Vol.2017-DPS-170 No.23
Vol.2017-CSEC-76 No.23
2017/3/3

不可視Webコンテンツ特徴に基づく Drive-by Download攻撃の検知と調査支援ツールの提案

荻野 貴大[1,a)]　高田 哲司[1,b)]

概要：サイバー攻撃の多くは「見えない化」されている．攻撃の動機が金銭的利益であり，より多くの利益を得るため，利用者やシステム管理者に気付かれることなく不正行為を継続実行できることが望ましいからである．この傾向は，Web を通じた攻撃にもあてはまる．例えば Drive-by Download 攻撃では，攻撃用コンテンツを微小サイズで埋め込んだり，閲覧可能領域外に配置するなどの手法がとられる．これに対し，URL の文字列特徴や HTTP のヘッダ情報に注目した従来の検知手法では対応に限界があり，前述のような特徴を攻撃の検知に活用できていないと言える．そこで本研究では，Web コンテンツ内の情報を攻撃の検知に活用し，不可視化された Web コンテンツ特徴に基づく検知手法について提案する．また，当該攻撃の調査を補佐しうる可視化システムについても議論する．

キーワード：Web セキュリティ，Drive-by Download 攻撃，不可視 Web コンテンツ，Web ページ改ざん，改ざん Web ページ検知，情報視覚化

Compromised Web page Detection Scheme based on features of invisible elements and its Investigation Support Tool by Visualization

Ogino Takahiro[1,a)]　Takada Tetsuji[1,b)]

Abstract: Cyber attacks have become “invisible”, because attackers are attempt to continue malicious activities for a longer period without being noticed by other person. It is said that it will bring more benefits to attackers, and this trend is also seen in a malware infection through a web page. In the case of Drive-by Download attack, attackers make use of following techniques that a user can not notice a compromised web page: a) embedding tiny size elements, b) putting element in outside of viewable area, and so on. Previous works for attack web page detection schemes do not make use of above features to detect attack pages. In this work, we propose an attack web page detection scheme that make use of the features in invisible elements in compromised web page. We also proposed a visual tool to support attack investigation in the compromised web page.

Keywords: Web security, Drive-by Download attack, Invisible web content, Web page alteration, compromised web page detection, information Visualization

1. はじめに

Web サイト閲覧時のクライアントを標的とした Drive-by Download 攻撃（以降，DbD 攻撃と記す）は Web 利用における大きな脅威である．IBM Security Services の「2016 年上半期 Tokyo SOC 情報分析レポート」[1] から，2016 年上半期の DbD 攻撃の検知数は 2015 年下半期と比較し減少したが，2015 年下半期の検知数は 2013 年上半期以降最多の検知数であったことがわかる．

DbD 攻撃の概要を図 1 に示す．攻撃者によって改ざんされた Web サイトが攻撃の起点となるケースが多く，ユーザは中継サイト・マルウェア配布サイトへと誘導され，マルウェア感染に至る．以降では，各サイトについて説明する．

[1] 電気通信大学
University of Electro-Communications, Tokyo, Japan
[a)] takahiro.ogino@uec.ac.jp
[b)] zetaka@computer.org

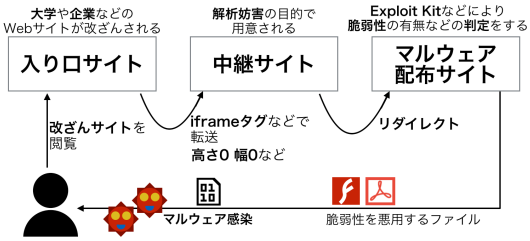

図 1 DbD 攻撃の概要図

(1) 入り口サイト
DbD 攻撃の起点となる Web サイトのことである．改ざん Web サイトには，iframe タグや script タグなどが挿入される．

(2) 中継サイト
解析妨害の目的で用意される Web サイトのことである．一般的に，Web ブラウザや Web ブラウザプラグインに特定の脆弱性がある場合のみ，マルウェア配布サイトへと誘導する．また，そのための JavaScript も難読化されることが多いため，攻撃手法の解析が困難となっている．

(3) マルウェア配布サイト
Web ブラウザや Web ブラウザプラグインなどの脆弱性が悪用され，ユーザの承認なしにマルウェアがダウンロードされる．

本研究では，攻撃の初期フェーズでの検知を目的とし，入り口サイトにおける不可視 Web コンテンツに着目した検知手法を提案する．不可視 Web コンテンツとは，微小サイズによる埋め込み・閲覧領域外への配置・カスケーディングスタイルシート（以降，CSS と記す）の設定の 3 手法により，ユーザからは見えないよう作り込まれる Web コンテンツのことである．

第 2 章では関連研究の問題点を整理し，本研究の利点について述べる．第 3 章では外部 Web サイトの投稿記事情報をもとに，攻撃に悪用される不可視 Web コンテンツに関する調査結果をまとめる．第 4 章では検知システムの実装について述べ，第 5 章ではその有用性について議論する．第 6 章では攻撃の調査を補佐しうる可視化システムについて述べ，第 7 章ではその有用性について議論する．第 8 章では本研究に関する考察について述べ，第 9 章では本研究のまとめをする．

2. 関連研究と本研究の提案手法

本章では，関連する先行研究を紹介するとともに，その問題点をふまえた提案手法について述べる．

2.1 関連研究

2.1.1 URL の文字列特徴に着目した検知手法

DbD 攻撃では，Exploit Kit と呼ばれる攻撃 Web サイト構築ツールが利用される傾向にある．笠間ら [3] は，Exploit Kit を利用した Web サイトの URL 文字列特徴に着目した検知手法を提案している．しかし，新種の Exploit Kit を利用した DbD 攻撃や，Exploit Kit を利用せず構築された攻撃サイトについては検知できない問題がある．

2.1.2 HTTP ヘッダ情報に着目した検知手法

酒井ら [4] は，HTTP ヘッダの PHP や取得コンテンツに関する情報に着目した検知手法を提案している．しかし，広告コンテンツとして Flash コンテンツが取得される Web サイトも多いため誤検知を誘発する問題がある．また，マルウェア配布サイトにおける特徴を利用しているため，攻撃者によりクローキングなどの検知回避技術が用いられる場合，酒井らの検知手法はクライアントハニーポットによる検知手法に応用できない．

2.1.3 Web コンテンツに着目した検知手法

望月ら [5] は，Web サイトの改ざん前と後の Web サイトの構成情報を比較することで，悪性 Web サイトの検知を試みている．Web サイトの構成情報の変化に着目することで，攻撃者 Web サイトまで誘導する iframe タグや script タグなどを見つけることができる見込みがある．しかし，望月らの検知手法では，改ざん前の情報が必要であり，初めてアクセスする Web サイトについては検知ができない．

西田ら [6] は，難読化 JavaScript に着目した攻撃の検知手法を提案している．98.84%と高精度で，良性と悪性の JavaScript を判別することが可能であるが，難読化 JavaScript を利用しない DbD 攻撃の検知はできない．

田村ら [7] は，width あるいは height が 0 で指定される Web コンテンツを DbD 攻撃の判定処理に利用している．これは本研究と同様に不可視 Web コンテンツに着目していると言えるが，それ以外の不可視 Web コンテンツについては考慮されていない．また，Web コンテンツのサイズ情報はあくまでも判定処理の一特徴として利用されているため，不可視 Web コンテンツに着目することにより，どの程度攻撃を検知できるかについては明らかでない．

2.2 本研究における提案手法

本研究では，入り口サイトの検知を目指し，不可視 Web コンテンツに着目した検知手法を提案する．この提案内容を考案した理由は 2 つある．1 つは入り口サイトの検知に着目した例が少ないためである．文献 [7] で一部の特徴を利用しているものの，入り口サイトの特徴を調査し，その結果をふまえた検知システムの提案が，我々の知るかぎり見当たらないためである．もう 1 つは，他の検知手法を補完する仕組みとなり，攻撃対策の強化になると考えるからである．既存の検知手法に加えて，入り口サイトを対象とした検知システムが実現できれば，複層による攻撃検知が可能になると考える．

3. 攻撃に悪用される不可視 Web コンテンツに関する調査

本研究では，入り口サイトで利用される攻撃者の手口を把握する目的で調査を実施した．以降では，その調査方法と調査結果について述べる．

Malware Traffic Analysis.net[2]（以降，MTAnet と記す）の投稿記事を対象に，調査を実施した．MTAnet は DbD 攻撃発生時の pcap ファイルや改ざん事例などの情報提供 Web サイトであり，関連研究 [10][11] において悪性データセットとしても利用されている．

MTAnet に掲載された 2016 年 7 月 1 日から 2016 年 12 月 31 日（2016 年下半期）の投稿記事を対象に調査を行った．この期間における投稿総件数は 229 件であったが，本調査目的に適さないと考える以下の投稿記事については除外した．

- malspam に関する投稿 43 件
  メールを悪用した攻撃に関する投稿であるため除外．
- data dump に関する投稿 25 件
  Exploit Kit における総括的な投稿であるため除外．
- ISC diary に関する投稿 9 件
  情報共有報告に関する投稿であるため除外．
- Android アプリケーションに関する投稿 1 件
  Android に関する投稿であるため除外．

上記投稿記事を除外した結果，調査対象の投稿は 229 – (43 + 25 + 9 + 1) = 151 件となった．これら 151 件について投稿記事情報を参考に，入り口サイトに関する明確な報告がある場合には攻撃に悪用されるタグを特定し，明確な報告がないものについては「その他」とした．また，本調査結果における iframe タグは，直接 Web サイトに仕込まれる iframe タグだけに限らず，スクリプトの実行により Web サイトに生成される iframe タグについても数に含まれている．この調査結果を表 1 に示す．

表 1 MTAnet の入り口サイトに関する調査結果

誘導手法	不可視化手法	件数	件数	内訳
iframeタグ	微小サイズによる埋め込み 閲覧領域外への配置 CSSの設定	30件 72件 0件	102件	67.5%
objectタグ あるいはembedタグ	CSSの設定 (opacityを0に設定)	18件	18件	11.9%
その他			31件	20.5%
投稿総件数			合計151件	

調査結果から，投稿記事の内，約 68%は iframe タグを悪用した攻撃であることがわかった．また，多くはないものの object タグ・embed タグを悪用した攻撃も存在することがわかった．「その他」に該当する攻撃については，公開 pcap ファイルをもとに独自に調査したところ，HTTP レスポンスヘッダの Location によるリダイレクトや入り口サイトにて脆弱性を悪用していると思われるコンテンツが見受けられた．

また，攻撃に悪用される 3 種類の HTML 要素はいずれも不可視化され，iframe タグについては微小サイズによる埋め込み，あるいは Web ブラウザの閲覧可能領域外への配置，object タグ・embed タグについては CSS の設定により透過処理されていた．これらの調査結果から，攻撃には特定の HTML 要素が使われ，かつ利用者にその存在を気づかれないよう 3 種類の不可視手法が用いられていることが明らかになった．

4. 攻撃 Web サイト検知システムの実装

前章の調査結果から，不可視 Web コンテンツに着目することにより DbD 攻撃サイトの検知が可能だと考える．本章では，このアイデアに基づき実装した攻撃 Web サイト検知システムについて説明する．図 2 は，攻撃 Web サイト検知システムの概要図である．攻撃 Web サイト検知システムは，情報収集処理と検知処理の 2 ステップで行われる．なおシステムは Firefox の拡張機能として実装している．

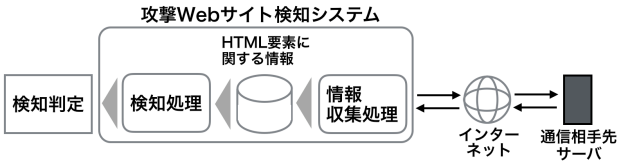

図 2 攻撃 Web サイト検知システムの概要図

4.1 情報収集処理

情報収集処理では，Web ブラウザ上において描画された Web コンテンツから検知に必要な情報を取得する．なお，前章の MTAnet の調査において，スクリプトが実行されることで動的に iframe タグが生成される事例を確認した．そこで本システムの情報処理では，単に静的な HTML ファイルを取得するのではなく，スクリプト実行後の状態の Web コンテンツを取得するようにした．

表 2 HTML 要素に関する情報

HTML 要素に関する情報	具体例
タグの種類	img
src，href，value	http://A.com
src，href，value のホスト部	A.com
src，href，value のホスト部の国情報	JP
X 座標	30
Y 座標	60
width	200
height	100
opacity	1
visibility	visible
display	inline

具体的な取得情報を表 2 に示す．収集対象の HTML 要素は (iframe，object，embed，a，img，html) の 6 種類に限定し

ている．iframe タグ・object タグ・embed タグは，MTAnet の調査から判明した入り口サイトで利用される傾向の高い HTML 要素である．a タグ・img タグは，検知に直接利用する情報ではない．しかし，a タグ・img タグの多くは良性コンテンツを指定すると考えられるため，これらのタグの指定先ホストの情報を活用することで，誤検知の抑制に応用できると考えた．html タグは描画領域のサイズ情報を利用することで，閲覧領域外に配置される不可視 Web コンテンツの検知に応用する．HTML 要素に関する情報を取得した後，各 HTML 要素の src・href・value 属性においてホスト名が取得できた場合には，さらに DNS と GeoIP[8] を利用してホストが存在する国情報を取得した．

4.2 検知処理

検知処理では，表 2 の情報を入力とし，攻撃 Web サイトの検知判定を行う．本研究では 2 つの検知ルールを策定した．1 つは「MTAnet 調査結果に基づく不可視」であり，もう 1 つは「包括的な不可視」である．どちらの検知ルールも，微小サイズによる埋め込み・閲覧領域外への配置・CSS の設定の 3 点に着目し判定を試みるが，判定のための閾値設定が異なっている．以降では，これらの検知ルールについて説明する．

4.2.1 微小サイズによる不可視

第 3 章で述べた調査により，width と height がともに 5px から 19px の微小サイズの iframe タグが確認できた．また object タグと embed タグのサイズは 50px 以下であった．これらの結果から「MTAnet 調査結果に基づく不可視」では，iframe タグと，object タグ・embed タグとで別々の閾値設定とした．一方「包括的な不可視」では言葉の通り，両者のルールを包括する閾値として，3 種の HTML タグに対して以下の閾値を設定した．

「MTAnet 調査に基づく不可視」ルール

`iframe` タグ:

$$5px <= width <= 20px \text{ and } 5px <= height <= 20px$$

`object, embed` タグ:

$$width <= 50px \text{ and } height <= 50px$$

「包括的な不可視」ルール

`iframe, object, embed` タグ:

$$width <= 50px \text{ or } height <= 50px$$

4.2.2 閲覧可能領域外への配置による不可視

MTAnet の調査結果より，iframe タグが閲覧可能領域の外に配置される場合，いずれも iframe タグの Y 座標に負の値が設定され，特に −500px と −1200px 付近の 2 つの値が確認された．したがって「MTAnet の調査に基づく不可視」の判定閾値は以下の設定とした．また，object タグ・embed タグではこの手法による事例が確認されなかったため，閾値は設定しないこととした．一方「包括的な不可視」では，この閾値を一般化し，不可視になりうる条件として X 座標あるいは Y 座標が負の値であることを判定閾値とした．また iframe タグに限らず object タグ・embed タグも判定対象とした．

「MTAnet 調査に基づく不可視」ルール

`iframe` タグ:

$$-1300px <= Y < -400px$$

`object,embed` タグ:

判定処理なし

「包括的な不可視」ルール

`iframe, object, embed` タグ:

$$X < 0px \text{ or } Y < 0px$$

4.2.3 CSS の設定による不可視

MTAnet の調査結果より，object タグおよび embed タグについて CSS の opacity 属性を 0.0 に設定する透過処理が明らかとなった．よって「MTAnet 調査結果に基づく不可視」では，この値を判定条件とした．一方「包括的な不可視」では，opacity 属性以外に知られている “visibility=hidden” や “display=none” による不可視化手法も判定ルールに含めるとともに，opacity の判定閾値も半透明に該当する 0.5 までゆるめた値とした．

「MTAnet 調査に基づく不可視」ルール

`iframe` タグ:

判定処理なし

`object,embed` タグ:

$$opacity = 0.0$$

「包括的な不可視」ルール

`iframe, object, embed` タグ:

$opacity < 0.5$ or “visibility=hidden” or “display=none”

5. 検知システムの有用性検証

5.1 MTAnet 投稿記事に基づく検知率の検証

MTAnet の 2016 年の下半期における調査対象の投稿記事 151 件中 120 件は不可視 Web コンテンツを悪用した攻撃であることが判明している．このことから，「包括的な不可視」,「MTAnet 調査結果に基づく不可視」に関して，攻撃の検知率を算出すると，120/151 = 0.795（約 79.5%）となる．

5.2 良性 Web サイトによる誤検知率の検証

誤検知率検証のために，DMOZ[9] にインデックスされた URL 群から無作為に 300 の URL を取得し，これらを良性 Web サイトの URL 群とした．DMOZ は，世界最大の Web ディレクトリである．前章にて説明した攻撃 Web サイト検知システムを Firefox にインストールし，Web ブラウザの操作自動化ツールである Selenium[12] を利用することで，Web サイトの巡回を行った．その後，リダイレクト

が発生した Web サイトおよび適切に情報取得ができなかった Web サイトを除外する目的で，以下の処理を行った．

- 同一ホストへリダイレクトした Web サイトについてはリダイレクト元 Web サイトを除外
- a タグ・img タグのどちらのタグも取得されなかった Web サイトについては除外

結果，284 の良性 Web サイトの HTML 要素に関する情報を取得した．「検知ルールなし」,「包括的な不可視」,「MTAnet 調査結果に基づく不可視」の 3 つの検出ルールに基づき，悪性 Web サイトとして判定された Web サイト数を表 3 に示す．

表 3 MTAnet の入り口サイトに関する調査結果

各定義	タグ	当該タグを含む Webサイト数	指定先が URL形式の Webサイト数	指定先が 外部ホストの Webサイト数
検知ルールなし	iframe	112	101	95
	object	13	3	3
	embed	8	8	4
包括的な不可視	iframe	88	74	74
	object	10	2	2
	embed	0	0	0
MTAnet調査結果に基づく不可視	iframe	8	5	5
	object	2	0	0
	embed	0	0	0

5.2.1 検知ルールなし

この「検知ルールなし」による調査を行った理由は，良性サイトにおいて，攻撃に悪用される HTML 要素 3 種がどの程度含まれているのかを検証するためである．結果，284 サイト中 112 サイトは iframe タグ，13 サイトは object タグ，8 サイトは embed タグを保持していた．またこれらの内，HTML 要素の src・value の URL 中ホスト部が外部ホストを参照しているものについて調査したところ，iframe タグ 95 サイト，object タグ 3 サイト，embed タグ 4 サイトであった．

5.2.2 包括的な不可視

包括的な不可視に該当した Web サイトは，iframe タグ 88 サイト，object タグ 10 サイト，embed タグは該当なしという結果になった．また，この定義において，外部の Web サイトに存在するコンテンツを描画する指定になっているものは，iframe タグ 74 サイト，object タグ 2 サイトであった．したがって，外部 Web サイトを参照する HTML 要素のみに限定し誤検知率を算出すると，76/284 = 0.268（26.8%）という結果になった．

5.2.3 MTAnet 調査結果に基づく不可視

MTAnet 調査結果に基づく不可視に該当した iframe タグは 8 サイト，object タグは 2 サイト，embed タグについては該当なしという結果となった．この定義において，外部の Web サイトに存在するコンテンツを描画する指定になっているものは，iframe タグ 5 サイト，object タグは該当なしという結果になった．したがって，外部 Web サイトを参照する HTML 要素のみに限定し誤検知率を算出すると，5/284 = 0.018（1.8%）という結果になった．

5.3 考察

包括的な不可視の適用により，不可視定義なしの Web サイト数と比較して，良性 Web サイト数を悪性 Web サイトとして誤判定する Web サイト数を 20 サイトほど減らすことができた．これは包括的な不可視が有効的に機能した結果であると肯定的な見方ができる一方，誤検知率が 26.8% と高い割合であるという否定的な見方もできる．そこで，誤検知に該当したタグについてさらに調査を行った．結果，包括的な不可視に該当した良性の Web コンテンツの多くは，タグ中の指定先が Google，Facebook，広告ネットワークなどに関連したホストであることが判明した．このことから，Web サイトの管理者が意図的に不可視 Web コンテンツを埋め込むのではなく，外部の Web サービスを経由して結果的に Web サイト上に iframe が生成されていると考えられる．よって，これらの外部サービス利用により生成される不可視 Web コンテンツについてはホワイトリストを設定することで軽減できる見込みがあると考える．

また，包括的な不可視と MTAnet 調査結果に基づく不可視とを比較すると，誤判定された Web サイト数が 80 サイトほど減少していることから，実データに基づき検知判定処理を実装することで，誤検知率を軽減できる見込みがあると言える．なお，誤検知率を軽減できた要因について調査したところ，良性の Web サイトでは CSS により display プロパティが none に設定された Web コンテンツが見受けられた．本研究では，HTML 要素情報の取得に getBoundingClientRect 関数を利用しているが，display が none に設定された Web コンテンツは，この関数によりサイズ・座標位置情報を取得した場合，width，height，X・Y 座標の値がいずれも 0 の値として取得される．width と height が共に 0 サイズの Web コンテンツは，MTAnet の調査結果に基づく不可視の検知ルールには該当しないため，このような結果になったと考える．

6. 可視化システム

前章では，不可視 Web コンテンツを検知するために，検知システムを実装し，その有用性について検証した．良性 Web サイトについて誤検知率の検証したところ，包括的な不可視では約 26.8%，MTAnet の調査結果に基づく不可視では約 1.8% という結果となり，誤検知の問題が残ることがわかった．こうした誤検知に該当する Web サイトについては，実データを検証し，src・href・value の指定先ホストや Web コンテンツの埋め込み状況などを調査することで正しく判定できる見込みがある．しかし，実データの解析は文字列ベースの調査となるため，手間・時間を要する懸念がある．加えて，一般ユーザが文字列ベースの調査を行うこ

とは困難であることが想定される．そこで本研究では，第4章4.1節で述べた検知システムの情報収集処理で取得したデータを可視化するシステムを提案することとした．

第3章におけるMTAnetの調査より，不可視Webコンテンツには微小サイズ・閲覧領域外・CSS設定の3手法があることが判明している．そのため，これら全ての情報をひと目で分かるよう可視化システムを設計する．加えて本可視化システム特徴として，src・href・valueの指定先ホストの国情報とその参照数に着目している．

図3は可視化システムの概要図である．衝立状のレイアウトとし，垂直面にはWebページのスケルトン図，水平面には地図を描画する．スケルトン図とは，Webサイトの描画状況について，各HTML要素の描画領域だけを線画として描いたものである．また両面の間の線は，Webサイト内のHTML要素の中でホスト情報を属性に持っているものに限り，そのホストが存在する国と各HTML要素との関係をリンク線として描画している．

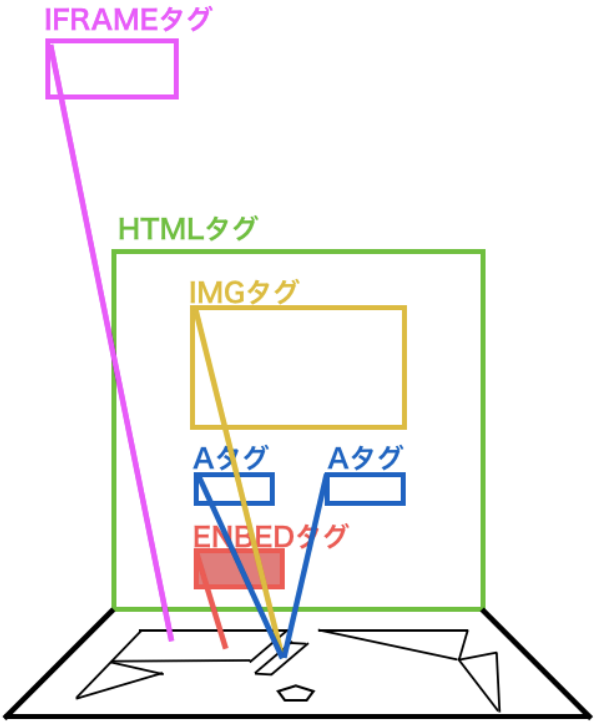

図3 可視化システムの概要図

7. 可視化システムの有用性検証

本章では，良性Webサイトと悪性Webサイトの可視化事例を提示することで，可視化システムの有用性について検証する．

7.1 良性Webサイトの可視化事例

図4は，良性Webサイトについて，システムで可視化した表示例である．

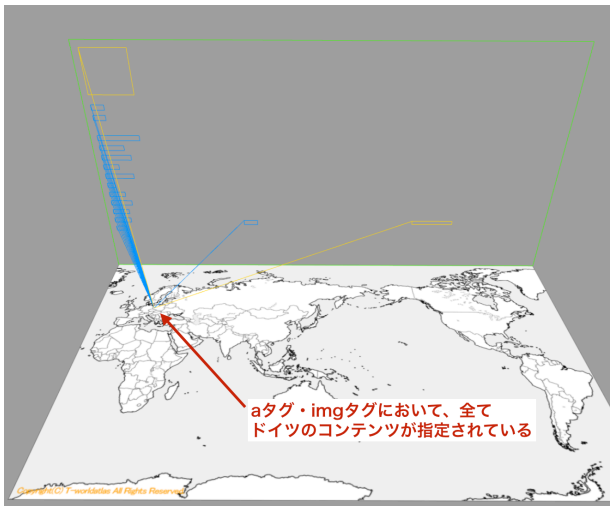

図4 DMOZ[9]より入手したWebサイト「http://irtel.uni-mannheim.de/pxlab/」における可視化

垂直面には，緑色の線で描かれた大きな四角形の中に青色と黄色による四角形が描かれている．大きな緑色の四角形はhtmlタグの描画領域を表している．つまりWebブラウザでユーザが見ることのできる描画領域である．青色・黄色の四角形は，aタグ・imgタグの描画領域であり，加えて攻撃Webコンテンツとは関係ないタグである可能性が高いものである．これらのことから，このWebページにはiframe，object，embedの3種のタグは存在しないことがわかる．

また，垂直面と水平面の間の線群に注目する．線群は，垂直面のaタグ・imgタグを表す四角形と，水平面にある地図で欧州域内の1点を結んでいる．閲覧中Webサイトのトップレベルドメインが「de」であり，ドイツを指し示していると推測される．またそれ以外の国へのリンクは見当たらない．このことからも疑わしい点はないことが見てわかる．

7.2 DbD攻撃の可視化事例

MTAnetからpcapファイルを入手し，HTML・JavaScriptファイルを抽出し，DbD攻撃において入り口サイトに埋め込まれるiframeタグ・objectタグ・embedタグの情報を取得した．本節では，第3章にて明らかになったDbD攻撃に悪用される3つの不可視Webコンテンツについて，システムによる可視化事例を提示する．

7.2.1 微小サイズの不可視Webコンテンツの可視化

図5は，微小サイズのiframeタグについて可視化したものである．緑色・黄色・青色・桃色の線は，それぞれhtmlタグ・imgタグ・aタグ・iframeタグを表している．

図中の線群は大きく2つに分類でき，日本から伸びる線と欧州域内から伸びる線である．日本から伸びる線はaタグ・imgタグなど複数確認できる一方で，オランダから伸びる線は1本しか確認できない．この1本だけ伸びる線

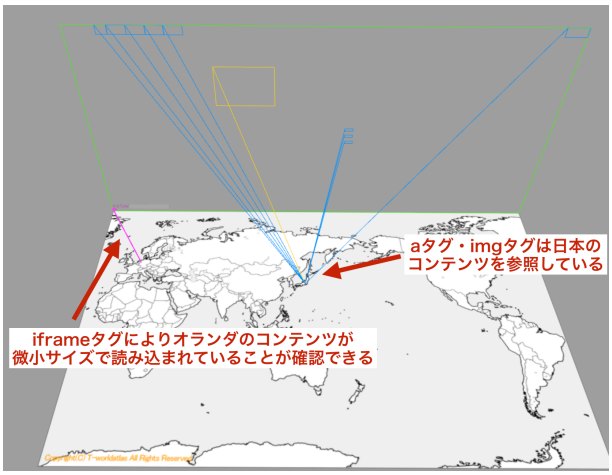

図 5 微小サイズの iframe タグの可視化

は，a タグや img タグが参照するホストの国情報とは異なることが見てわかる．また，可視化において描画される桃色の枠に着目すると，描画される四角形のサイズが小さいことがわかる．これらのことから，微小サイズの iframe タグを用いた入り口サイトの可能性が高いと判断することができる．

7.2.2 閲覧領域外上方への配置による不可視

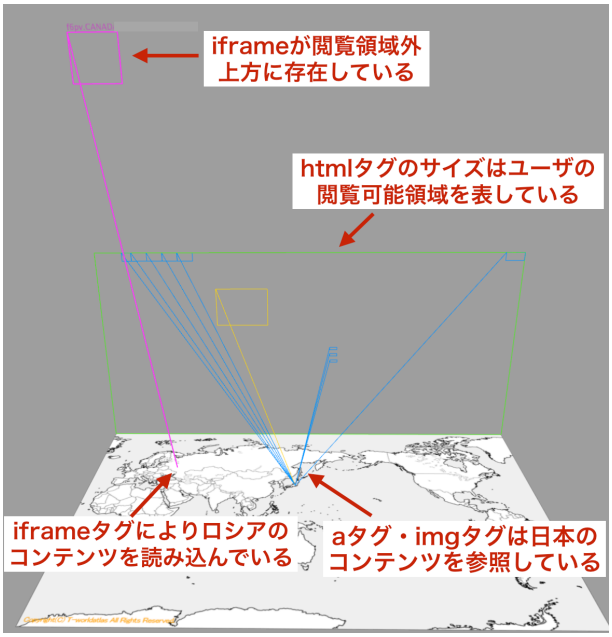

図 6 閲覧領域外上方へ配置された iframe タグの可視化

図 6 中左上に，桃色の枠線による四角形が見てとれる．これは線の色から iframe によるコンテンツであることがわかり，垂直面と水平面の間のリンク線から，ロシアのホストにおけるコンテンツを描画していると推測される．また，この桃色の四角形は，HTML の描画領域を示す緑色の四角形の外側に描画されている．このことから閲覧可能領域の外に描画がされていることも一目で理解できる．これらの情報から，iframe タグを利用し，閲覧可能領域の外側に描画させる攻撃コードが入り口サイトに挿入されていると推測できる．

7.2.3 CSS の透過処理による不可視

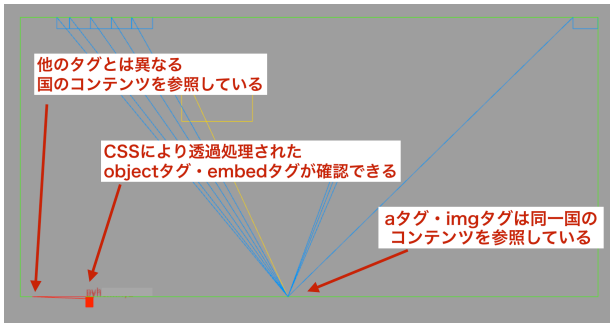

図 7 透過処理された object タグ・embed タグの可視化

図 7 は，CSS の透過処理による不可視について可視化したものである．html の閲覧領域を示す緑色の四角形の左下部に微小サイズの赤色の四角形が確認できる．赤色コンテンツは，object タグあるいは embed タグを表している．また，四角形の内側が塗りつぶされていることが確認できる．本可視化システムでは CSS 設定による不可視 Web コンテンツを明らかにする目的で，これらの Web コンテンツの場合には描画される四角形の枠内を塗りつぶすようにしている．つまり，この object タグ・embed タグは CSS 設定により不可視化されていると判断できる．また，図の例は可視化システムの正面からの図であるため詳細な国情報は把握できないが，青色・黄色線群と赤色の線とは異なる国のコンテンツを取得していると考えることができる．可視化システムにより得られた情報を整理すると，透過処理された object タグあるいは embed タグが含まれた Web サイトであり，入り口サイトである可能性が高いと判断することができる．

8. 考察

本章では，提案システムの考察について述べる．

8.1 本研究における制限

本研究では不可視 Web コンテンツに着目しているため，不可視 Web コンテンツを利用せずに行われる DbD 攻撃については検知することができない．具体的には，リダイレクトによる攻撃者 Web サイトへの誘導や，不可視ではなく可視コンテンツを利用した攻撃などが考えられる．これらの攻撃を検知するためには，不可視 Web コンテンツへの着目だけでは検知が困難であると考える．そのため，リダレクトによる誘導の場合には，HTTP レスポンスヘッダの Location 情報を活用し，検知に応用することを検討している．

8.2 今後の課題

今後の課題として，3 点ほど議論する．

1 つ目は，情報取得システムの改良である．現在の情報取得は Web サイト読み込み完了時点での HTML 要素の情報を取得している．そのため，非同期により生成される Web コンテンツの情報については取得できていない．一定時間ごとに HTML 要素の変化を監視するなどして，これらの Web コンテンツについても情報が取得できるよう改善を試みる予定である．また，親要素の CSS により透過処理される Web コンテンツの情報を取得するために，offsetParent 関数を利用している．offsetParent 関数は，absolute 属性が設定されている親要素を取得する関数である．しかしながら，absolute 属性が設定されていない親要素により透過処理が行われる場合には，現時点での実装では検知できない．これについては，親要素を取得する別の関数を利用することで改善できる見込みがある．

2 つ目は，検知システムの改良である．MTAnet 調査結果に基づく不可視を定義することで誤検知率を軽減できることが判明した．よって，検知システムの敷居値を改善することで，さらに誤検知を減らすことができると考える．

3 つ目は，他の Web ブラウザへの実装である．今回，Firefox の拡張機能により HTML 要素情報取得システムを実装したが，HTML 要素情報の取得については Firefox 固有の関数ではなく，全て JavaScript の関数を利用することで取得している．そのため，他の Web ブラウザにおいても同様の情報取得システムの実装が可能だと考える．

9. おわりに

本研究では，微小サイズによる埋め込み・閲覧領域外への配置・CSS 設定の 3 手法における不可視 Web コンテンツに着目することで，DbD 攻撃の検知を目指した．攻撃 Web サイト検知システムでは，MTAnet の投稿記事情報をもとにした検知率が 79.5%となった．また，DMOZ から無作為に URL 取得することで誤検知率についても検証し，包括的な不可視では 26.8%，MTAnet 調査調査に基づく不可視では 1.8%という結果となった．これらの結果から，不可視 Web コンテンツに着目することにより，攻撃 Web サイトと良性 Web サイトとの判別可能性を評価を通じて明らかにした．

提案システムの使用用途は大きく 3 つの利用方法が考えられる．1 つ目は一般ユーザによる利用である．提案システムによる検知システムと可視化システムを利用することで，DbD 攻撃が発生する Web サイトを特定できる見込みがある．2 つ目は Web サイト管理者による利用である．Web サイト管理者は，定期的に自身の Web サイトが改ざんされたかどうか調べることは容易でない．提案システムを導入することで，攻撃者により改ざんされていないかどうか調べることが可能となる．3 つ目はクローラによる利用である．提案システムは，入り口サイトの Web コンテンツに着目した検知手法であるため，クローキングなどの攻撃者による検知回避技術の影響を受けにくいと考える．また，マルウェア配布サイトが一時的に停止している等の場合にも，本手法であれば検知可能であると考える．

今後は，第 8 章にて議論した課題解決に向けて研究を行う．

参考文献

[1] IBM Security Services：2016 年上半期 Tokyo SOC 情報分析レポート，IBM Security Services（オンライン），入手先〈https://www-935.ibm.com/services/jp/ja/it-services/soc-report/〉(参照 2017-01-30).
[2] Malware-Traffic-Analysis.net: Malware-Traffic-Analysis.net homepage, Malware-Traffic-Analysis.net（オンライン），入手先〈http://www.malware-traffic-analysis.net〉(参照 2017-01-21).
[3] 笠間貴弘，神薗雅紀，井上大介：Exploit kit の特徴を用いた悪性 Web サイト検知手法の提案，コンピュータセキュリティシンポジウム 2013 論文集，Vol.2013，No.4，pp.603-610（2013）.
[4] 酒井裕亮，佐々木良一：Drive By Download 攻撃に対する HTTP ヘッダ情報に基づく検知手法の提案，研究報告マルチメディア通信と分散処理（DPS），Vol.2013-DPS-154，No.29，pp.1-6（2013）.
[5] 望月翔太，高田哲司：Web ページ内リンク情報の変化に基づく Web 改ざん検知の有効性検証，コンピュータセキュリティシンポジウム 2015 論文集，Vol.2015，No.3，pp.504-511（2015）.
[6] 西田雅太，星澤裕二，笠間貴弘，衛藤将史，井上大介，中尾康二：文字出現頻度をパラメータとした機械学習による悪質な難読化 JavaScript の検出，研究報告コンピュータセキュリティ（CSEC），Vol.2014-CSEC-64，No.21，pp.1-7（2014-02-27）.
[7] 田村佑輔，甲斐俊文，佐々木良一：ユーザ標的型 Web サイト改ざんに対する検索エンジンを用いた検知手法の提案，情報処理学会論文誌，Vol.51，No.1，pp.191-198（2010）.
[8] MAXMIND: MAXMIND homepage，MAXMIND（オンライン），入手先〈https://www.maxmind.com/en/home〉(参照 2017-01-22).
[9] dmoz: Welcome to DMOZ! It's the Web, Organized.，dmoz（オンライン），入手先〈https://www.dmoz.org〉(参照 2017-01-23).
[10] 小林峻，寺田成吾，瀬戸口武研，道根慶治，山下康一：Drive-by Download 攻撃検知手法の継続的評価と Exploit Kit に対する考察，コンピュータセキュリティシンポジウム 2016 論文集，pp.964–970（2016）.
[11] 佐藤祐磨，中村嘉隆，高橋修：エクスプロイトキットで利用される文字列特徴を用いた悪性 URL 検出手法の提案，研究報告コンピュータセキュリティ（CSEC），Vol.2016-CSEC-72，No.25，pp.1-6（2016）.
[12] Selenium: Selenium homepage，Selenium（オンライン），入手先〈http://www.seleniumhq.org〉(参照 2017-01-25).