RIG Exploit Kitにおける攻撃傾向の調査

山田 道洋[1]　小池 倫太郎[2]　菊池 浩明[2]　黄 緒平[3]

概要：Webサイトを閲覧したユーザに対してWebブラウザ等の脆弱性を突く攻撃コードを送り，マルウェアをダウンロード・実行させるDrive-by Download攻撃が蔓延している．その背景には攻撃のための専用ツールとしてExploit Kitが存在し，攻撃者は特別な技術や知識がなくとも容易にDrive-by Download攻撃を仕掛けることが出来るようになっている．本稿では2017年2月から2017年8月にかけて観測したRIG Exploit Kitを用いたDrive-by Download攻撃のデータを元に，攻撃に使用されたサイトやExploit Kitの攻撃傾向の調査結果を報告する．

キーワード：MWS，Drive-by Download攻撃，Exploit Kit，不正サイト解析

Michihiro Yamada[1]　Rintaro Koike[2]　Hiroaki Kikuchi[2]　Xuping Huang[3]

1. はじめに

近年，公開サーバへのサイバー攻撃は増加の一途を辿り，深刻な被害を出している．例えば2017年3月頃に発生したApache Struts 2の脆弱性CVE-2017-5638では外部から任意のコードが実行可能となり，多くのサイトで情報漏えいが発生した．このような公開サーバへの攻撃は日々高度化しており[1]，脅威への対応が求められている．高度化している攻撃の1つとして，Drive-by Download攻撃が挙げられる．Drive-by Download攻撃は改ざんされた一般のWebサイトや不正なWeb広告を閲覧したユーザに対して攻撃者が用意した攻撃サーバへ誘導し，利用者のWebブラウザ等の脆弱性を突くことでマルウェアに感染させる．Drive-by Download攻撃では多くの場合，専用ツールExploit KitによってWebブラウザの脆弱性を突きマルウェアに感染させる．攻撃者はExploit Kitにユーザを誘導するだけでDrive-by Download攻撃を仕掛けることが可能であり，攻撃の難易度は低くなっている．中でも，RIG Exploitは複数の機関から注意喚起が行われており，その被害が深刻である[2][3]．

RIG Exploit Kitで用いられるドメインやIPアドレスは数時間で変更され[4]，IPアドレス等を用いた単純なブラックリストではRIG Exploit Kitとの通信を遮断することは困難である．また，利用されるURLの特徴も頻繁に変化し[5]，URLから検知用のシグネチャを作成することも容易ではない．加えて，解析や追跡を妨害するために，攻撃に用いられるコードが多重に難読化されていたり，アクセス制御が行われているため，RIG Exploit Kitを解析することは困難である．

そこで我々はRIG Exploit Kitを利用する複数の攻撃キャンペーンを探索するプログラムを実装し，Rig Exploit Kitの特徴を調査した．この観測手法や難読化の詳細については，我々の研究グループによる報告[10]に示す．

本稿では，2017年2月～8月に観測されたDrive-by Download攻撃のデータから決定木の作成を行い，キャンペーン毎の特徴の分析を行う．

2. 背景

本章では，Drive-by Download攻撃の攻撃フローとExploit Kitの概要について述べた後，関連する研究について述べる．

2.1 Rig Exploit Kitを用いたDrive-by Download攻撃

RIG Exploit KitによるDrive-by Download攻撃の流れ

1 明治大学大学院先端数理科学研究科
Graduate School of Advanced Mathematical Sciences，Meiji University
2 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences，Meiji University
3 明治大学　研究・知財先着機構
Strategic Coordination of Research and Intellectual Properties, Meiji University

は図 1 の様に，5 つの段階に分けることができる．

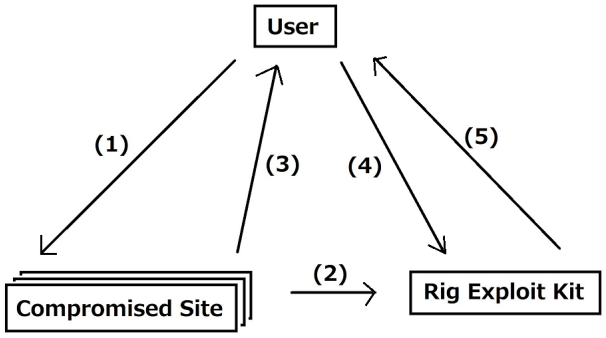

図 1 Drive-by Download 攻撃の流れ

(1) 脆弱性などを利用し攻撃者によって改ざんされた一般の Web サイト (以下，Compromised サイト) や不正な Web 広告 (Malvartising) をユーザが閲覧する．(2) 改ざん時に挿入されたコードなどによって RIG Exploit Kit へ繋がる URL が生成される．(3) その URL へ誘導するようなコードをユーザに読み込ませることで，ユーザを RIG Exploit Kit へ誘導する．(4) そうして誘導されると，RIG Exploit Kit はユーザの使用している Web ブラウザ等の脆弱性を突くようなコードを含む難読化された JavaScript コードを生成し，ユーザへ送り込む．(5) ユーザの Web ブラウザはそれらのコードを実行し，マルウェアをダウンロード・実行する．

RIG Exploit Kit の特徴は (3) の，攻撃サイトと (4) のマルウェア配布サイトをまとめて RIG Exploit Kit サイトと呼ぶ．このような仕組みを Exploit Kit as a Service と呼ぶ [7].

Exploit Kit as a Service の性質上，オンプレミスで Exploit Kit を設置している場合よりも更新が容易で，高頻度で Exploit Kit の サーバや攻撃コードが更新され，解析や防衛を困難にしている．

2.2 攻撃キャンペーン

Exploit Kit への誘導方法には用いる脆弱性の種類や対象とする CMS のバージョンなどに関するいくつかのキャンペーンがある．例えば，pseudo-Darkleech では，iframe タグを用いて Web のドライブをするのに対し，EITest では javascript で動的にコードを生成する．本研究で観測したキャンペーンの入口サイトは大きく分けて Compromised サイト，Malvertising の 2 種類であった．誘導方法と観測したキャンペーンの関係を表 1 に示す．

2.3 関連研究

笠間らは，Exploit Kit によって構築された悪性 Web サイトの特徴を用いて Drive-by Download 攻撃を検知する手法を提案している [8]．RIG Exploit Kit についてユーザ環境から調査した文献として [4][9] がある．嶌田らは，セキュリティベンダが保有する Web アクセスログからユーザ環境における RIG Exploit Kit のログ分析を行い，RIG Exploit Kit で用いられているドメインの活動期間が数時間であると報告している [4]．NTT セキュリティのグループでは，RIG Exploit Kit で用いられている攻撃手法について詳細に述べつつ，URL やドメインや IP アドレスについて分析を行い，RIG Exploit Kit の特徴について報告している [9].

表 1 Exploit Kit への誘導方法とキャンペーンの関係

入口サイト	キャンペーン
Compromised サイト	pseudo-Darkleech
	EITest
	Good Man
	Fake Chrome Popup
	DecimalIP
Malvertising	Seamless
	HooAds
	Despicale

我々の研究グループでは，RIG Exploit Kit を利用する複数の攻撃キャンペーンを探索するプログラムを実装し，それらの特徴を調査した [10]．その結果得られた中継サイトを継続的に観測し，RIG Exploit Kit の変化を明らかにした．

3. 攻撃の観測 [10]

[10] では，Drive-by Download 攻撃のデータ収集のために次の 2 種類の調査を行った．Compromised サイト，Exploit Kit の収集期間とデータ件数を表 2 に示す．これらのデータを元に特徴などの分析を行う．(1) クローラによる改ざんされた Web サイトの収集．(2) 中継サイトのポーリングによる RIG Exploit Kit の追跡こうして収集した

3.1 Compromised サイトの収集

2017 年 2 月 24 日〜4 月 10 日の間，Alexa Top 1 Million に対して HTTP で接続し，Web サイトのソースコードをダウンロードする．攻撃キャンペーンと Exploit Kit のシグネチャを用いてパターンマッチングを行い，マッチした Web サイトから，攻撃者によって挿入されたであろうコードと Exploit Kit に関する情報を収集した．

3.2 RIG Exploit Kit の追跡

2017 年 5 月 4 日〜8 月 15 日の間，RIG Exploit Kit に対して 1 時間間隔でアクセスを行い，マルウェアのダウンロードが行われた時刻と IP アドレスを記録した．RIG Exploit Kit の URL は Seamless と呼ばれるキャンペーンの中継サイトから取得した．

表 2　各データセットの収集期間と件数

	Compromised サイト (ユニーク)	Exploit Kit
期間	2017/2/24〜4/10	2017/5/4〜8/15
N	252	2182

4. データの分析

4.1 実験目的

本実験の目的は次の 2 つである．

(1) キャンペーンの観測時期や RIG Exploit Kit の中継サイトの変更時間などから Drive-by Download 攻撃の特徴などを調査する．

(2) Compromised サイトのデータを元に各キャンペーンの決定木を作成することで，キャンペーン毎の Compromised サイトの傾向を明らかにする．

4.2 実験方法

4.2.1 Contents Management System

収集したキャンペーンが対象とする固有の Contents Management System(以下，CMS) がある．そこで，Compromised サイトの収集を行う際に Web ページを作成した際に使用された CMS の情報も調査した．これにより，より広く脆弱性を利用されている CMS を明らかにする．

4.2.2 アクセスランキング

Compromised サイトへのアクセスが多ければ多いほど，Drive-by Download 攻撃の成功数は増える．そのため，アクセスランキング上位のサイトに Compromised サイトが多く存在するのではないかと考えた．そこで，収集した 252 の Compromised サイトについて Alexa のアクセスランキングの調査を行った．

4.2.3 RIG Exploit Kit の変更間隔

収集した RIG Exploit Kit サイトの IP アドレスと時刻から RIG Exploit Kit の IP アドレスの平均変更間隔時間やその分散を明らかにする

4.2.4 決定木

決定木は，ターゲットである属性を決定する論理条件を明らかにする機械学習アルゴリズムであり，根に近い属性が最も識別に重要な条件となる属性である．キャンペーン毎の Compromised サイトの傾向を明らかにするために各キャンペーンについて，決定木解析を試みる．

4.3 実験結果

4.3.1 キャンペーンと CMS

入口サイトの種類でキャンペーンを分類した URL の分布を図 2 に示す．各キャンペーンの月ごとの観測数を図 2 に示す．5 月までは pseudo-Darkleech，EITest といったキャンペーンが観測されていたが 5 月以降は DecimalIP や Seamless といったキャンペーンが観測され，Compromised サイトを利用していたキャンペーンが 5 月 6 月頃から Malvertising のキャンペーンに変化しているのがわかる．また，キャンペーンが実施される期間には重なりがあり，複数のキャンペーンが並列に展開されていることがわかる．

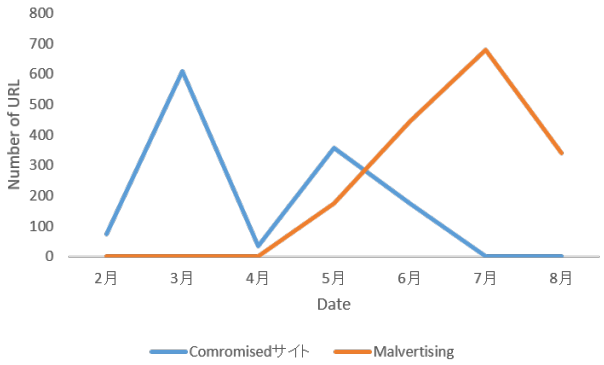

図 2　入口ページ毎の Campaign 観測数

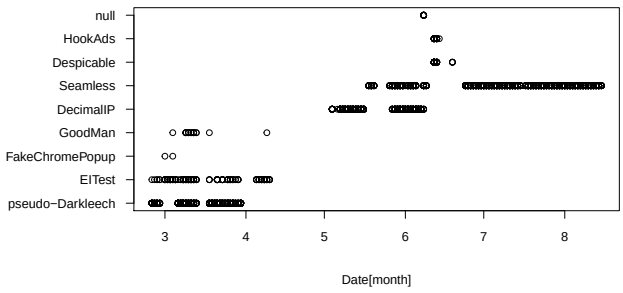

図 3　Campaign 毎の観測数

表 3　Camapaign 毎の Compromised サイトの使用 CMS の総計

	pseudo-Darkleech	EITest	FakeChromePopup	GoodMan	総計
WordPress1	23	17	1	0	41
WordPress2	29	5	0	0	34
WordPress3	8	5	0	0	13
WordPress4	1	21	1	0	23
Diviv. 3. 0. 10	1	0	0	0	1
Drupal7	23	1	0	0	24
Joomla!	61	7	0	5	73
MicrosoftFrontPage5. 0	0	0	0	1	1
MODx	1	0	0	0	1
PrestaShop	0	1	0	0	1
www. site5. com	0	1	0	0	1
N/A	24	8	0	7	39
総計	171	66	2	13	252

また，キャンペーン毎の Compromised サイトの使用 CMS の総計を表 3 に示す．WordPress1〜4 のバージョン情報を表 4 に従って一般化して集計している．最も多く観測された Campaign は psuedo-Darkleech の 171 件であった．また，WordPress のバージョン毎に観測件数を見ると，WordPress1〜3 では psuedo-Darkleech がで最も多く利用されていることが観測されていたが，WordPress4 では EITest が最も多い．最も多くのキャンペーンで対象にされていたのは，Joomla!の 73 件である．

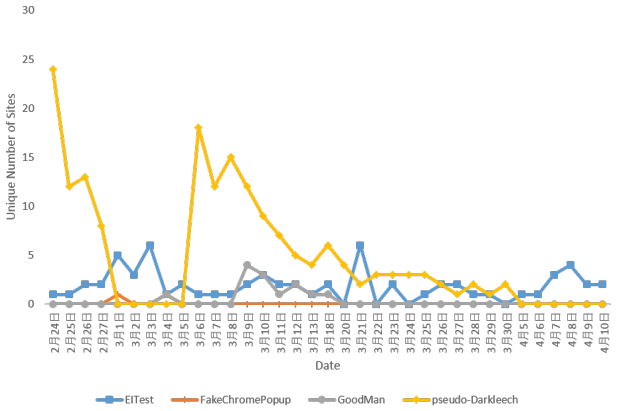

図 4　Compromised サイトを利用した攻撃キャンペーンの日毎の観測サイト数

表 4　WordPress の Ver 内訳

WordPress1	Ver 情報不明
WordPress2	3.5.1，3.6.1，3.8.16 4.0.1，4.0.15，4.0.16 4.1，4.1.1，4.1.15，4.1.16 4.2.12，4.2.13，4.2.3 4.3.5，4.3.8，4.3.9 4.4.2，4.4.3，4.4.7，4.4.8 4.5.3，4.5.6，4.5.7
WordPress3	4.6.1，4.6.3，4.6.4
WordPress4	4.7.1，4.7.2，4.7.3

表 5　Compromised サイトの Alexa アクセスランキングの統計量

最高順位	13,495
Max	19,837,000
計測不能	9 件
平均順位	846,828
中央値	1,749,177

4.3.2　アクセスランキング

252 の Compromised サイトの Alexa でのアクセスランキングの統計量と分布を表 5，表 6 にそれぞれ示す．50 万位以内のサイトは 54 サイトで全体の約 21.4% と上位のサイトが特に多いわけではなかった．

4.3.3　RIG Exploit Kit サイトの変更間隔

Exploit Kit は一定期間で IP アドレスを変化させながら活動を行っている．表 7 に RIG Exploit Kit サイトの変更間隔の統計量を示す．図 5 に RIG Exploit Kit の変更間隔とその割合を示す．平均で約 53 分，3 時間以内に 9 割，1 時間のうちに 5 割以上の RIG Exploit Kit が IP アドレスを変更されていた．

4.3.4　決定木

特定の Campaign か否かをターゲット属性として，R のパッケージ “rpart” により学習した決定木を図 6 図 7 図 8 にそれぞれ示す．入力した説明変数はサイトのドメイン，CMS，Alexa のアクセスランキングである．

ここで，「CMS=WordPress」等の分岐の条件を各節点の上に示し，左側の枝が条件にあてはまる．「特定の Campaign 以外のサイト数/特定の Campaign の数」を各節点の下に示す．例えば，図 6 の木では表示されている CMS かどうかが最も大きな条件であり，その CMS を利用しているサイトのうち Campaign が pseudo-Darkleech であったサイトは 1 つだけであった．また，図 7 からも WordPress4 を利用しているかどうかが，EITest の大きな条件であることがわかる．

表 6　2017/8/18 の Alexa アクセスランキング

順位	件数
1-100000	4
100001-200000	10
200001-300000	10
300001-400000	15
400001-500000	15
500001-600000	15
600001-700000	20
700001-800000	18
800001-900000	15
900001-1000000	19
1000001-1100000	11
1100001-1200000	11
1200001-1300000	9
1300001-	80

表 7　Exploit Kit の変更時間の統計量

平均時間	0.888 時間
最大	20 時間
標準偏差	1.582
中央値	0 時間

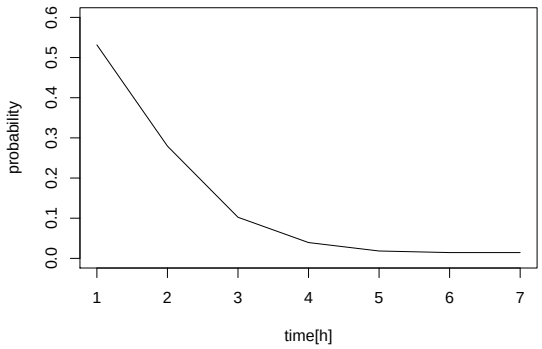

図 5　Exploit Kit サイトの更新間隔の分布

5.　考察

Compromised サイトの CMS の分布，決定木から psuedo-Darkleech は Joomla!や WordPress2，Drupal7 が，EITest は WordPress4 を Compromised サイトに利用していた点が他のキャンペーンと比較した時の大きな特徴であり，こ

れらの CMS の脆弱性を主にターゲットとしたキャンペーンであることが予想される．しかし，WordPress4 は最新のバージョンであり，脆弱性は少ないと考えられる．そのため，EITest は CMS の脆弱性ではなく他の脆弱性を狙ったキャンペーンであると考えられる．

Compromised サイトのアクセスランキングの分布は上位のサイトのほうが多いという予想であったが，実際は上位のサイトが特に多いわけではなかった．これは，より上位のサイトになるほど運営元が大きく脆弱性の対策が迅速に行われるためであると考えられる．

[9] によると，RIG Exploit Kit でドメインが利用されていた時期には同一ドメインを約 50 分，IP アドレスは約 550 分継続して利用されていた．しかし，現在では IP アドレスだけが利用されており平均約 53 分，5 割の RIG Exploit Kit が 1 時間以内に IP アドレスを変更するように変化している．これは IP アドレスを高頻度で変化させることで，解析やフィルタリングを困難にしているのではないかと考えられる．

6. おわりに

本稿では，RIG Exploit Kit による Drive-by Download 攻撃の傾向を調査した．また，決定木の作成により攻撃 Campaign は主に Web サイトの作成に使用された CMS の脆弱性を利用し Web サイトを改ざんし行われており，キャンペーン毎にターゲットとなる CMS が異なることを明らかにした．

今後の目標は，Exploit Kit の中継サイトなどの IP アドレスや URL の変化の特徴を明らかにすることである．

参考文献

[1] 株式会社 LAC：Apache Struts 2 における脆弱性 (S2-045、CVE-2017-5638) の被害拡大について，入手先 ⟨https://www.lac.co.jp/lacwatch/alert/20170310_001246.html⟩ (2017．08．23)．

[2] 日本サイバー犯罪対策センター：RIG-EK 改ざんサイト無害化の取組，入手先 ⟨https://www.jc3.or.jp/topics/op_rigek.html⟩ (2017．08．23)．

[3] 警察庁：ウイルス感染を目的としたウェブサイト改ざんの対策について，入手先 ⟨https://www.npa.go.jp/cyber/policy/pdf/rig.pdf⟩ (2017．08．23)．

[4] 嶌田一郎，太田敏史，岡田晃市郎，山田明，“ユーザ環境における RIG Exploit Kit の実態調査方法の提案”，第 78 回コンピュータセキュリティ・第 24 回セキュリティ心理学とトラスト合同研究発表会，July 2017．

[5] 株式会社 LAC：CYBER GRID VIEW Vol．3 猛威を振るう RIG Exploit Kit の全貌と対策，入手先 ⟨https://www.lac.co.jp/lacwatch/pdf/20170202_cgview_vol3_f001t.pdf⟩ (2017．08．23)．

[6] RSA：SHADOWFALL，入手先 ⟨https://blogs.rsa.com/shadowfall/⟩ (2017．08．23)．

[7] トレンドマイクロ株式会社：サービスとしてのエクスプロイトキット，入手先 ⟨http://about-threats.trendmicro.com/RelatedThreats.aspx?language=jp&name=Exploits+as+a+Service⟩ (2017．08．23)．

[8] 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”，マルウェア対策研究人材育成ワークショップ 2013 (MWS2013)，2013．

[9] NTT セキュリティ・ジャパン株式会社：RIG エクスプロイトキット解析レポート，入手先 ⟨https://www.nttsecurity.com/-/media/nttsecurity/files/resource-center/what-we-think/rigek-analysis-report.pdf⟩ (2017．08．23)．

[10] 小池 倫太郎，菊池浩明，“Drive- by Download 攻撃における RIG Exploit Kit の解析回避手法の調査”，コンピュータセキュリティシンポジウム 2017 (CSS2017)，2017

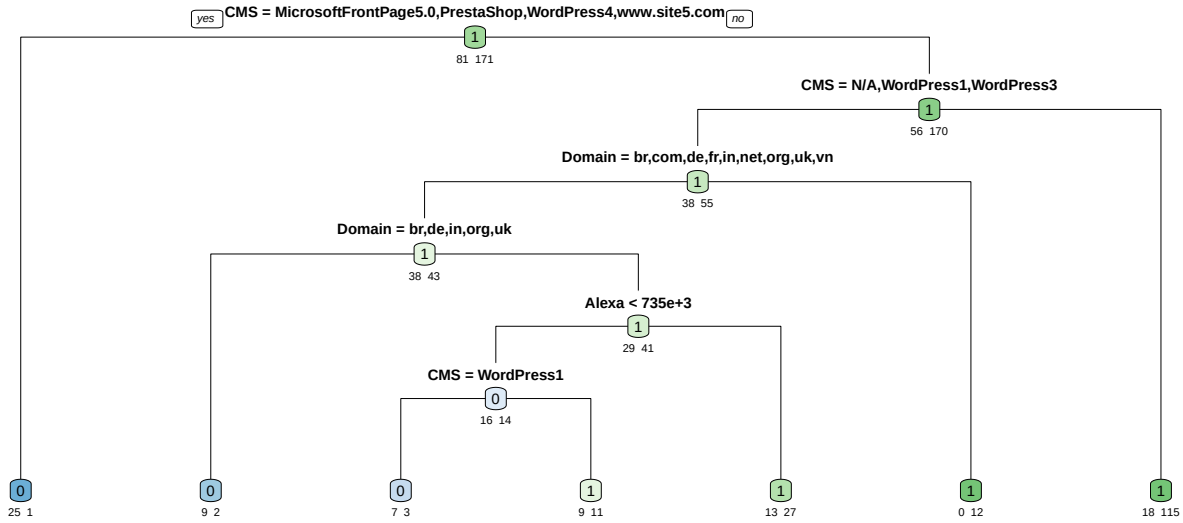

図 **6**　pseudo-Darkleech の決定木

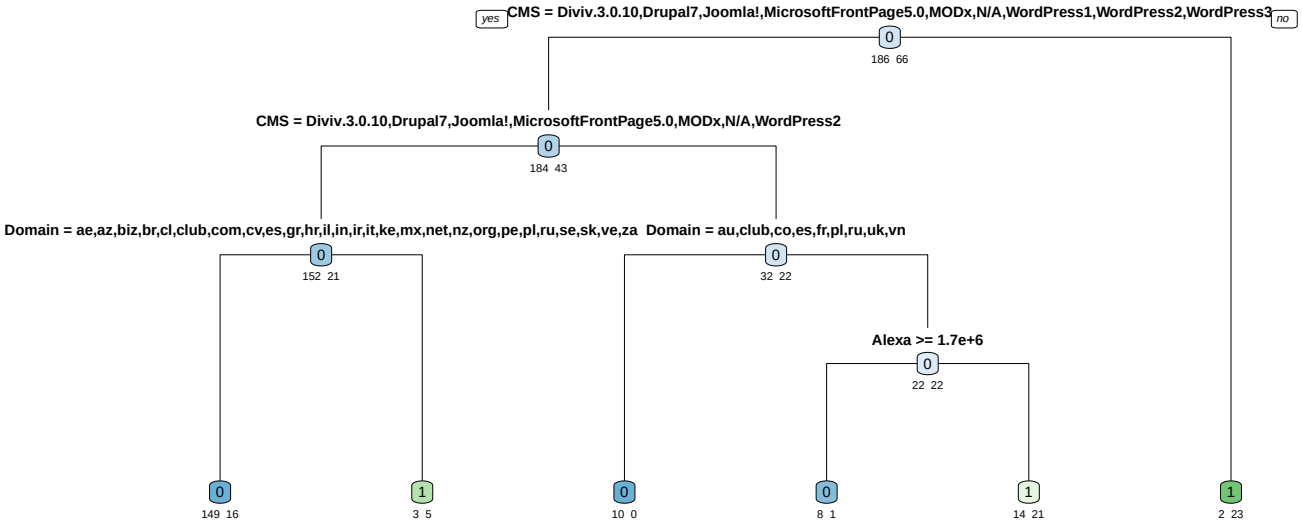

図 **7**　EITest の決定木

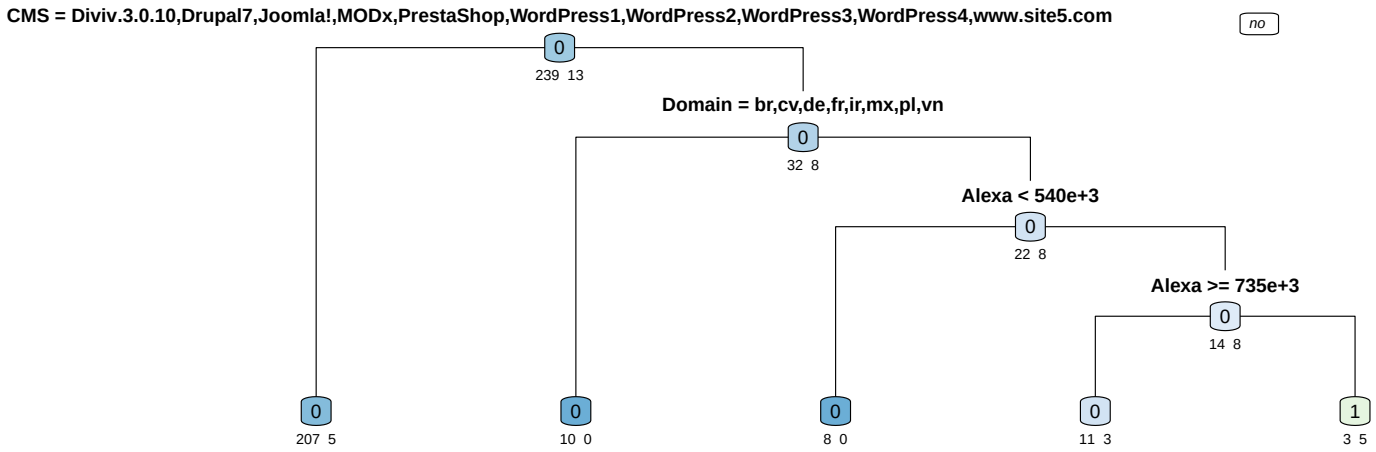

図 8　GoodMan の決定木