DNS ログからの不正 Web サイト抽出について －解析手法とその匿名化－

田中 晃太郎 †　　長尾 篤 †　　森井 昌克 †

† 神戸大学大学院工学研究科
657-8501 兵庫県神戸市灘区六甲台町 1-1
ko-tanaka@stu.kobe-u.ac.jp
a.nagao@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

あらまし　インターネット上の不正 Web サイトには様々なマルウェアが仕掛けられており，被害を抑えるためには新たな不正 Web サイトを早急に発見する必要がある．不正 Web サイトのドメインは正常なドメインと比較して特異な性質を持つことが多い．本研究では不正 Web サイトを検知するためにマルウェアの挙動に着目し，DNS ログの解析によって未だ報告されていない不正 Web サイトのドメインを抽出する手法を提案する．提案手法を用いた評価実験では未報告の不正 Web サイトである可能性の高い悪性のドメインを抽出することができた．さらに本稿では提案手法を匿名化された DNS データを用いて適用する場合について考察を与える．

Extracting Malicious Website from DNS Log -Analysis Method and Anonymity-

Kotaro Tanaka†　　Atsushi Nagao†　　Masakatu Morii†

†Graduate School of Engineering, Kobe University,
1-1 Rokkodai, Nada, Kobe, 657-8501, Japan.
ko-tanaka@stu.kobe-u.ac.jp
a.nagao@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

Abstract　Website which have been created maliciously have various kinds of malware. Detecting new malicious Website immediately prevent the Internet user from damage. Domains of malicious Websites often have abnormal properties compared to benign domains. We propose a method to extract new domains of malicious Websites by analyzing DNS log. We focus on behavior of malware. As a result, we extracted unreported malicious domains. In addition, we also consider the proposed method against anonymized DNS data.

1　はじめに

Domain Name System(DNS) はインターネットを利用する上で不可欠な機能となっている．通常ユーザが利用する Web サイトの URL はアルファベットや数字で記されたドメインから構成される．多くの場合ドメインはユーザに対して馴染み深い文字列で示されており，DNS は Web サイトに割り当てられたドメイン名と IP アド

レスを対応付ける役割を担っている．DNS によってドメイン名から IP アドレスを求めることを名前解決と呼ぶ．

DNS は多くのインターネットユーザにとって不可欠である一方で，Fast-Flux 攻撃や DNSamp といった DNS を悪用した攻撃が存在する．これらの場合において名前解決が要求される不正 Web サイトのドメインは悪意のあるものによって管理されており，これを悪性ドメインと呼ぶ．悪性ドメインは正常なドメインと比較して名前解決時に異なる特徴が見られる場合が多い．DNS キャッシュサーバの通信を観測することにより，セキュリティ上の異常を検出する研究がこれまでに行われてきた [1, 2]．しかし既存の研究では悪性ドメインが名前解決される際の時間的周期や TTL 値といった特徴に着目したものが多く，マルウェアの挙動に基づいた研究はあまり行われていなかった．

本研究ではマルウェアが通信を行う際の特徴に着目することで，DNS 通信の観測を通じて新たな不正 Web サイトの抽出方法を提案する．マルウェアが通信を行う際の特徴として，マルウェア感染クライアントは複数の不正 Web サイトへとアクセスを行うことが知られている．ある不正 Web サイトへアクセスを行ったクライアントは他の不正 Web サイトへもアクセスを行っている可能性が高い．そこで，DNS 通信において既知の悪性ドメインへアクセスを行ったクライアントのログを抽出する．複数のクライアントから重複して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悪性ドメインだと考えられ，新たな不正 Web サイトの発見に繋がる．

また本研究では DNS 通信の観測データを匿名化した上で解析を行うことも目的としている．本稿では提案手法を匿名化された DNS データを用いて適用する場合を想定し，匿名化すべき要件についても簡単に述べる．

2　関連技術

本章では DNS 通信ログを用いた解析に関する既存研究と，データの匿名化に関する技術について紹介する．

2.1　DNS 通信ログの解析

DNS に関わるセキュリティ問題の増加を背景として，DNS 通信ログからマルウェアに関連した異常を検知する研究が増えてきている．2011 年，Bilge らは特定の攻撃や問題にとらわれず，DNS に関わる問題全般に対応した解析手法を提案した [1]．Bilge らは正常なドメインと悪性ドメインを区別するための複数の特徴を提示した．具体的には名前解決動作の周期性，様々な悪性ドメインとの IP アドレスの共有，TTL 値の特異性，ドメイン名に乱数が含まれる，といった項目が悪性ドメインの特徴として挙げられている．これらの特徴を利用して数ヶ月にわたる DNS 通信ログの分析を行い，高い精度で悪性ドメインが検出できることを示した．さらにこの手法を取り入れた解析システムを ISP に導入し，リアルタイムで解析を行った場合にも未報告の悪性ドメインを検出できることを示した．論文 [2] では DNS 通信ログを利用し，悪意あるクライアントによって構築された可能性の高い権威 DNS サーバを検出する手法が Minn らによって示された．Minn らはまず DNS 通信ログ中で名前解決が要求されたドメインを全て抽出し，それぞれのドメインに対する権威 DNS サーバを求めた．多くの既知悪性ドメインとつながりのある権威 DNS サーバを抽出し，マルウェア報告サイトにおけるデータとの比較によって検証を行った．検証の結果，Minn らが抽出した権威 DNS サーバ群には悪性の権威 DNS サーバが多く含まれていることが示された．その他にも DNS 通信ログを用いた研究は近年盛んに行われており [3, 4, 5]，セキュリティ上の異常を検知する上で役立つと考えられる．

2.2　DNS 通信ログの匿名化

本研究では DNS 通信の観測データを匿名化した上で解析することも目的としている．匿名化はデータを所持している機関がデータを分析する他機関へデータを提供する際に用いられる

技術である．単一または複数の機関が所持するデータを収集・分析し，新たな知見を得ることをデータマイニングと呼ぶ．しかし，データ中にプライバシに関わる情報が含まれる場合は，データの取り扱いに細心の注意を払う必要がある．特にデータを所持する機関と分析する機関の間に十分な信頼関係が結ばれていない場合は，ユーザのプライバシに関わる情報が意図しない用途で用いられる危険性がある．そこでユーザのプライバシに関わる情報を秘匿した上でデータマイニングを行う技術として，プライバシ保護データマイニング (Privacy-Preserving Data Mining) [6] が必要とされている．プライバシ保護データマイニングでは，ユーザのプライバシを保護した上で提供データを最大限に活用できるような匿名化手法が必要となる．

DNS 通信ログについても，複数の機関が所持する DNS 通信ログを収集し解析することで，様々な攻撃やマルウェアによる被害を未然に防ぐことが期待される．しかし DNS 通信ログにはパケットを送信したクライアントのプライバシにつながる情報が含まれているので，DNS 通信ログを他機関へ提供する際には適切な匿名化を施す必要がある．

3 DNS 通信ログを利用した不正 Web サイトの抽出

我々は不正 Web サイトにアクセスを行うマルウェアの挙動に着目し，DNS 通信ログを利用して不正 Web サイトを抽出する手法を提案する．本章では不正 Web サイト抽出につながるマルウェアの挙動について述べる．

3.1 マルウェア感染コンピュータ

通常インターネットを使用しているユーザが自らの意思で不正 Web サイトへアクセスすることは稀である．しかしマルウェアに感染しているコンピュータからはユーザの意思に関わらず特定の Web サイトへとアクセスが行われる場合がある．マルウェアが感染コンピュータを特定の Web サイトへアクセスさせるためである．マルウェア感染コンピュータは通常のユーザと比較して，不正 Web サイトへのアクセス頻度が増加する．そのため不正 Web サイトへアクセスを行っているコンピュータはマルウェアに感染している可能性が高いと考えられる．

3.2 不正 Web サイト抽出

マルウェアは感染を拡大させるために複数の不正 Web サイトへアクセスを試みると考えられる．このような不正 Web サイトは同時に他のマルウェア感染コンピュータからもアクセスが行われている可能性が高い．そこで我々はマルウェア感染コンピュータから重複してアクセスが行われている Web サイトを抽出することで，新たな不正 Web サイトが特定できると考えた．特に通常のユーザからアクセスが少ないにも関わらず，マルウェア感染コンピュータからは重複してアクセスされている Web サイトには，不正 Web サイトが高い割合で含まれていると考えられる．

4 悪性ドメインを用いた未知の不正 Web サイト抽出法

本章では不正 Web サイトを抽出するための手法について述べる．4.1 節で DNS 通信ログを用いたマルウェア感染クライアントの検出法について説明し，4.2 節でマルウェアの挙動に基づく不正 Web サイトの抽出法について説明する．

4.1 マルウェア感染クライアントの検出

キャッシュDNS サーバの通信ログを用いたマルウェア感染クライアントの抽出法を図 1 に示す．図 1 における既知悪性ドメインには，Malware domain list(MDL)[7] や MalwareURL(M-URL)[8] において報告されている明らかな悪性ドメインを使用する．リスト中の悪性ドメインは実際に不正な活動が確認されており，通常のユーザが自らの意思でこれらのドメインと結びつく Web サイトへアクセスを行うとは考え難い．図 1 ではクライアント 1 とクライアント 3

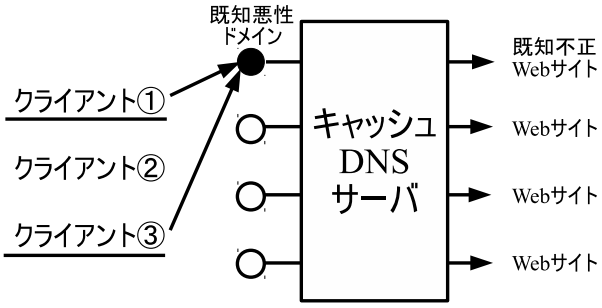

図 1: マルウェア感染クライアントの検出

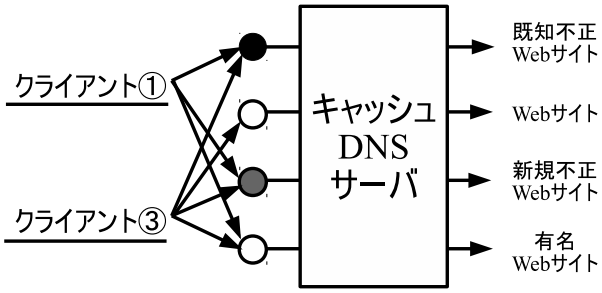

図 2: 不正 Web サイトの抽出法

が悪性ドメインに対して名前解決要求を行っている．クライアント 1,3 はクライアント 2 と比較してマルウェアに感染している可能性が高い．したがってクライアント 1,3 をマルウェア感染クライアントとして検出する．

4.2 マルウェアの挙動に基づく不正 Web サイトの抽出

マルウェアはクライアントを他のマルウェアにも感染させるために，特定の Web サイトへアクセスすることが知られている 1．4.1 節で検出したマルウェア感染クライアントの情報を用いた不正 Web サイトの抽出法を図 2 に示す．図 2 におけるクライアント 1,3 は，4.1 節で検出したマルウェア感染クライアントである．クライアント 1,3 は多数のドメインに対して名前解決要求を行うが，この中にはマルウェアが Web サイトへアクセスする際に生じた名前解決要求が含まれる．特に複数のマルウェア感染クライアントから重複してアクセスが行われた Web サイトはマルウェアがアクセスを試みた Web サイトである可能性が高い．マルウェアがアクセスを試みたドメインを抽出するため，図 2 においてクライアント 1,3 から重複して名前解決要求が行われたドメインを抽出する．またクライアントはマルウェアに完全に支配されているとは考え難く，ユーザの操作により生じた正常な Web サイトへの名前解決要求を除外する必要がある．そのため，Alexa[9] で公開されている上位ドメインを重複したドメインリストから除外する．

以上の手順により抽出したドメイン群には未知の悪性ドメインが含まれると思われる．さらに新規悪性ドメインの発見精度を上げるため，マルウェア感染クライアントの数は多い方が望ましい．重複したドメイン群から既知悪性ドメインを抽出し，新たな感染クライアントとして検出する．感染クライアントから重複してアクセスされたドメインを抽出する．この手順を複数回繰り返した結果において，数多く検出されたドメインを悪性ドメインとして抽出する．

5 評価実験

本章では実際のキャッシュDNS サーバの通信ログを用いた実験により，提案手法の有効性を検証する．

5.1 分析対象の通信ログ

本稿では実際のキャッシュDNS サーバの通信ログを用いて実験を行った．本実験では 2013 年 3 月の 22 時台における 5 分間の通信ログにおいて正引きを要求したクエリを用いた．またその際，RFC1035[10] の定義から外れるクエリは除外した．5 分間のうちでドメインに対して正引きを要求したクエリ数は約 500 万件であった．

5.2 実験方法

本節では実験の手順を述べる．本実験では STEP2 から STEP4 の手順を繰り返すことで未知の悪性ドメインの抽出を試みる．

STEP1 マルウェア感染クライアントの検出: キャッシュDNS サーバの通信ログにおいて，既知の悪性ドメインである”tapjoyads.com”に対して名前解決要求を行った送信元 IP アドレスを抽出する．名前解決要求が多いクライアント 10 件をマルウェア感染クライアントとして検出する．

STEP2 ドメインの抽出: ログ中で複数のマルウェア感染クライアントから重複して名前解決要求が行われたドメインを重複件数ごとに抽出する．抽出したドメインから Alexa で公開されている上位 100 万件以内のドメインとセキュリティチェックに絡むドメインは除外する．重複件数の閾値を増加させ，ドメイン数が 50 を下回るように重複ドメインを抽出する．

STEP3 既知悪性ドメインの抽出: 抽出したドメインの内，ドメイン名またはその正引き後 IP アドレスが MDL や M-URL で報告されているドメインを既知悪性ドメイン群とする．なお本実験で用いたログは 2013 年 3 月のデータであるため，MDL と M-URL についても 2013 年 3 月以前のリストを用いた．

STEP4 マルウェア感染クライアントの検出: 全ての既知悪性ドメインについて，メインドメインが一致するものをログ全体から抽出し，既知悪性ドメイン群に加える．既知悪性ドメインそれぞれに対して名前解決要求を行ったクライアントを，メインドメインごとに名前解決要求回数を基準として上位 10 クライアントを抽出する．抽出した全てのクライアントをマルウェア感染クライアントとする．

STEP5 提案手法による未知ドメインの抽出: STEP2〜4 を繰り返し，複数回にわたって出現した重複ドメインを最終的な重複ドメインとする．最終的な重複ドメイン数を表 1 に示す．

5.3 実験結果の検討

表 1 の実験結果において，出現回数が多いドメインほど悪性である可能性が高い．まず 10 回の実験において 5 回以上出現した 33 ドメインについて検討を行う．33 ドメインから既知の悪性ドメイン 16 個を除いた 17 個のドメインに対して，virustotal[11] のデータセット内検索を用いてマルウェアとの関連性を検索した．検索の結果，3 個のドメインがマルウェアのダウンロード URL との結びつきが報告されており，これら 3 個のドメインは不正 Web サイトに結びついた悪性ドメインである．また virustotal においてマルウェアとの関連が報告されていない残りの 14 個のドメインについても健全でないドメインが含まれると考えられる．14 ドメインについて内訳を検索したところ，アドウェアやアップローダサイト，広告サイトなどが含まれ，推定不可なものも存在した．アップローダサイトに関しては悪意のあるものによってマルウェアがアップロードされ，提案手法の定義における不正 Web サイトとして抽出されたと推測できる．また推定不可なものに関しては悪性である可能性が高いと考えられる．33 ドメインに対する検討結果を表 2 に示す．

表 1: 重複ドメイン数

出現回数	ドメイン数	既知悪性ドメイン数
2 回以上	79	22
3 回以上	52	19
4 回以上	39	18
5 回以上	33	16
6 回以上	22	10
7 回以上	12	6
8 回以上	9	5
9 回以上	4	3
10 回	3	2

次に 10 回の実験全ての場合において出現した 3 ドメインについて検討を行う．この 3 ドメインを表 3 に示す．表 3 において 3 ドメインの内 2 個は既知悪性ドメインであるが，”bid.socdm.com”に関しては悪性ドメインとして報告されておらず，また virustotal を用いてもマルウェアとの関連性を見つけることはできなかった．しかし既知の悪性ドメインとの

表 2: 5 回以上出現したドメイン

ドメイン名	悪性報告	内訳（推定）
adadvisor.net	MDL	-
adc.media-rep.com	報告なし	広告
all.adadvisor.net	MDL	-
an.tacoda.net	MDL	-
as.a.bypass.jp	報告なし	アドウェア
b4.i.adimg.net	報告なし	アップローダ
beacon.krxd.net	MDL	-
bid.socdm.com	報告なし	不明
cdn.c-team.jp	報告なし	アドウェア
cdn.krxd.net	MDL	-
cf.eco-tag.jp	報告なし	**マルウェア**
g1.panthercdn.com	MDL	-
i.adingo.jp.gslb.idc.jp	報告なし	不明
idsync-ext.rlcdn.com	報告なし	**マルウェア**
img.adplan-ds.com	MDL	-
img.fluct.jp.eimg.jp	報告なし	広告
img.newswatch.jp	報告なし	ニュース配信
imgcdn.ptvcdn.net	M-URL	-
loadm.exelator.com	報告あり	-
m.addthisedge.com	報告なし	**マルウェア**
myup.jp	報告なし	アップローダ
pict-navi.net	報告なし	アダルト
poolus.exelator.com	MDL	-
recruit.112.2o7.net	MDL	-
rt.legolas-media.com	MDL	-
secure-us.imrworldwide.com	MDL	-
sh.adingo.jp.gslb.idc.jp	報告なし	不明
tg.socdm.com	報告なし	不明
ws.tapjoyads.com	MDL	-
www.adadvisor.net	MDL	-
www.bmmetrix.com	MDL	-
www.grail.bz.gslb.idc.jp	報告なし	不明
www31.tracer.jp	報告なし	クッキー

表 3: 10 回出現したドメイン

ドメイン名	悪性報告
bid.socdm.com	報告なし
img.adplan-ds.com	MDL
ws.tapjoyads.com	MDL

結びつきの強さから，このドメインが不正 Web サイトに結びついた悪性ドメインである可能性は高いといえる．

6 匿名化データに対する提案手法の適用

本章では匿名化された DNS 通信ログを用いて提案手法を適用するための考察を与える．提案手法を用いた解析の際に必要なデータは，マルウェア感染クライアントの IP アドレスと名前解決が要求されたドメイン名である．クライアントの IP アドレスに関しては同一性を確保する必要があるが，具体的な値は利用しない．したがって自明な匿名化の方法としてハッシュ関数を用いることが考えられる．IP アドレスをハッシュ関数に入力する際，解析データ毎に異なる冗長部分を加えることで，より匿名性を高めることができる．さらにプライバシ保護に重点をおいて匿名化する場合には，IP アドレスの下位ビットを削除する Sanitization が考えられる．しかし Sanitization による匿名化はデータの一部を切り落とすため，提案手法による解析の有効性が著しく低下する可能性がある．Sanitization を行う際に切り落とすデータの範囲は，解析後データの有用性について十分検討した上で決定する必要がある．IP アドレスに対する Sanitization については Burkhart らによって考察が行われている [12]．また，ドメイン名に関しては既存の研究手法を容易に適用することができない．ドメイン名に関する情報を匿名化してしまうと不正 Web サイトの報告ができなくなるためである．また提案手法を適用するためには，匿名化する際にドメイン名とその正引き IP アドレスの情報を結びつけて行う必要があると考えられる．今後の課題として有効な匿名化手法を検討する必要がある．

7 まとめ

本稿では，マルウェアの挙動に着目することで新たな不正 Web サイトを抽出する手法を提案した．提案手法の評価実験では実際のキャッシュ DNS サーバの通信ログを用いて，未だ報告されていない不正 Web サイトのドメインを抽出することができた．また提案手法を匿名化された DNS データを用いて適用する場合について，匿名化を行う上での要件について考察を行った．匿名化の要件に関しては十分な検討を行えていないので，提案手法を用いた解析に対して有効な匿名化手法について今後検討を行う必要がある．

謝辞

本研究の一部は、総務省による国際連携によるサイバー攻撃予知・即応プロジェクト『PRACTICE』の一環である.

参考文献

[1] L.Bilge, E.Kirda, C.Kruegel, and M.Balduzzi “EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis”. Proceedings of NDSS, 2011.

[2] Y.Minn, D.Makita, K.Yoshioka, and T.Matsumoto “Finding Malicious Authoritative DNS Servers”. ICSS, pp.25-30, 2013.

[3] 牧田 大佑, Y.Minn, 吉岡 克成, 松本 勉 “名前解決動作の類似性に基づくマルウェア感染ホストの特定”. SCIS2013, pp.1-6, 2013.

[4] 田辺 瑠偉, 鉄 穎, 水戸 慎, 牧田 大佑, 神薗 雅紀, 星澤 裕二, 吉岡 克成, 松本 勉 “長期動的解析によるマルウェアの特徴的な DNS 通信の抽出”. CSS2012, pp.712-719, 2012.

[5] 金井 文宏, 吉岡 克成, 松本 勉 “動的解析による Android マルウェアの DNS 通信の観測”. ICSS2013, pp.31-36, 2013.

[6] C.C. Aggarwal, and P.S. Yu. “Privacy-Preserving Data Mining: Models and Algorithms”. 2008.

[7] Malware domain list.
`http://www.malwaredomainlist.com/`

[8] MalwareURL.
`http://www.malwareurl.com/`

[9] Alexa Web Information Company.
`http://www.alexa.com/`

[10] P.Mockapetris “Domain names - implementation and specification”. IETF RFC-1035, 1987.

[11] virustotal.
`https://www.virustotal.com/`

[12] M.Burkhart, D.Brauckhoff, M.May, andE.Boschi “The Risk-Utility Tradeoff for IP Address Truncation”. ACM Workshop 2008, pp.23-30, 2008.