ドメイン情報に着目した悪性 Web サイトの活動傾向調査と関連性分析

福島 祥郎 †‡　　堀 良彰 †‡　　櫻井 幸一 †‡

† 九州大学大学院　システム情報科学府
〒 819-0395 福岡市西区元岡 744
‡ 財団法人九州先端科学技術研究所
〒 814-0001 福岡市早良区百道浜 2-1-22 福岡 SRP センタービル 7 階
yfukushima@itslab.csce.kyushu-u.ac.jp, {hori,sakurai}@inf.kyushu-u.ac.jp

あらまし　正規サイトの改竄によってユーザを攻撃者が用意した攻撃サイトに誘導し，Web ブラウザの脆弱性を突くことで感染を広げる Web 感染型マルウェアの脅威が増加している．Web 感染型マルウェア対策には，その活動形態を調査し攻撃者の手口を解明することが重要となる．そこで本研究では，悪性 Web サイトのドメイン情報に着目し，DNS レコードや whois の登録情報などをもとに，各サイトの活動傾向の調査と関連性分析を行った．実験により，攻撃者の悪性 Web サイト設置における手口を明らかにし，特に使用されるレジストラに偏りがあることがわかった．

Study of Malicious Web Site Activities and their Relational Analysis Based on Domain Information

Yoshiro Fukushima†‡　　Yoshiaki Hori†‡　　Kouichi Sakurai†‡

†Graduate School of Information Science and Electrical Engineering, Kyushu University
744 Motoka, Nishi-ku, Fukuoka 819-0395, Japan
Institute of Systems and Information Technologies and Nanotechnologies
2-1-22 Momochi-hama, Sawara-ku, Fukuoka 814-0001, Japan
yfukushima@itslab.csce.kyushu-u.ac.jp, {hori,sakurai}@inf.kyushu-u.ac.jp

Abstract　The threat of web based malware which leads users to attacker's sites from infected legitimate Web sites and exploits user's Web browser is increasing. It is important to investigate malicious activities of web based malware and clarify attacker's methodology for tackling it. In this paper, we investigate malicious Web site's activities and analyze relationship between them based on domain information such as DNS records and whois information. Our results reveal attacker's methods on deploying malicious Web sites and find that they use specific registrars.

1　はじめに

近年，コンピュータウィルスやワーム、トロイの木馬などといったマルウェアによる被害が深刻化してきている．特に近年の攻撃者は金銭獲得を主な目的としており，ユーザの機密情報の搾取や，ユーザをボットネットに組み込み，スパムメールや DoS 攻撃などの不正活動を行う踏み台として悪用している．さらに，攻撃者はより多くのユーザをマルウェアに感染させるために様々な感染手段を取るようになっている．従来はワームの感染活動のような，攻撃者が起点となって攻撃が開始され，OS の脆弱性などを悪用して感染を試みる「能動型攻撃」が主流で

あった．しかし，最近はパーソナルFWやNAT機能を備えたBBルータの普及により，能動型攻撃によるマルウェア感染の脅威は減少しつつある．そのため，近年では能動型攻撃に加えて，主に悪性Webサイトへのアクセスなどの，ユーザの行動が起点となって攻撃が開始される「受動型攻撃」が増加している [7, 8, 9, 12]．受動型攻撃では主に，悪性Webサイトにユーザを誘導し，Webブラウザなどの脆弱性を突くことでマルウェアに感染させる．

受動型攻撃により，Webブラウザなどの脆弱性を悪用して感染を広げるマルウェアはWeb感染型マルウェアと呼ばれる．また，この脆弱性攻撃からマルウェア感染までの一連の攻撃は，drive-by-download攻撃と呼ばれる．悪性Webサイト対策には，その攻撃の検知と活動形態の調査を行うことが重要であり，そのためにWebクライアントハニーポットに関する研究 [7, 8] や悪性Webサイトの実態調査 [3, 4] に関する研究が行われている．しかし，より効果的な対策に向けて，悪性Webサイト間の関連性を分析し，攻撃者の悪性Webサイト設置における手口と意図の解明を行い，それを今後の対策に結び付けることが重要となる．

本論文では，攻撃者の悪性Webサイト設置における手口と意図把握のため，悪性Webサイトのドメイン情報に着目してその活動傾向調査と関連性分析に取り組む．特に，悪性Webサイトを正規サイトの改竄によるものと攻撃者が自ら設置したものに分類し，後者に着目することで攻撃者の不正活動に関連する特徴抽出を行う．また，マルウェア配布サイトの特徴分析についても試みる．

2 Drive-by-download攻撃

2.1 攻撃手法

drive-by-download攻撃の一連の流れを図1に示す．攻撃者はまず，正規のWebサイトをSQLインジェクションやFTPアカウントの盗用などにより改竄し，攻撃サイトへの誘導コードを挿入する．攻撃サイトには，ユーザのWebブラウザや各種プラグインの脆弱性を突く攻撃コードが設置されており，ユーザにそれらの攻撃コードを送信する．攻撃が成功すると，ユーザの意図しないところで勝手にマルウェアがダウンロード・インストールされる [7]．本論文では，ユーザに実際に攻撃を行うサイトを攻撃サイト，攻撃サイトへの誘導を行うサイトを誘導サイト，マルウェアダウンロード先のサイトをマルウェア配布サイトと呼ぶことにする．また，これらを総称して単に悪性Webサイトと呼ぶ．

2.2 攻撃の巧妙性と対策の困難性

Web感染型マルウェアによる脅威が深刻化している要因として，その感染時の攻撃手法の巧妙性と対策の困難性が挙げられる．誘導サイトや攻撃サイトで用いられる不正コードは複雑に難読化されており [1]，単純なパターンマッチでは検知は困難である．また，標的となる脆弱性は既知のものだけではなく，ゼロデイ状態のものも狙われ，セキュリティ意識の高いユーザであっても感染する恐れがある [5]．さらに，誘導サイトは通常改竄された正規サイトであるため，より多くのユーザを効率的かつ気づかれることなくマルウェアに感染させることができる．また，正規サイト（誘導サイト）が幾重にも踏み台とされ攻撃者の追跡も困難である [3]．加えて，不正コードやマルウェアバイナリが頻繁に更新され，ウィルス対策ソフトの検知率が低いことも要因の1つである．

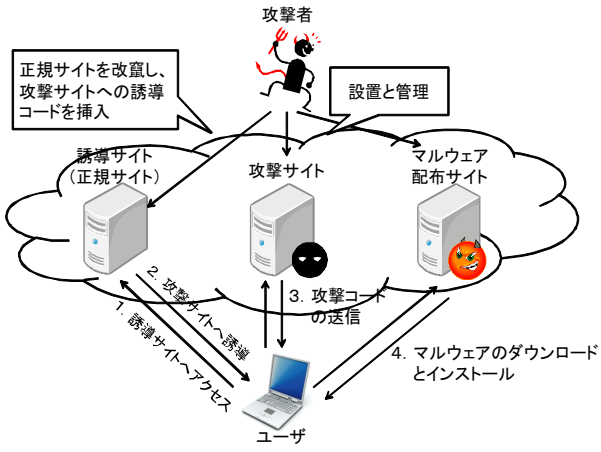

図 1: Drive-by-download攻撃の流れ

3 関連研究

Web 感染型マルウェア対策のために，文献 [7] ではその実態調査のための要件を挙げ，その点について取り組んでいる．本節ではその実態調査要件に沿って，関連研究について述べる．まず 1 つ目の要件は，攻撃の検知とその攻撃手法の詳細把握である．そのために，クライアントハニーポットに関する研究 [7, 8] や大規模実態調査 [3, 4, 9]，また異常検知 [1] に基づく攻撃検知に関する研究が行われている．実態調査では，主に悪性 Web サイトの活動の時系列推移や悪用される脆弱性の特定などを行っている．2 つ目の要件は，攻撃サイト URL の収集である．Malware Domain List（MDL）[2] や HoneyWhales[6] などの Web サービスでは，各種悪性 Web サイトの URL が公開されている．3 つ目の要件は，感染後の挙動解析であり，能動型マルウェアと受動型マルウェアの挙動の相違に関していくつか研究が行われている [12, 7]．最後の 4 つ目の要件は，悪性ネットワーク追跡による大本のサイト特定である．これらの研究では主に，悪性 Web サイト間の接続関連性 [7] や，地理的分布の可視化 [10] について取り組んでいる．本研究では，攻撃者の悪性 Web サイト設置に関する手口やその意図を解明するために，単なる接続関連性や分布ではなくて，その他の視点からの特徴分析に取り組む．

4 悪性 Web サイトの特徴分析

4.1 実験データ

本実験では，D3M 2010[11] の攻撃通信データを用いて実験を行った．攻撃通信データには，クライアントハニーポットの巡回対象となる悪性 Web サイトの URL リストが 3 日間分と，その URL へのアクセスからマルウェア検体ダウンロードまでの通信データが含まれる．

4.2 本研究で取り組むこと

本研究の目的は，悪性 Web サイトの活動傾向とサイト間の関連性を分析することで，攻撃者の悪性 Web サイト設置における手口と意図を解明し，それを今後の対策に役立てることである．そのために本研究では悪性 Web サイトのドメイン情報に着目し，以下の項目について調査と分析を行う．

悪性 Web サイトのドメイン情報と生存性調査

悪性 Web サイトの URL から FQDN を抽出し，その FQDN に対して DNS レコードと whois 情報を取得する．それらの情報から，情報取得時（2010 年 8 月上旬）における悪性 Web サイトの活動傾向に関して，登録情報の有無や悪性 Web サイトの生存性などの観点から特徴分析を行う．また，取得データを利用して以下の 2 点について取り組む．

攻撃者設置サイトの特定と特徴分析

通常，悪性 Web サイトは正規サイトが改竄されたものと攻撃者が自ら設置したものの 2 つに分類できる．攻撃者の悪性 Web サイト設置における手口を把握するためには，前者は除外し後者だけに着目して特徴抽出を行う必要がある．そこで，悪性 Web サイトのドメインの whois 情報を利用し，ドメインの登録日に着目した攻撃者設置サイト特定について検討する．またその際，我々の特定手法と MDL[2] 登録情報に基づいた特定について比較と評価を行う．そして，特定した攻撃者設置サイトから攻撃者の不正活動に関係する特徴を，レジストラ情報などの観点から抽出し，その手口や意図の把握を試みる．

攻撃サイト/マルウェア配布サイト間の特徴分析

本研究では，攻撃サイトアクセス後に接続するサイトをマルウェア配布サイトと考える．そのためにまず，MDL 登録情報を利用して巡回対象 URL から攻撃サイトの URL を抽出する．そしてその URL アクセス後に接続するサイトの FQDN を抽出し，それをマルウェア配布サイトとする．その後，攻撃サイトとマルウェア配布サイトについて，ドメイン情報の観点から特徴の違いや関連性分析などを試みる．

5　実験結果の分析

5.1　悪性 Web サイトのドメイン情報

D3M 2010 における巡回対象 URL3 日間分の統計を表 1 にまとめる．同じ FQDN を持つホストに異なるパスでアクセスしているため，巡回 URL 数に対して FQDN 数は少ない．また，FQDN 数に対してユニーク IP アドレス数が少ないことから，複数の FQDN に同じ IP アドレスが割り当てられていることがわかる．このような特徴は文献 [9] でも報告されている．また，各日の巡回 URL は大部分が共通していたため，以降では 3 月 8 日分の URL にのみ着目する．

次に，3 月 8 日の各巡回 URL の FQDN について，その DNS レコードと whois 情報を取得した（表 2）．なお，これらの情報取得は 2010 年 8 月上旬と，3 月 8 日から時間が経っており，すでに閉鎖されたサイトも存在する．そのため，8 月上旬で名前解決可能であった（A レコードが存在する）FQDN の数は 74 個と全体の半分であった．悪性 Web サイトの生存性については 5.3 節で詳しく議論する．一方，whois 情報は 8 割以上が取得可能であり，攻撃者の手口や意図把握する上で有効な情報となりえると言える．

表 1: 巡回対象 URL の統計データ

	3 月 8 日	3 月 9 日	3 月 11 日
巡回 URL 数	205	180	169
FQDN 数	148	136	123
ユニーク IP アドレス数	108	97	95

表 2: 3 月 8 日巡回 URL の DNS と whois 情報

FQDN	DNS 登録なし	DNS 登録あり		whois 登録あり
		A	ANY	
148	48	74	91	127

5.2　攻撃者設置サイトの特定

攻撃者の悪性 Web サイト設置における手口と意図把握のため，巡回 URL のサイトの中から改竄により悪性となった正規サイトを除外し，攻撃者が自ら設置したサイトのみ抽出する．そのために，以下の仮定の下，悪性 Web サイトの whois 情報内のドメイン登録日に着目する．

攻撃者設置サイトのドメイン登録日は新しい

攻撃者は攻撃サイト構築ツールなどを用いて，新規に悪性 Web サイトを設置する．そのため，ドメイン登録からそれが MDL[2] などにより悪性と判断されるまでの期間が短い．

正規サイトのドメイン登録日は古い

正規サイトは改竄前は正常に運用されてきたものと考える．よって，ドメイン登録から改竄の影響で悪性と判断されるまでの期間は長い．

本研究では具体的には，D3M 2010 のデータ取得が 2010 年 3 月上旬ということを勘案し，2010 年のドメイン登録日を持つものを攻撃者設置サイトと考えた．

3 月 8 日の巡回 URL の FQDN について，取得可能であったものについてそのドメイン登録日を集計した（図 2）．2010 年 1 月〜3 月の登録日を持つものが多く，取得可能全体数 120 個に対して 57 個がその期間に登録されていた．我々の基準ではそれら 57 個の FQDN が攻撃者設置サイトとなる．次に，各巡回 URL の FQDN について，MDL[2] の登録情報を参照し，それが悪性 Web サイトと判断された理由を調査した．その結果，全 148 個の FQDN の内，攻撃コードの設置によるものが 93 件，誘導コードの設置によるものが 25 件，その他の理由が 11 件，登録なしが 19 件であった．我々の判断により特定した攻撃者設置サイトの FQDN57 個は，登録なし 1 個を除いて，全てが攻撃コードが設置された攻撃サイトという結果となった．

5.3　悪性 Web サイトの生存性

上記により特定した攻撃者設置サイトと改竄により悪性となった正規サイトについて，それらのサイトの生存性（A レコードの有無）について調査した．我々のドメイン登録日に基づいた判定では 57 個の FQDN が攻撃者設置サイトと判定されたが，ドメイン登録日が取得できなかったものを除いた 63 個の FQDN を，ここでは正規サイトと考える．一方で，MDL 登録情報に基づいて，93 個の FQDN が攻撃サイトと

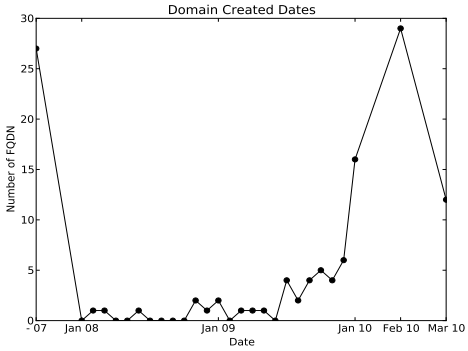

図 2: 巡回対象 URL（3 月 8 日）におけるドメインの登録日（2007 年以前，それ以降は月ごと）

判定されたが（5.2 節），攻撃サイトは攻撃者によって設置されると仮定すると，MDL 情報からは 93 個の FQDN が攻撃者設置サイトと判断される．また，悪性判断理由としてその他の項目を持つものを除外した，44 個の FQDN を MDL 情報による判断では正規サイトと考える．

以上の分類に基づいて，攻撃者設置サイトと正規サイトの生存性について調査した（表 3）．この結果から，正規サイトの方が生存性が高いことがわかる．これは，正規サイトは復旧後元通り通常運用されるが，攻撃者設置サイトはそのまま使い捨てられることが多いためと考えられる．また，MDL 情報により攻撃者設置サイトと判定された FQDN の方が，我々の判定によるものよりも生存性が高い．我々の判定で攻撃者設置サイトに含まれなかった 36 個の FQDN の内，24 個の FQDN に A レコードが割り当てられており，MDL により攻撃サイトと判定されたサイトの中に元々正規サイトであったものも含まれていたからと推測する．つまり，攻撃者は正規サイトの改竄により，誘導コードだけでなく攻撃コードも挿入する場合がある．よって，厳密に攻撃者設置サイトを特定する上で，我々の判定手法の方が有効性が高いと考える．

5.4 レジストラ情報の抽出

5.2 節において抽出した攻撃者設置サイトから，それらのドメインのレジストラ情報を取得した．表 4 にその他 63 ドメイン（我々の判断では元々正規サイト）の結果も併せて示す．

表 3: 悪性 Web サイトの生存性

	攻撃者設置サイト		正規サイト	
	FQDN	A あり	FQDN	A あり
我々の判定	57	11	63	46
MDL 判定	93	35	44	33

表 4 の結果からもわかるように，正規サイトと比べて攻撃者設置サイトのドメインのレジストラは種類数が少なく，特定のレジストラに集中している．これは，管理が不十分で不正目的で悪用しやすい，登録料が安いなどの理由で，攻撃者が不正活動を行う上で都合が良いからと考えられる．なお，攻撃者設置サイトにおける登録数上位のレジストラは中国やインドの会社のものであった．よって，悪性 Web サイトに対策を取る上で，各レジストラの登録管理を徹底化することが重要となる．実際，.cn や.ru などのいくつかのトップレベルドメインでは個人名義での登録が禁止されるなど，その管理が厳格化され，一定の効果を上げている [5].

表 4: 両サイトにおけるレジストラ情報の比較

	攻撃者 設置サイト	正規 サイト
登録レジストラ数	57	59
登録種類数	11	32
登録数 1 位	A 社/16 個	D,E 社/5 個
登録数 2 位	B 社/11 個	-
登録数 3 位	C,D 社/7 個	A,C,F,G 社/3 個

5.5 マルウェア配布サイトの特定

MDL 登録情報により攻撃サイトと判明した 93 個の FQDN から，サイトアクセス後に接続するサイトをマルウェア配布サイトと考えその抽出を行った．その結果，93 個中 78 個の FQDN については，新たにドメインの名前解決を行ってそこへ接続する挙動は観測されなかった．IP アドレスを直接指定して接続を行うものもいくつか存在したが，全体として攻撃サイトがマルウェア配布サイトの役割も担っている傾向にあった．攻撃者は管理の容易さからか，同じ 1 つのサイトで攻撃からマルウェア配布までを行って

いると言える．そのため，抽出できたマルウェア配布サイトの数は少なく，特徴抽出や攻撃サイトとの関連性分析までは至らなかった．今回着目した悪性 Web サイトの数はあまり多くなく，今後はより大規模な実験が必要となる．

6 おわりに

本研究では，近年のマルウェア感染経路の多様化によってその脅威が増加してきている，Web 感染型マルウェア対策のために，悪性 Web サイトのドメイン情報に着目しその活動傾向調査と関連性分析に取り組んだ．またその際，悪性 Web サイトを正規サイトの改竄によるものと攻撃者が自ら設置したものに分類するために，ドメインの登録日に着目した判断手法を提案し，その有効性を示した．さらに，特定した攻撃者設置サイトから，攻撃者によって使用されるレジストラに偏りがあることを発見した．

今後の課題として，より大規模な調査実験を通して，今回の調査結果の一般性を明らかにする必要がある．また，ドメイン情報だけでなく，使用される攻撃コードの特徴や，各種情報の時間的変化などに着目して，より詳細に悪性 Web サイト間の関連性を分析し，攻撃者の手口や意図を解明していく必要がある．

謝辞

本研究の一部は，独立行政法人情報通信研究機構が実施するインシデント分析の広域化・高速化技術に関する研究開発の支援を受けている．

参考文献

[1] Cova, M., Kruegel, C., and Vigna, G., “Detection and Analysis of Drive-by-Download Attacks and Malicious JavaScript Code”, *Proc of the 19th international conference on World Wide Web*, pp.281-290, 20010.

[2] Malware Domain List, `http://www.malwaredomainlist.com/mdl.php`

[3] Provos, N., Mavrommatis, P., Rajab, M.A., and Monrose, F., “All Your iFRAMEs Point to US”, *17th USENIX Security Symposium*, pp.1-15, 2008.

[4] Provos, N., McNamee, D., Mavrommatis, P., Wang, K., and Modadugu, N., “The Ghost In The Browser Analysis of Web-based Malware”, *Proc of the First Workshop on Hot Topics in Understanding Botnets (HotBots)*, 2007.

[5] 2010 年 上半期 Tokyo SOC 情報分析レポート, `http://www-935.ibm.com/services/jp/its/pdf/tokyo_soc_report2010_h1.pdf`

[6] Web ベースマルウェア調査サービス HoneyWhales, `http://honeywhales.com/`

[7] 秋山満昭, 川古谷裕平, 岩村誠, 伊藤光恭, “クライアントハニーポットを用いた Web 感染型マルウェアの実態調査”, MWS2008, 2008.

[8] 秋山満昭, 岩村誠, 川古谷裕平, 青木一史, 伊藤光恭, “クライアントハニーポットにおける攻撃検知手法の実装と評価”, MWS2009, 2009.

[9] 秋山満昭，佐藤一道，岩村誠，伊藤光恭, “Gumblar の長期観測による分析”，信学技報 IEICE Technical Report IA2010-13, 2010.

[10] 金子博一, 松木隆宏, 新井悠, “通信トラフィックの分析による Gumblar 感染 PC の可視化”, 信学技報 IEICE Technical Report IA2010-1, 2010.

[11] 畑田充弘，他：マルウェア対策のための研究用データセット ～MWS 2010 Datasets～，MWS2010（2010 年 10 月）

[12] 水谷正慶, 武田圭史, 村井純, “Web 感染型悪性プログラムの分析と検知手法の提案”, 電子情報通信学会論文誌 B Vol.J92-B No.10, pp.1631-1642, 2009.