に基づく DGA マルウェアが生成したドメインの例を示す．これらドメインはコールバック先の C&C に依存するため，その生成には各マルウェアで異なる辞書とアルゴリズムが用いられることとなる．たとえば，文献 [21] と [22] で報告されている Gozi と Matsnu の辞書を比較したところ，単語数は 975 と 1,391 であり，

表 1 辞書に基づく DGA マルウェアと悪性ドメインの例
Table 1 Examples of domain names generated by dictionary-based DGA malware.

Banjori	earnestnessbiophysicalohax.com
	pbmnestnessbiophysicalohax.com
Gozi	williamseasily.com
	printingthatlabel.com
Matsnu	shoulderracerecognizeblue.com
	emergencyadaptselectdoubt.com
Suppobox	windowtherefore.net
	severadifference.net

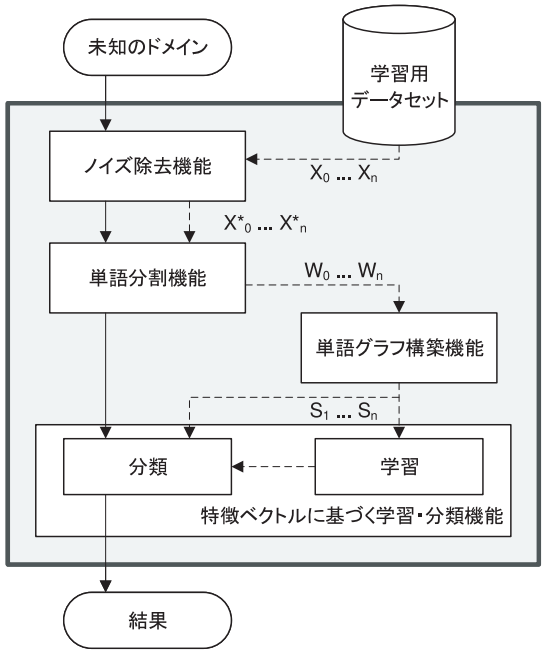


図 2 ドメインの文字列を構成する単語間の関係性に基づく悪性ドメイン判別手法の概要

Fig. 2 Overview of the proposed approach for detecting dictionary-based DGA malware based on word relations in domain names.

それら辞書間の重複は 240 のみであった[*2]．それに加え，マルウェアが機械的に生成したドメインは特定の辞書の単語から構成されるため，人為的に生成したドメインと比較して使用される単語に大きな偏りが生じると予想される．そのドメイン文字列で頻出する単語や共起する単語に明確な差異が現れるという仮定をふまえ，我々はドメインの文字列を構成する単語の関係性に基づく悪性ドメイン判別手法を提案する．本手法の特徴は，(1) 文字列のみからドメインの良性と悪性を判別すること，(2) ドメイン文字列をなす単語群の関係を一般的なグラフ理論における重み付き無向グラフで表現すること，(3) グラフにおける各頂点の中心性により各単語の重要性を測ることにある．最終的に，その指標に基づく特徴ベクトルに機械学習アルゴリズムを適用することで，ドメインの良性と悪性を判別する．

図 2 に提案手法の概要を示す．本手法は，(1) ノイズ除去機能，(2) 単語分割機能，(3) 単語グラフ構築機能，(4) 特徴ベクトルに基づく学習・分類機能により構成される．3.1 節で学習用データセットについて，それ以降の節で各機能の詳細について述べる．なお，本手法では，2.1 節で述べた DGA マルウェアにおけるコールバック通信の特徴をふまえ，良性と悪性を判別するドメインの数を大きく絞り込んでいる．具体的には，DNS の名前解決要求がドメイン名からアドレスへの変換であること，その結果として NXDOMAIN を応答することの，両条件を満たすもののみを未知のドメインと学習用データセットとして用いる．

[*2] 文献 [21] で記載のある Ursnif は Gozi の別称である．

3.1 学習用データセット

提案手法において，特定のドメインのみを含む数種のデータセット $X_i, i \in 0 \cdots n$ を学習のために用いることとなる．ここで，X_0 に属するのは良性ドメイン，X_1 から X_n に属するのは n 種の DGA マルウェアに起因する悪性ドメインとする．特筆すべきは，各ドメインをプライマリドメインまで短縮する点である．プライマリドメインは登録可能な最高レベルのサブドメインである [23]．具体的には，`www.ipsj.or.jp` と `smtp.isc.kyutech.ac.jp` のプライマリドメインは，それぞれ `ipsj.or.jp` と `kyutech.ac.jp` となる．

3.2 ノイズ除去機能

本機能は，まずホワイトリストと合致したものを自明な良性ドメインと判断する．ホワイトリストに含まれるのは，`kyutech.jp` や `kyutech.ac.jp` などの自組織が有するドメイン，`uribl.com` や `dnswl.org` などの DNSBL・DNSWL（DNS Blacklists and Whitelists）に代表される特定のサービスが利用するドメイン，`trendmicro.com` や `barracudacentral.org` などのセキュリティ製品に関連付けられたドメインである．それに加え，DNS の仕様に違反する文字列からなるドメインを [24]，入力ミスや設定ミスに起因するノイズと見なす．なお，本手法は国際化ドメインに非対応であるため [25]，ここであわせて `xn--` を接頭辞とするドメインを除外する点を留意されたい．その残りのデータセット $X_i^*, i \in 0 \cdots n$ は，さらなる分析のために次の機能に渡される．

3.3 単語分割機能

本機能は，辞書 $\mathbb{D}$ に基づいてドメイン x のプライマリレベルの文字列を単語群 w に分割する．辞書 $\mathbb{D}$ に含まれるのは，クローリングで作成したコーパスと英語辞書である．その分割を次式で示す．

$$\mathcal{F}(x) = \arg\max_{w \in \mathbb{W}(x)} \frac{1}{m} \prod_{j=1}^{m} \mathcal{P}(w_j)$$

$$\mathcal{P}(w_j) = \begin{cases} 1 & (w_j \in \mathbb{D}) \\ 1/|\mathbb{D}|^{|w_j|} & (w_j \notin \mathbb{D}) \end{cases}$$

ここで，$\mathbb{W}(x)$ はドメイン x のプライマリレベルの文字列における全分割候補の集合，w は単語 $w_1, \cdots, w_j, \cdots, w_m$ からなる候補の単語群，$|w_j|$ は単語 w_j の文字長，$|\mathbb{D}|$ は辞書 $\mathbb{D}$ の総単語数をそれぞれ意味する．また，$\mathcal{P}(w_j)$ は，単語 w_j が辞書 $\mathbb{D}$ に含まれるか否かに基づいて，単語 w_j の選択率を導出する関数である．この結果は，(1) 各単語の文字長が最大かつ単語数が最小になること，(2) 極端な選択率の差により辞書に含まれる単語を優先することの，両条件を満たす分割となる．なお，文字列 `localdomain` の分割が `local` と `domain` の 2 単語からなる場合，その結

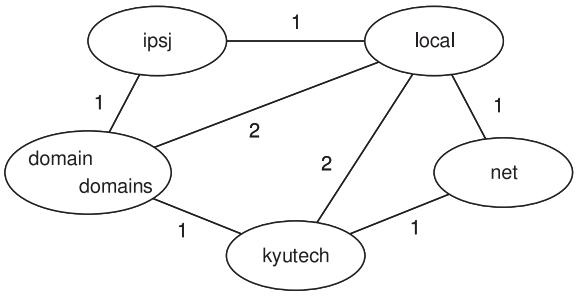


図 3 ドメイン文字列をなす単語群を用いた単語グラフの構築例
Fig. 3 Example of constructing a word graph using a set of word groups.

果を {local, domain} と表記することとする．最終的に，本機能は学習用データセット X_i^* におけるドメインから，そのプライマリレベルの文字列をなす単語群の集合 W_i を出力する．

3.4 単語グラフ構築機能

本機能は，単語グラフを用いることで，プライマリレベルの文字列をなす単語群の集合 W_i の関係性を示す．単語グラフ G_i は，頂点と辺の集合からなる重み付き無向グラフである．頂点は集合 W_i における単語，辺の重みは単一ドメインにおいて 2 つの単語が共起する頻度を意味する．ここで留意すべきは，単語の活用を考慮するため，文字列が類似した単語を同一頂点に集約する点である．その類似性の導出には編集距離比を，集約には閾値 th の重心法に基づく階層型クラスタリングを採用した [26]．**図 3** に，{kyutech, local, domain}，{local, kyutech, net}，{ipsj, domains, local} の単語群を用いた単語グラフの構築例を示す．この単語グラフおいて，domain と domains は文字列の類似性から同一頂点に集約されることになる．

次いで，単語グラフにおける処理としてグラフの結合を，指標として単語の重要性を定義する．結合は，単語グラフ G_i と G_j を構築するために用いた単語群の集合 W_i と W_j の和から，単語グラフ $G_{i:j}$ を再構築する処理である．重要性は，単語グラフにおける中心性に関連付けられた指標である．任意の単語群 w の重要性を次式で示す．

$$\mathcal{S}_i(w) = \sum\nolimits_{w_j \in w} |w_j| \big(\mathcal{C}_{0:i}(w_j) - \mathcal{C}_0(w_j) \big)$$

ここで，$|w_j|$ は単語 w_j の文字長を意味する．また，$\mathcal{C}_0(w_j)$ と $\mathcal{C}_{0:i}(w_j)$ は，単語グラフ G_0 と $G_{0:i}$ における単語 w_j の中心性を導出する関数である．機械的に生成した悪性ドメインと人為的に生成した良性ドメインでは，その文字列で頻出する単語や共起する単語に明確な差異が現れるがゆえに，良性と悪性の判別に効果的な単語は単語グラフにおいて中心的な役割を担うこととなる．その観点をふまえ，良性データセットから構築した単語グラフ G_0 を基準として，それと悪性データセットから構築した単語グラフ G_i の結合による中心性の変化量を単語 w_j の重要性とした．その中心性の導出には，単語グラフが無向かつ非連結になることを勘案して PageRank を採用した [27]．

3.5 特徴ベクトルに基づく学習・分類機能

本機能は，まず学習用データセットのドメインから特徴ベクトルを導出する．単語グラフにおける重要性に基づいて，単語群 w からなるドメイン文字列の特徴ベクトルを次式で示す．

$$\vec{w} = \big(\mathcal{S}_1(w), \cdots \mathcal{S}_i(w), \cdots \mathcal{S}_n(w) \big)$$

次いで，それら特徴ベクトルに機械学習アルゴリズムを適用することで学習モデルを構築する．機械学習アルゴリズムには，その汎化性能と判別性能を勘案して SVM（Support Vector Machine）を採用した [28]．最終的に，未知のドメインはノイズの除去，ドメイン文字列の分割と特徴ベクトルの導出を経て，その学習モデルに基づくことにより良性と悪性が判別される．

4. 評価

本章では，実験を通じた提案手法の評価により，DNS に対する膨大な数の名前解決から悪性ドメインを判別できること，その結果に基づくことで辞書に基づく DGA マルウェアのコールバックを高精度で検出できることを示す．4.1 節で実験の諸元について述べた後，4.2 節と 4.3 節で結果について議論する．

4.1 諸元

表 2 に実験に用いたデータセットを示す．良性ドメイン X_0 は，キャンパスネットワークの RDNS で観測された，名前解決要求がドメイン名からアドレスへの変換であり，その結果として NXDOMAIN を応答したドメイン群である．キャンパスネットワークは計 6,000 人を超える学生と職員が利用しており，そのネットワークに接続する端末の名前解決を RDNS が担っている．計測の期間は 2018 年 8 月の 1 カ月間で，条件を満たす応答の数は全名前解決要求の 1.5%にあたる 3,021,124 であった．それらドメイン群に悪性ドメインが含まれないことの調査として，複数のセキュリティ製品により感染が疑われる端末を選定した後に，その端末が文献 [6], [17], [18] で報告のある “短期間に見慣れないドメインの名前解決と NXDOMAIN 応答が生じる” という DGA マルウェアの特徴と合致しないことを確認した．この調査で悪性ドメインの混在がないことを完全に担保できるわけではない．しかしながら，良性ドメインに混在する悪性ドメインの数は極微量となるため，実験結果に対して大きな影響を及ばさないと考えられる．悪性ドメイン X_1 から X_7 として，一般に公開されている 7 種の DGA マルウェアにより生成されたドメイン群を利用

表 2 各データセットにおけるドメインの数
Table 2 Numbers of benign and malicious domains in the datasets.

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Benign	Banjori	Gozi	Matsnu	Pizd	Rovnix	Sisron	Suppobox
3021124	30000	30000	30000	30000	30000	30000	30000

表 3 実験結果
Table 3 Experimental results.

	Recall	Precision
Anderson et al. [5]	0.9977	0.9305
Pereira et al. [20]	0.8873	0.6380
Our work	0.9977	0.9869

した [29], [30]．ここで，Banjori と Sisron は単語にランダムな文字列を結合することでドメインを生成するマルウェア，その他は自身が有す辞書の単語を結合することでドメインを生成するマルウェアである．それら良性と悪性ドメインに対して 5 分割交差検証を適用することで，各データセットの 20%を検証用，残りを学習用とした．

提案手法との比較のため，ドメインの文字列のみから悪性ドメインを判別する 2 種類の手法を実装した．第 1 の実装は，ドメイン文字列に対して LSTM（Long Short-Term Memory）モデルを適用することで判別する手法であり [5]，第 2 の実装は，マルウェアのコールバック先に基づいて構築した辞書の単語が，ドメイン文字列に閾値 2 以上含まれるか否かで判別する手法である [20]．

提案手法における各設定は次のとおりである．単語分割機能の辞書 $\mathbb{D}$ として，Aspell [31] とコーパス [32] に登録されている単語を利用した．その単語の総数は 500,000 を超える．単語グラフ構築機能におけるクラスタリングの閾値 th を経験的に 0.85 とした．また，SVM のカーネルとして RBF（Radial Basis Function）を採用し，そのハイパーパラメタとコストを 1.0 と 5.0 とした．これらの最適化は今後の課題とする．

4.2 定量的評価

各手法における悪意ドメインの判別性能を定量的に評価するために，一般的な 2 つの指標を用いた．再現率は，悪性ドメインの総数に対する悪性と判別されたドメインの数の比率であり，適合率は，悪性と判別されたドメインの総数に対する真に悪性であるドメインの数の比率である．

実験結果を**表 3** に示す．ここで，各値は交差検証における 5 回の試行の平均である．この結果から，提案手法は 0.9977 の再現率と 0.9869 の適合率を達成しており，2 つの実装よりも高い精度を示すことが見てとれる．2 つの実装の精度が低下した理由は次のとおりである．まず，文献 [5] に基づく実装における，単純な文字の並びのみからドメインの良性と悪性を判別することの限界である．特筆すべきは，8 文字以上のアルファベットのみからなるドメインを悪性と判別する傾向が見られ，それがゆえに多数の良性ドメインで誤判が生じることとなった．次いで，文献 [20] に基づく実装は，ドメイン文字列における単語の重要性を画一的に測るため，良性ドメインと悪性ドメインの一部で誤判が多発したことが原因である．提案手法は，良性と悪性のドメイン文字列をなす単語の差異を考慮すること，単語グラフにおいて中心的な役割を担う単語を重視することで，これらの要因を除外できたと考えられる．

各データセットにおいて誤判したドメインの数を**表 4** に示す．ここで，各値は交差検証における 5 回の試行の合計である．提案手法は，Banjori と Sisron が生成したドメインである X_1 と X_6 に対して非常に優秀な判別を実現した．その一方，良性を含む他のデータセットにおいては合計で 3,241 の誤判が生じる結果となった．その誤判の多くは次の 4 種，(a) 単一の単語のみからなるドメイン，(b) 学習用データセットにおいて，出現頻度が極端に少ない単語からなるドメイン，(c) 一般的な単語を含むドメイン，(d) 辞書に含まれない単語からなるドメインであった．本手法は，ドメイン文字列を構成する単語の関係性に着目しており，良性と悪性の判別には教師あり機械学習アルゴリズムを利用している．それらの特性上，(a) と (b) のドメインの正確な判別は困難となる．(c) のドメイン文字列を調査したところ，`domain`，`network`，`host`，`local` など，良性と悪性を問わず頻出する単語を含んでいた．そのため，自然言語処理におけるストップワードを参考に文献 [33]，ドメインにおける一般的な単語を除外することで改善が期待できる．(d) の誤判は，辞書における語彙数の不足により，ドメインを無意味な短い文字列に分割することが原因であった．具体的には，略語や頭字語を含むドメイン，非アルファベットをアルファベット表記したドメイン，固有名詞からなるドメインなどの単語分割で顕著な誤りが見られた．その改善には，より網羅的な辞書の準備が必要となる．

以上の議論より，いくつかの課題があるにしても提案手法が 0.9977 の再現率と 0.9869 の適合率で悪性ドメインを判別できることを確認した．この結果は，これら悪性ドメインの名前解決を予兆として，ネットワークに内在する辞書に基づく DGA マルウェアを高精度で検出できることを示唆している．

4.3 定性的評価

表 5 に提案手法と既存手法との定性的な比較を示す．そ

表 4 各データセットにおけるドメインの誤判数
Table 4 Numbers of misidentified domains in the datasets.

	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Anderson et al. [5]	15631	0	154	0	62	27	0	220
Pereira et al. [20]	105706	0	285	26	23286	28	0	36
Our work	2772	0	295	31	29	32	0	82

表 5 提案手法と既存手法との定性的な比較
Table 5 Qualitative comparison of our work with other well-known detection methods.

	(1) DGA malware detection	(2) Dict-DGA detection	(3) Robust to encryption	(4) Network scale independent	(5) Real-time detection
Soldo et al. [10]			✓	✓	✓
Gu et al. [12]				✓	✓
Rahbarinia et al. [15]			✓	✓	✓
Berger et al. [16]			✓		
Wang et al. [17]	✓	✓	✓		
Truong et al. [4]	✓		✓	✓	✓
Anderson et al. [5]	✓		✓	✓	✓
Pereira et al. [20]		✓	✓	✓	✓
Our work		✓	✓	✓	

の比較の観点は，(1) DGA マルウェアの検出性能，(2) 辞書に基づく DGA マルウェアの検出性能，(3) 暗号化に対する頑健性，(4) ネットワークの規模に対する依存の有無，(5) 検出の実時間性であり，各項目における対応の可否をレ点の有無により記している．なお，表中の評価は，主に 2.2 節における議論をとりまとめたものとなっている．

前節で述べたように，提案手法は辞書に基づく DGA マルウェアの高精度な検出を実現している．その検出の特徴として，Gu らの BotHunter [12] とは異なり，通信の暗号化による制限を受けないこと，Berger らの DNSMap [16] や Wang らの DBod [17] とは異なり，大規模なネットワークの観測を必要としないことがあげられる．その一方で，一般的な DGA マルウェアのための検出機能を有しない．これは辞書に基づく DGA マルウェアに特化しているためであり，それゆえに実際の運用では Anderson らの手法 [5] などを用いた補完が必須となる．

他のドメイン文字列に基づく検出と比べ [4], [5], [20]，提案手法は計算時間を要する仕組みとなっている．これは 3.3 節で述べたドメイン文字列の単語分割において，最適な候補を総あたりで見つけ出すことが原因である．この緩和のため，DNS の名前解決要求がドメイン名からアドレスへの変換であること，その結果として NXDOMAIN を応答することの，両条件を満たすもののみを対象とすることでドメインの判別数を大きく絞り込んでいる．ここで，NXDOMAIN 応答はドメインの未登録を意味するため，その名前解決に起因したデータ通信は発生しないことを留意されたい．ゆえに，本手法による判別には厳格な実時間性は要求されず，その実装の最適化を図るのみで運用に耐えうる性能を達成可能であると考えられる．

以上の議論により，辞書に基づく DGA マルウェアの検出における提案手法の優位性を明らかにした．具体的には，通信の暗号化やネットワークの規模に制限されることなく，感染端末の迅速な排除が可能であり，それゆえにネットワークの運用における安全性への貢献が期待できる．

5. おわりに

本稿では，DNS に対する膨大な数の名前解決要求から，辞書に基づく DGA マルウェアにより生成されたドメインの判別を試みた．その実現に向け，ドメインの文字列を構成する単語の関係性に基づく悪性ドメイン判別手法を提案した．また実験を通じて，提案手法が 0.9977 の再現率と 0.9869 の適合率で悪性ドメインを判別可能であること，それゆえに，悪性ドメインの名前解決を予兆として辞書に基づく DGA マルウェアを高精度で検出できることを確認した．この結果から，ネットワークに内在する多様なマルウェアへの迅速な対処が可能となるため，ネットワークの運用において安全性の向上が期待できる．今後は，大規模なネットワークで観測した通信を対象に，判別精度の継続的な評価を予定している．

謝辞 本研究は JSPS 科研費 JP18K11296 の助成を受けたものである．ここに深く謝意を示す．

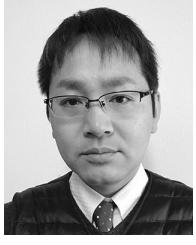
参考文献

[1] Lewis, J.A.: Economic Impact of Cybercrime — No Slowing Down, available from ⟨https://www.csis.org/analysis/economic-impact-cybercrime⟩ (accessed 2020-09-10).
[2] Fu, Y. et al.: Stealthy Domain Generation Algorithms, *IEEE Trans. Information Forensics and Secu-*

rity, Vol.12, No.6, pp.1430–1443 (2017).
[3] Satoh, A. et al.: Estimating the Randomness of Domain Names for DGA Bot Callbacks, *IEEE Communications Letters*, Vol.22, No.7, pp.1378–1381 (2018).
[4] Truong, D. et al.: Detecting Domain-Flux Botnet based on DNS Traffic Features in Managed Network, *Security and Communication Networks*, Vol.9, No.14, pp.2338–2347 (2016).
[5] Anderson, H.S. et al.: DeepDGA: Adversarially-Tuned Domain Generation and Detection, *Proc. ACM Workshop on Artificial Intelligence and Security*, pp.13–21 (2016).
[6] Sood, A.K. et al.: A Taxonomy of Domain-Generation Algorithms, *IEEE Security & Privacy*, Vol.14, No.4, pp.46–53 (2016).
[7] Kim, D.: Potential Risk Analysis Method for Malware Distribution Networks, *IEEE Access*, Vol.7, pp.185157–185167 (2019).
[8] Dwyer, C. et al.: Malvertising — A Rising Threat to the Online Ecosystem, *Journal of Information Systems Applied Research*, Vol.10, No.3, pp.29–37 (2017).
[9] Andriesse, D. et al.: Highly Resilient Peer-to-Peer Botnets Are Here: An Analysis of GameOver Zeus, *Proc. International Conference on Malicious and Unwanted Software*, pp.116–123 (2013).
[10] Soldo, F. et al.: Blacklisting Recommendation System: Using Spatio-Temporal Patterns to Predict Future Attacks, *IEEE Journal on Selected Areas in Communications*, Vol.29, No.7, pp.1423–1437 (2011).
[11] Freudiger, J. et al.: Controlled Data Sharing for Collaborative Predictive Blacklisting, *Proc. International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pp.327–349 (2015).
[12] Gu, G. et al.: BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation, *Proc. USENIX Conference on Security Symposium*, pp.167–182 (2007).
[13] Parvat, T.J. et al.: Performance Improvement of Deep Packet Inspection for Intrusion Detection, *Proc. IEEE Global Conference on Wireless Computing & Networking*, pp.224–228 (2014).
[14] Cisco Systems, Inc.: Cisco 2018 Annual Cybersecurity Report, available from ⟨https://www.cisco.com/c/en/us/products/security/cybersecurity-reports.html⟩ (accessed 2020-09-10).
[15] Rahbarinia, B. et al.: Efficient and Accurate Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks, *ACM Trans. Privacy and Security*, Vol.19, No.2, pp.4:1–4:31 (2016).
[16] Berger, A. et al.: Mining Agile DNS Traffic Using Graph Analysis for Cybercrime Detection, *Computer Networks*, Vol.100, pp.28–44 (2016).
[17] Wang, T.S. et al.: DBod: Clustering and Detecting DGA-based Botnets using DNS Traffic Analysis, *Computers & Security*, Vol.64, pp.1–15 (2017).
[18] Plohmann, D. et al.: A Comprehensive Measurement Study of Domain Generating Malware, *Proc. USENIX Conference on Security Symposium*, pp.263–278 (2016).
[19] Vinayakumar, R. et al.: Evaluating Deep Learning Approaches to Characterize and Classify the DGAs at Scale, *Journal of Intelligent and Fuzzy Systems*, Vol.34, No.3, pp.1265–1276 (2018).
[20] Pereira, M. et al.: Dictionary Extraction and Detection of Algorithmically Generated Domain Names in Passive DNS Traffic, *Proc. International Symposium on Research in Attacks, Intrusions, and Defenses*, pp.295–314 (2018).
[21] Koren, A.: Ursnif Malware: Deep Technical Dive, available from ⟨https://arielkoren.com/blog/2016/11/01/ursnif-malware-deep-technical-dive/⟩ (accessed 2020-09-10).
[22] Skuratovich, S.: Matsnu: A Deep Dive, available from ⟨https://blog.checkpoint.com/2015/07/02/matsnu-a-new-malware-discovery/⟩ (accessed 2020-09-10).
[23] Sahoo, D. et al.: Malicious URL Detection using Machine Learning: A Survey, arXiv:1701.07179, pp.1–21 (2017).
[24] Mockapetris, P.: Domain Names — Implementation and Specification, IETF Request for Comments 1035 (1987).
[25] Costello, A.: Punycode: A Bootstring Encoding of Unicode for Internationalized Domain Names in Applications (IDNA), IETF Request for Comments 3492 (2003).
[26] Müllner, D.: fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python, *Journal of Statistical Software*, Vol.53, No.9, pp.1–18 (2013).
[27] Csárdi, G. et al.: The igraph Software Package for Complex Network Research, *InterJournal Complex Systems*, No.1695 (2006).
[28] Karatzoglou, A. et al.: Support Vector Machines in R, *Journal of Statistical Software*, Vol.15, No.9, pp.1–28 (2006).
[29] Bader, J.: Some Results of My DGA Reversing Efforts, available from ⟨https://github.com/baderj/domain_generation_algorithms⟩ (accessed 2020-09-10).
[30] Fraunhofer FKIE: DGArchive, available from ⟨https://dgarchive.caad.fkie.fraunhofer.de⟩ (accessed 2020-09-10).
[31] Atkinson, K.: GNU Aspell, available from ⟨http://aspell.net⟩ (accessed 2020-09-10).
[32] Norvig, P.: Natural Language Corpus Data: Beautiful Data, available from ⟨http://norvig.com/ngrams/⟩ (accessed 2020-09-10).
[33] Nothman, J. et al.: Stop Word Lists in Free Open-source Software Packages, *Proc. Workshop for NLP Open Source Software*, pp.7–12 (2018).

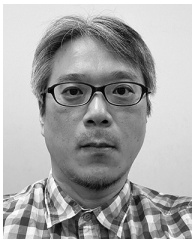
佐藤 彰洋 （正会員）

九州工業大学情報基盤センター助教．2011 年東北大学大学院情報科学研究科博士後期課程修了．博士（情報科学）．ネットワーク運用技術，ネットワークセキュリティに関する研究に従事．電子情報通信学会会員．



福田 豊 （正会員）

九州工業大学情報基盤センター准教授．2005 年九州工業大学情報工学研究科博士後期課程修了．博士（情報工学）．情報ネットワーク，無線 LAN に関する研究に従事．IEEE，電子情報通信学会各会員．



井上 純一

九州工業大学飯塚キャンパス技術部技術専門職員．コンピュータの利活用とネットワークの運用に関する業務に従事．

中村 豊 （正会員）

九州工業大学情報基盤センター教授．2001 年奈良先端科学技術大学情報科学研究科博士後期課程修了．博士（工学）．インターネット計測技術，ネットワーク運用技術，ネットワークセキュリティに関する研究に従事．電子情報通信学会会員．