Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

悪性 Web サイト間の関連性に着目した信頼性評価によるブラックリスト方式の検討

福島祥郎[†1,†2] 堀良彰[†1,†2] 櫻井幸一[†1,†2]

近年，Web ブラウザなどの脆弱性を突いてユーザにマルウェアを感染させる悪性 Web サイトの脅威が深刻化してきている．悪性 Web サイト対策にはそれらの URL やドメインのブラックリスト化が重要となるが，攻撃者はその回避のために URL やドメインを短期的に変更するため，たとえ未知の悪性 Web サイトであっても対応可能なブラックリストが重要となる．そこで本研究では，悪性 Web サイトが属す IP アドレスブロックとドメイン登録に用いたレジストラの関連性に着目し，それらの信頼性評価に取り組む．そして，信頼性の低い要素の組み合わせを用いたブラックリスト方式を提案し，その手法の有効性について評価を行う．

Blacklisting by Reliability Evaluation Based on Similarities among Malicious Web Sites

Yoshiro Fukushima,[†1,†2] Yoshiaki Hori[†1,†2] and Kouichi Sakurai[†1,†2]

The threats of malware infection via malicious Web sites which exploit vulnerabilities of Web browser are increasing. It is important to blacklist URLs or domains of the malicious Web sites for filtering them. However, attackers attempt to frequently change the URLs or domains to avoid the blacklist. Thus, a blacklist which can filter even unknown malicious Web sites is significant. In this paper, we focus on simiralities of IP address blocks and registrars used by malicious Web sites, and evaluate reliability of them. Then, we propose a blacklisting scheme using combinations of the two data with low reliability as the candidates for filtering, and evaluate effectiveness of our proposal.

†1 九州大学大学院　システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
†2 財団法人九州先端科学技術研究所
Institute of Systems and Information Technologies and Nanotechnologies

1. はじめに

近年，悪性 Web サイト経由でのマルウェア感染が問題となっている．従来はマルウェア感染手段として，ワームの感染活動のような，攻撃者が視点となって攻撃を開始し，OS などの脆弱性を突く「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 機能を持つ BB ルータの普及により，その脅威は減少しつつある．そのため，攻撃者は悪性 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」を用いるようになった[16]．受動型攻撃では主に，ユーザを攻撃者が用意した不正なサイトに誘導し，Web ブラウザやそのプラグインなどの脆弱性を突くことでマルウェア感染を試みる．この悪性 Web サイトへのアクセスから脆弱性攻撃，マルウェア感染までの一連の流れは drive-by-download 攻撃と呼ばれる．

悪性 Web サイト対策にはまず，drive-by-download 攻撃に関する悪性 Web サイトを検知することが重要となる．悪性 Web サイト検知には，Web を自動巡回し drive-by-download 攻撃の検知を行うクライアントハニーポットが有効である[15),16)]．そして，検知した悪性 Web サイトの URL やドメインをブラックリスト化し，ユーザのそれらへのアクセスを未然に防ぐことが重要になる．しかしながら，攻撃者はブラックリスト回避を目的として，URL やドメインを短期的に変更するため，たとえ未知の悪性 Web サイトであっても対応可能なブラックリストが必要となる．

そこで本研究では，悪性 Web サイト間の関連性に着目し，ドメイン情報の観点から信頼性評価を行い，その結果に基づいたブラックリスト方式を提案する．そのためにまず，ドメイン，IP アドレス，AS 番号，レジストラといったドメイン情報の観点から悪性 Web サイトの特徴を分析する．次に，悪性 Web サイト群が持つ IP アドレスブロックとレジストラ情報の関連性に着目し，各要素について信頼性評価を行う．その評価結果により，攻撃者の悪性 Web サイト設置に強く関連している要素データを特定する．そして，その要素データを用いたブラックリスト方式を提案し，その有効性について評価する．

2. 悪性 Web サイト経由のマルウェア感染

悪性 Web サイトを介した脆弱性攻撃から，マルウェアのダウンロードとインストールまでの一連の攻撃は drive-by-download 攻撃と呼ばれる[3)]．本節では，その攻撃手法と特徴，

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

対策の困難性について説明する．

2.1 drive-by-download 攻撃

drive-by-download 攻撃に関連する悪性 Web サイトは，主に誘導サイト，攻撃サイト，マルウェア配布サイトの 3 つからなる（図 1）．ただし，攻撃サイトがマルウェア配布サイトの役割を担う場合もある[16),21)]．誘導サイトは元々は正規に通常運用されてきた Web サイトである．しかし，攻撃者による SQL インジェクションや FTP アカウントの盗用などにより Web ページが改ざんされ，攻撃サイトへ誘導するための不正なコードが挿入される．攻撃サイトは，誘導サイトによって誘導されてきたユーザに対して，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を送信する．ユーザ側の環境で適切な脆弱性対策が行われていない場合は，その攻撃が成功し，マルウェア配布サイトを介して，ユーザの意図しないところで自動でマルウェアがダウンロード・インストールされる．

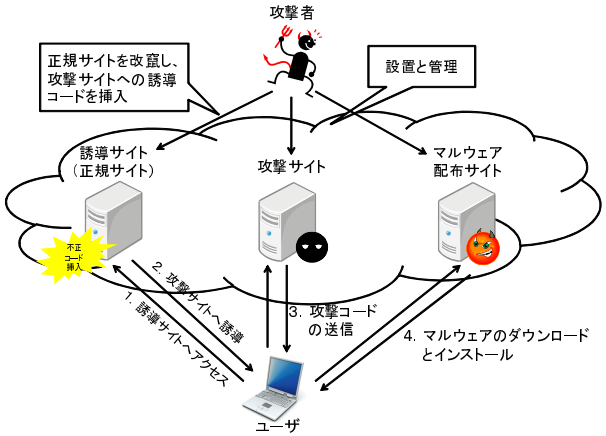

図 1 drive-by-download 攻撃の流れ

2.2 対策の困難性

悪性 Web サイト経由のマルウェア感染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[2)]，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[13)]．さらに，誘導サイトは元々は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[9)]．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[9),11)]．

3. 関 連 研 究

drive-by-download 攻撃を検知するために，様々なクライアントハニーポットが考案されている[1),6)]．クライアントハニーポットは，ユーザのブラウザ環境をエミュレート，もしくは実環境を用いて Web を自動巡回し，脆弱性を突くような攻撃が観測された場合に，そのサイトを悪性 Web サイトとして検知する．また，クライアントハニーポットを利用し，悪性 Web サイトの活動の時系列的な変化や悪用される脆弱性の種類などに関する実態調査も行われている[9),10),16)]．こういった悪性 Web サイトの URL 等に関する情報は，Malware Domain List（MDL）[8)] や HoneyWhales[14)] といった Web サイトで公開されている．

悪性 Web サイト対策には，クライアントハニーポット等を利用して取得した悪性 Web サイトの URL やドメインをブラックリスト（以下，BL）化することが重要になる．事前に悪性 Web サイトを BL に追加しておくことで，ユーザのそれらへのアクセスを未然に遮断することができる．しかしながら，BL はすでに悪性と判定された既知のものにしか対応できない．さらに，攻撃者は BL 回避のために，悪性 Web サイトの URL やドメインを短期的に変更，あるいは新規に悪性 Web サイトを設置しようとする．そのため，既存の BL 方式では，増え続ける悪性 Web サイトに追随できず，BL の効率的な更新が必要となる．そういった課題を解決するために，既知の悪性 Web サイトに関する情報を基に，それと高い関連性を持つ Web サイトを BL 候補とする研究が行われている[4),7),17),18)]．秋山らは悪性 Web サイト URL の部分的変化に対応可能な，悪性 Web サイト URL の類似性に着目した BL 手法を提案している[17),18)]．しかし，これらの手法は悪性 Web サイト間の URL の類似性に基づいているため，攻撃者が毎回全く異なる URL（ドメインやディレクトリパス）を用いることで回避されてしまう．また，Felegyhazi らは，ドメインの登録時間の同時性に着目し，すでに悪性と判断されたサイトと同じ時間にドメインが登録されたサイトを BL 候補とする手法を提案している[4)]．ただし，本研究が drive-by-download 攻撃に関連する悪性 Web サイトを対象としているのに対し，この研究は主にフィッシングサイト等の詐欺サイトを対象としている．加えて，同じドメイン登録時間を持つ Web サイト特定のために，そ

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

の登録に携わるレジストラの協力が必要であり，全てのレジストラから同じような協力を得ることは難しいと言える．さらに，Ma らは不正な Web サイトの URL の文字列的特徴と，そのサイトの IP アドレスや whois 情報，DNS レコード情報等を用いた学習により，それらの検知を行う手法を提案している[7]．こちらも対象はフィッシングサイトやスパムに関連するサイトであり，我々の対象とは異なる．しかしその一方で，こういった関連研究のアプローチは，我々が提案する BL 方式の精度を向上させるために利用できると考えられる．

本研究では，悪性 Web サイトの IP アドレスやレジストラといったドメイン情報に着目し，それらについて信頼性評価を行う．ドメイン情報の信頼性評価に関する研究として，須藤らは CCC DATAset 2010 の攻撃元データ[20] から攻撃元 IP アドレスを抽出し，それらのアドレスブロックや AS 等について信頼性評価を行っている[19]．しかし，CCC DATAset 2010 の攻撃元データは，サーバ型のハニーポットで収集されたものであり，基本的に能動型攻撃に関連する情報しか含んでいない．一方，本研究では受動型攻撃によりマルウェア感染を試みる悪性 Web サイトを対象としており，その特徴や傾向等に違いが見られると考えられる．また，Stone-Gross らは長期的な不正活動に関わる ISP を悪性と判断する手法を提案し，実際に各 AS の悪性度について評価を行っている[12]．そして，特に悪性である可能性が高い AS については，その情報を Web 上で公開している[5]．こういった情報は，我々が信頼性評価に取り組む上で有益なものであると言える．

4. ドメイン情報を用いた特徴分析

4.1 本研究の目的

本研究では，まず攻撃者の悪性 Web サイト設置における手口とその特徴を解明するために，ドメイン情報の観点から特徴分析を行う．特に，悪性 Web サイト間の関連性に着目し，攻撃者によって頻繁に利用される，AS，IP アドレスブロック，IP アドレス，レジストラといったドメイン情報に関する要素データを特定する．ISP やホスティング事業者の中には，不正目的での利用を容認しているところもあり[12][19]，攻撃者によって頻繁に利用される要素データを信頼性が低いと判断し，それに関連する Web サイトを広く遮断することが，効果的な悪性 Web サイト対策につながると考えられる．そのため本研究では，ドメイン情報の信頼性評価結果に基づいて，攻撃者の悪性 Web サイト設置に強く関連している要素データをブラックリスト（BL）に追加する手法を提案し，その評価を行う．

4.2 着目するドメイン情報

本研究では，ドメイン情報として以下の表 1 に示す 5 つの要素を考える．これら 5 つの要素は，図 2 のような階層構造で表現することができる．1 つの AS にはいくつかのアドレスブロックが割り当てられ，そのアドレスブロック以下に IP アドレス群が存在する．そして，その IP アドレスに対してドメインが割り当てられ，そのドメインはとあるレジストラを介して登録される．

このようなドメイン情報の階層構造に着目して特徴分析することで，攻撃者が積極的に利用する AS，IP アドレスブロック，IP アドレス，レジストラといった情報の関連性を明らかにすることができる．

表 1 着目するドメイン情報の要素

AS 番号	IP アドレスブロック	IP アドレス	ドメイン名	レジストラ

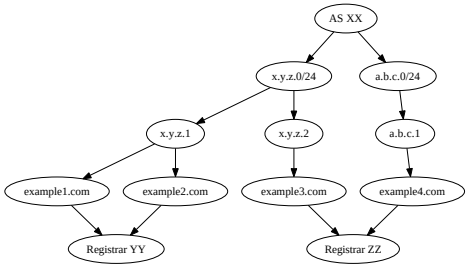

図 2 ドメイン情報の階層構造の例

表 2 図 2 における各要素の次数

	次数
AS XX	2
x.y.z.0/24	2
a.b.c.0/24	1
x.y.z.1	2
x.y.z.2	1
a.b.c.1	1
Registrar YY	2
Registrar ZZ	2

4.3 特徴分析手法

まず，悪性 Web サイト群から，それらの AS 番号，IP アドレスブロック，IP アドレス，ドメイン名，レジストラ情報を抽出し，図 2 にならって，ドメイン情報の階層構造グラフを作成する．ただし，この時のグラフの根は AS 番号とする．次に，各要素ごとにその要素以下に存在する子ノード数（出次数）を次数と定義し，その値を求める．ただし，ドメイン名とレジストラの関係は 1 対 1 であるため，ドメイン名については次数は考慮しない．また，レジストラについてはその要素以下に存在する子ノード数ではなく，その親ノード数（入次数）を次数とみなす．

図 2 における階層構造グラフから，各要素の次数を求めると表 2 のようになる．この場

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

合，次数が大きい要素ほど，攻撃者の悪性 Web サイト設置に強く関連している要素と考えることができる．例えば，攻撃者が同じ IP アドレスブロック以下に異なる IP アドレスを持つ悪性 Web サイトを多数設置した場合は，その IP アドレスブロックの次数が大きくなる．また，ドメイン情報の階層構造に着目することで，悪性 Web サイト間の関連性を容易に把握することが可能となる．本研究では，悪性 Web サイト群からドメイン情報の階層構造グラフを作成し，次数に関する統計情報を算出するとともに，特に次数が高い要素に関して詳細に説明する．

4.4 データセット

本研究では，Malware Domain List (MDL)[8] で公開されている不正な Web サイトに関する情報から，drive-by-download 攻撃に関連すると思われるおよそ 4000 個の悪性 Web サイトを抽出し，それを用いて特徴分析を行う．ただし，取得した情報は 2010 年 1 月から 2010 年 11 月上旬に公開されたものであり，また，同じ IP アドレスと同じドメイン名を持つ項目については，同じ 1 つの情報として集約した．drive-by-download 攻撃に関連するかどうかは，その Web サイトが MDL に登録された理由を参照し，“exploit” や “redirect” といったキーワードを含むかどうかで機械的に判断した．

MDL で公開される情報は，主に悪性 Web サイトの URL と IP アドレス，AS 番号，悪性と判断した理由などである．レジストラ情報は含まれていないため，ドメイン名を基に whois 情報を参照して新たに取得した．ただし，多数のドメインについてそのレジストラ情報が，whois 情報に正しく登録されていない，あるいは時間経過のためか消失していた．また，IP アドレスブロックに関しては IP アドレスを基に算出し，今回は簡単のため全て/24 サイズで考える（例えば，1.1.1.1 に対して，1.1.1.0/24 とする）．

4.5 特徴分析結果

4000 個ほどの悪性 Web サイトに関する情報から，前述の方法にならってドメイン情報の階層構造グラフを構築し，AS 番号，IP アドレスブロック，IP アドレス，レジストラの 4 つの要素についてその次数を算出した．

まず，図 3〜図 6 に，4 つの要素における次数の累積分布グラフを載せる．図 3〜図 6 より，次数が 1 であるものが大半ではあるが，次数が大きいものも存在していることがわかる．それぞれの図から，次数が 10 以上であるような要素の割合を求めると，およそ 3%，1.5%，3.5%，20%であった．こういった次数が大きい要素データは，攻撃者の悪性 Web サイト設置に強く関連していると考えられる．

次に，表 3 に各要素ごとの次数の平均と分散の値を載せる．AS，IP アドレスブロック，

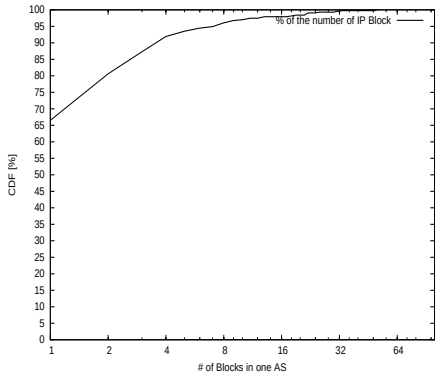

図 3 AS 以下に存在する IP アドレスブロック数の累積分布

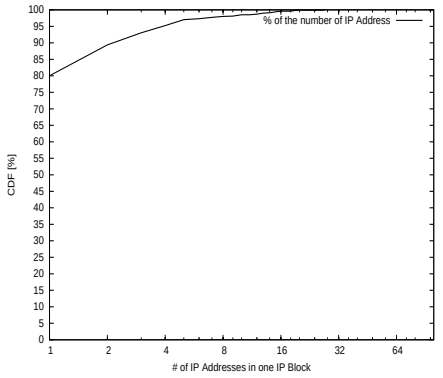

図 4 IP アドレスブロック以下に存在する IP アドレス数の累積分布

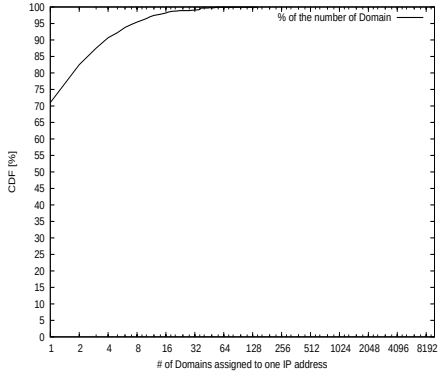

図 5 IP アドレスに割当てられたドメイン数の累積分布

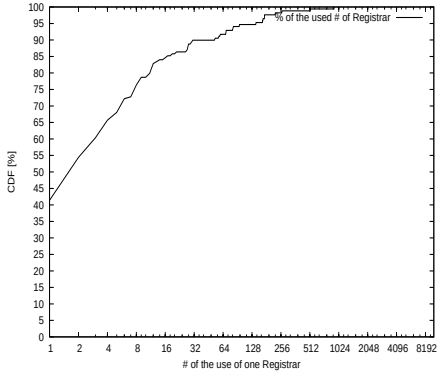

図 6 各レジストラの使用回数の累積分布

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

IP アドレスについては次数の平均はそれほど大きくない一方で，レジストラについてはその値が大きい．これは，レジストラの種類が他の要素と比べて少ないため，1 レジストラあたりの使用回数が多くなっているからと思われる．次に分散を見ると，レジストラの分散値がかなり大きいことがわかる．これは，特定のレジストラが攻撃者によって集中的に使用されているからである．実際，いくつかの研究により攻撃者が不正な Web サイト設置時に用いるレジストラには偏りがあることがわかっている[4),7),21)]．IP アドレスの分散もやや大きいが，これはいくつかの攻撃者がドメインを使い捨てる傾向にあり，同じ IP アドレスに多数のドメインが割り当てられるためである．AS の分散も同様にやや大きいが，これは各 AS が所有するアドレスブロックの数がそもそも異っていることが要因だと思われる．IP アドレスブロックについては分散は大きくはないが，図 4 が示すように平均から大きく外れるものもわずかながら存在している．

表 3 各ドメイン情報における次数の平均と分散（レジストラは Not Found を除く）

	AS	IP アドレスブロック	IP アドレス	レジストラ
次数の平均	2.349	1.630	2.453	15.510
次数の分散	17.775	4.617	32.920	1969.541

最後に，攻撃者の悪性 Web サイト設置に強く関連していると思われるドメイン情報について説明する．表 4 は AS 番号，IP アドレスブロック，IP アドレス，レジストラの 4 つのドメイン情報要素のそれぞれについて，上位 5 つの次数を持つものを示したものである（ただし，一部伏字にしている）．AS 番号については，それぞれが所有するアドレスブロック数が異なるため一概には言えないが，こういった要素が攻撃者の悪性 Web サイト設置に強く関連していると考えられる．特に，IP アドレスブロック “a.b.59.0/24” については，ユニークな IP アドレスを持つ悪性 Web サイトの数が多い（27 個）．また，その中でも IP アドレス “a.b.59.55” を持つ Web サイトはドメインの使い捨てにより，多数のドメインが割り当てられている（55 個）．このような IP アドレスブロック以下に存在する Web サイトは信頼性が低いと言える．同様に，攻撃者によって集中的に利用されるレジストラを介して登録されたドメインを持つ Web サイトも，信頼性が低いと言える．

表 4 各ドメイン情報における次数上位 5 つの詳細情報（カッコ内はその次数）

	1	2	3	4	5
AS	AS 1 (50)	AS 2 (33)	AS 3 (31)	AS 4 (25)	AS 5 (22)
IP アドレスブロック	a.b.59.0/24 (27)	c.d.152.0/24 (20)	e.f.179.0/24 (19)	g.h.190.0/24 (19)	i.j.143.0/24 (16)
IP アドレス	k.l.127.197 (142)	a.b.59.55 (55)	m.n.240.211 (53)	o.p.159.35 (52)	g.h.191.41 (45)
レジストラ	Not Found (1435)	レジストラ W (262)	レジストラ X (262)	レジストラ Y (255)	レジストラ Z (224)

5. 信頼性評価

5.1 評価対象とするドメイン情報

評価対象としては，AS 番号，IP アドレスブロック，IP アドレス，レジストラの 4 つが考えられる．しかしながら，AS 番号については，AS ごとに所有する IP アドレスブロック数が異なるため，悪性 Web サイトが存在するアドレスブロックを多数所有しているという判断だけでは，単純にその AS の信頼性が低いと判断することは難しい．また，AS という広い範囲で考えると，悪性でない多数の正規の Web サイトを誤って信頼性が低いと判断する恐れもある．同様に，IP アドレス単体で考えた場合は，悪性であると判断されるのはドメインの使い捨てにより，多数のドメインが割り当てられた IP アドレスに限られる．そのため本研究では，信頼性の評価対象として IP アドレスブロックとレジストラに着目する．

信頼性評価にあたって，IP アドレスブロックとレジストラの AND 結合を取った組み合わせを評価対象の要素とする．AND 結合を取る理由は，各要素単体で評価した場合に起こりえる，正規の Web サイトを誤って信頼性が低いと判定する誤検知を減らすためである．仮に，ある IP アドレスブロックに属する悪性 Web サイトの数が多くても，そのブロック全体を信頼性が低いと判断すると，そのブロック内に存在する悪性でない Web サイトも誤って遮断する恐れがある．そこで，攻撃者は悪性 Web サイト設置時に特定のレジストラを集中的に利用するという特徴に着目し，IP アドレスブロックとレジストラの 2 つの AND 結合を取ることで，誤検知の可能性を抑える．特に，同じ IP アドレスブロック内に存在する悪性 Web サイトは，同じ攻撃者によって設置された可能性が高く，より特定のレジストラが使用されやすいと考える．

5.2 IP アドレスブロックとレジストラの組み合わせによる信頼性評価

まず，悪性 Web サイトが存在する全ての IP アドレスブロックを抽出し，その次数（つ

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

まり，そのブロックに属す悪性 Web サイトのユニークな IP アドレス数）を求める．次に，その IP アドレスブロックに属す全ての悪性 Web サイトのドメインから，使用されるレジストラの種類とその次数（使用回数）を求める．ここで，レジストラ使用回数については全ての悪性 Web サイトで使用された回数ではなく，着目するアドレスブロック以下に存在する悪性 Web サイトによって使用された回数とする．これら 2 つの要素データと次数の組み合わせを以下のような形式で表し，その次数の値に基づいて信頼性評価に取り組む．

$$(IPBlock_i, Registrar_j) = (v_i, v_j)$$

例えば図 2 の場合は，$(x.y.z.0/24, RegistrarYY) = (2, 2)$，$(x.y.z.0/24, RegistrarZZ) = (2, 1)$，$(a.b.c.0/24, RegistrarZZ) = (1, 1)$ のような評価値を持つ．

実際に，前章で用いた悪性 Web サイト群について，IP アドレスブロックとレジストラの組み合わせによりその信頼性評価を行った．図 7 にその結果を示す．ただし，次数が大きいほど信頼性が低いと判断するため，ある値よりも大きい次数を持つ組み合わせのみを考える．今回は，IP アドレスブロックの次数が 5 以上かつレジストラの使用回数が 10 以上の組み合わせのものだけを考えることにする．また，レジストラ情報が登録されていないものについては図から除外している．

図 7 では，上述の閾値設定のもと，計 27 通りの IP アドレスブロックとレジストラの組み合わせについて信頼性が低いと判断された．今回，閾値の設定による違いについては特に考慮はしないが，おおよそこれらの組み合わせを持つような Web サイトは信頼性が低いと考えられる．こういった IP アドレスブロックを所有する ISP やドメイン登録を受け付けるレジストラはセキュリティ対策が十分に考慮されていない，あるいは利用コストが低い，といった理由で攻撃者によって頻繁に利用されていると思われる．このような信頼性が低い IP アドレスブロックとレジストラを用いた Web サイトを広く遮断することで，未知の悪性 Web サイト対策につなげることができると考える．

6. 信頼性評価に基づいたブラックリスト方式

6.1 ブラックリスト作成方法

前章の考えのもと，悪性 Web サイト群からその IP アドレスブロックとレジストラの AND 結合を取った組み合わせについて信頼性評価を行い，信頼性が低いと判断された要素を BL に追加する．例えば，図 7 において，$(a.b.59.0/24, RegistrarB)$ を 1 つの要素として BL に追加する．ある Web サイトの信頼性が低いかどうかを判断する際は，そのサイトが属する IP アドレスブロックと使用レジストラの組み合わせを BL 内の要素と比較し，一致するものがあればそのサイトを信頼性が低いと判断し遮断する．この場合，例えば IP アドレス $a.b.59.100$ を持ち，レジストラとして $RegistarB$ を使用している全ての悪性 Web サイトを遮断する．

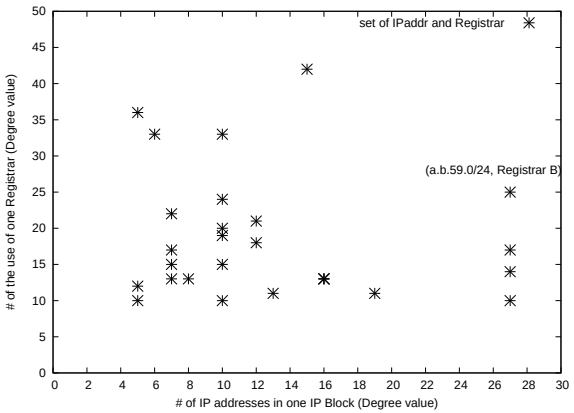

図 7 IP アドレスブロックとレジストラの組み合わせにおける次数分布

6.2 有効性評価

信頼性が低い IP アドレスブロックとレジストラの組み合わせを用いた BL 方式の有効性について評価を行う．有効性を評価するにあたって重要となるのは，ある時点で追加した BL 要素と同じ要素を持つ悪性 Web サイトが，将来にわたってどれくらい出現する可能性があるかである．同じ要素を持つ悪性 Web サイトが，長期間にわたって継続的に設置されるようならば，早期にその要素を BL に追加することで，未知の悪性 Web サイトであっても未然に防ぐことができる．

前章で信頼性が低いと判断した 27 個の組み合わせについて，それらと同じ要素を持つ悪性 Web サイトの出現期間に着目し，有効性評価に取り組んだ．表 5 に，図 7 における 27 個の IP アドレスブロックとレジストラの組み合わせに合致する悪性 Web サイトの，最初に出現した日から最後に出現した日までの出現期間日数の統計を示す．最小で出現期間が 6 日しかない組み合わせが存在したが，平均でおよそ 95 日の間，同じ組み合わせを持つ悪性 Web サイトが継続的に出現していた．比較的長期間，同じ組み合わせを持つ悪性 Web サイトが出現していたと言える．

しかし，出現期間が長期に亘る一方で，信頼性が低い組み合わせと同じ要素を持つ悪性

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

Web サイトの数は，1 つの組み合わせあたり平均でおよそ 19 個と必ずしも多いとは言えない結果であった．図 8 に，信頼性が低いと判断した 27 個の組み合わせと同じ要素を持つ悪性 Web サイトの出現日時をまとめる．横軸は出現日時（日ごと，2010 年 1 月 1 日を 0，1 月 2 日を 1，以下同様），縦軸は組み合わせ ID（1〜27 の 27 通り）を意味している．例えば，前述の組み合わせ $(a.b.59.0/24, RegistrarB)$ は ID23 であり，それと同じ要素を持つ悪性 Web サイトは比較的長期間（5 カ月程度）出現しているが，そのサイト数自体は 25 個であった．つまり，同じ要素を持つ悪性 Web サイトは長期的に出現はしていたが，比較的散発的な出現であった．1 つの信頼性の低い組み合わせだけで対応できる悪性 Web サイトの数は多いとは言えないが，それでも多数の組み合わせに着目することで，より対応範囲の広い BL を実現できると考える．特に，このような信頼性の低い組み合わせを早期に BL に追加することで，より多くの悪性 Web サイトに対して効果的に対策することが可能となる．

表 5 図 7 の組み合わせと同じ要素を持つ悪性 Web サイトの出現期間

	平均	最大	最小
出現期間（日数）	95.22	221	6

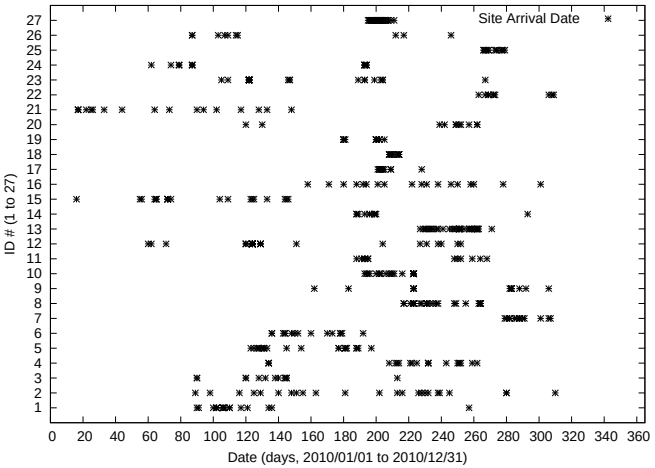

図 8 信頼性が低い 27 個の組み合わせと同じ要素を持つ悪性 Web サイトの出現日時（日ごと）

6.3 考　　察

6.3.1 本研究における課題とその解決策

上記の評価結果から，我々の信頼性評価に基づく BL 方式は一定の有効性を持つと言える．しかしながら，信頼性評価によって選定できた，信頼性の低い IP アドレスブロックとレジストラの組み合わせはわずか 27 通りであり，これだけでは全ての悪性 Web サイトに対応することは難しい．そのため，評価対象要素の追加や他の評価手法の検討が必要となる．今回は IP アドレスブロックとレジストラの 2 つに着目したが，今後はドメイン登録日の新しさや，Web サイトのページ内容の類似性などを評価対象とすることを検討している．また今回は，要素間の AND 結合を取って信頼性評価を行ったが，OR 結合にも着目することで，BL の性能を向上できる可能性がある．

さらに，本研究で用いた悪性 Web サイト情報は MDL で公開された限定的なものであり，その数も 4000 個程度と決して多くはない．そのため，本研究の結果が悪性 Web サイト全体の特徴を反映できているとは必ずしも言えない．今後は，クライアントハニーポットを用いる等してより多くの悪性 Web サイトについて信頼性評価に取り組む必要がある．

6.3.2 提案手法の回避策とその実現性

我々の BL 方式の回避を目的として，攻撃者が毎回異なる IP アドレスブロックとレジストラを用いて悪性 Web サイトを設置する可能性がある．この場合，悪性 Web サイトは様々な IP アドレスブロックに分散し，また様々なレジストラが使用されるため，信頼性の低い要素を特定することは難しい．一方で，こういった分散設置方式は攻撃者の利便性の低下にもつながると考えられる．攻撃者が特定の IP アドレスブロックやレジストラを用いる理由としては，使用する ISP やレジストラの利用コストが安い，あるいはセキュリティ対策が十分に行われていない，などが考えられる．そのため，分散設置方式は，攻撃者にとっても金銭的コストを増加させ，また不正利用自体を難しくさせると考えられる．実際，.cn や.ru など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，管理がより厳格化されるようになり，一定の成果を挙げている[13]．

7. おわりに

悪性 Web サイト経由でのマルウェア感染が深刻化し，その効果的な対策手法が求められている．その対策の 1 つとして，悪性 Web サイトの URL やドメインのブラックリスト（BL）化が挙げられる．しかし，そういった BL は既知の悪性 Web サイトにしか対応できない．

Vol.2011-DPS-146 No.38
Vol.2011-CSEC-52 No.38
2011/3/11

加えて，攻撃者は BL 回避のために，悪性 Web サイトの URL やドメインを短期的に変更，あるいは新規に悪性 Web サイトを設置する傾向にある．そのため，たとえ未知の悪性 Web サイトであっても対応可能な BL 方式が重要となる．そこで本研究では，悪性 Web サイトが属す IP アドレスブロックとドメイン登録のレジストラの関連性に着目し，その信頼性評価を行うことで攻撃者の悪性 Web サイト設置に強く関連している要素の特定に取り組んだ．そして，信頼性が低い IP アドレスブロックとレジストラの組み合わせを用いた BL 方式を提案した．評価実験により，信頼性が低い要素の組み合わせを持つ悪性 Web サイトが，比較的長期間にわたって継続的に出現しており，提案手法の一定の有効性が確認できた．

しかしながら，今回特定できた信頼性の低い要素の組み合わせの数は少なく，全ての悪性 Web サイトに対応することは難しいと考える．今後は，他の信頼性評価要素の追加と，それらの組み合わせ方の工夫などに取り組み，提案手法の改善をはかる必要がある．加えて，より多くの悪性 Web サイトデータを用いて実験していく必要がある．また，提案手法の性能について，悪性でない Web サイトに対する誤検知の観点から評価を行う必要もある．

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) Capture-HPC, `https://projects.honeynet.org/capture-hpc`
2) 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, 2010.
3) Egele, M., Kirda, E., and Kruegel, C. “Mitigating Drive-by Download Attacks: Challenges and Open Problems,” *Proceedings of Open Research Problems in Network Security Workshop (iNetSec)*, pp.52-62, 2009.
4) Felegyhazi, M., Kreibich, C., and Paxson, V., “On the Potential of Proactive Domain Blacklisting,” *Proceedings of the 3rd USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET’10)*, 2010.
5) FIRE: FInding RoguE Networks, `http://maliciousnetworks.org/`
6) HoneyC, `https://projects.honeynet.org/honeyc`
7) Ma, J., Saul, J.K., Savage, S., and Voelker, G.M., “Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs,” *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’09)*, pp.1245-1254, 2009.
8) Malware Domain List, `http://www.malwaredomainlist.com/mdl.php`
9) Provos, N., Mavrommatis, P., Rajab, M.A., and Monrose, F., “All Your iFRAMEs Point to US,” *17th USENIX Security Symposium*, pp.1-15, 2008.
10) 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.
11) Rajab, M.A, Ballard, L., Mavrommatis, P., Provos, N, and Zhao, X. “The Nocebo Effect on the Web: An Analysis of Fake Anti-Virus Distribution,” *Proceedings of the 3rd USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET’10)*, 2010.
12) Stone-Gross, B., Kruegel, C., Almeroth, K., Moser, A., and Kirda, E. “FIRE: FInding Rogue nEtworks,” *Computer Security Applications Conference (ACSAC’09)*, pp.231-240, 2009.
13) 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`
14) Web ベースマルウェア調査サービス HoneyWhales, `http://honeywhales.com/`
15) 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.
16) 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.
17) 秋山満昭, 八木毅, 伊藤光恭, “悪性 URL 群の木構造に着目した URL フィルタリングの粒度決定,” 信学技報, vol.110, no.266, ICSS2010-53, pp.53-58, 2010.
18) 秋山満昭, 八木毅, 伊藤光恭, “検索エンジンを利用した悪性サイト url の近隣探索によるブラックリストの拡張,” CSS2010, 2010.
19) 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,” MWS2010, 2010.
20) 畑田充弘, 中津留勇, 秋山満昭, 三輪信介, “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.
21) 福島祥郎, 堀良彰, 櫻井幸一, “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,” MWS2010, 2010.