Vol.2013-IOT-20 No.17
2013/3/14

単純なハニーポットによるウエブアクセス動向調査

長谷川明生[†1]

IOT-16,No.51 ウエブ・サーバに対する不正と思われるアクセスに関して予備的調査を行い報告した．今回は、調査用プログラムに改善を加え、2012 年 2 月より 2012 年 12 月の期間に収集した 1500 件を超えるアクセス・データについて検討した結果を報告する．

Analyses of the Monitoring data Captured by a Simple Web Honeypot

AKIUMI HASEGAWA[†1]

A report on the preliminary analysis of the monitoring data captured by a simple dummy Web server was presented in IOT-16, No.51. Since then, the program was rewritten, and is running for over ten month to capture Web requests. Over 1500 accesses were observed till the end of the year 2012. In the report, analyses of the monitoring data captured by the revised program were made.

1. はじめに

2012 年 3 月の IOT16 No.51 において，Perl を用いて実装した簡単な Web ハニーポットで予備的に収集したデータについて分析を行って報告した．その後，当該プログラムについて改善をおこない，2012 年 2 月 23 日より 12 月 26 日の間の約 10 か月にわたりデータ収集をおこない 1559 件のアクセスを記録した．

これらのデータについてアクセス元の IP アドレスやそのドメインの調査，アクセスの規則性，アクセス対象や複数アクセスのあるアドレスについてのアクセスの規則性等について解析した．

2. データ収集と解析手法

データ収集の詳細については IOT16, No.51 に記したが，Perl で記述した簡単な Web サーバプログラムで 80 番ポートに接続に来た Web リクエストをすべて記録している．予備的調査から，要求される URL によって応答コードを変えるようにプログラムを変更している．具体的には，ドキュメント・ルートおよび index.html については 200 とともにプレーン・テキストを，php スクリプトへのアクセス には internal server error を，それ以外の要求には 404 を応答するようにした．このプログラムを走らせているサーバの IP アドレスに対して一切外部からリンクを張っていない．また，データ収集プログラムは IPv4 にのみ対応している．

(1) データの収集

本研究で用いた Web サーバプログラムは当初標準のバッファありでデータを出力していたために 7 月 22 日から 8 月 7 日の間のデータについて，瞬停のための欠測が存在する．現在は，バッファなし出力にプログラムを変更したので，データの欠測の問題は，ほぼ解決されている．

(2) データの解析手法

収集したデータを Perl スクリプトにより処理し，IP アドレスとその逆引き結果および国別コードを含む CSV ファイルと日時，IP アドレス，コマンド等を CSV 化した観測データに変換し，Access によりデータベース化して解析のための基礎データとした．この際に，単純にすべての Web アクセスを均等にあつかったデータベースと，同じソースから短時間（数分以内）集中的来るアクセスをまとめて 1 件と数えるものと 2 通りのデータベースを用意した．このような手法を採用したのは，1500 件程度のアクセスに対して，ツールを用いた短時間集中型のアクセスによる影響が過大に評価されるのを避けるため，およびインシデントのソースとしては短期集中アクセスを 1 件と数えた方が実態を反映していると考えたからである．ツールによるデータアクセスを集約して計数した場合のアクセス件数は 610 件となる．このようにアクセスを集約することの問題点は，同じツールからのアクセスがアクセスの間隔によって独立か集約かに分けられてしまうことであるが，現状のデータでは集約型かそうでないかは明確に分離可能である．また，短期集中アクセスには，ほぼ ZmEu と呼ばれるツールが使用されている．

3. 記録されたデータから読み取れること

観測データについて，時間，ソースの国別，HTTP コマンド等の切り口から検討する．HTTP バージョンについては，1.1 と 1.0 が大半をしめており，比率は 3 対 1 であるが，

†1 中京大学
Chukyo University

ごく少数の 0.9 によるアクセスが存在する．

3.1 予備調査との比較

今回の調査で観察されたアクセス元の IP アドレスの個数は 349 個であった．予備調査での 69 件に比べると期間に対して件数の伸びはそれほど大きくない．また，予備調査と今回の調査に共通して出現するアドレスの個数は 6 個であった．

3.2 時間軸からの観察

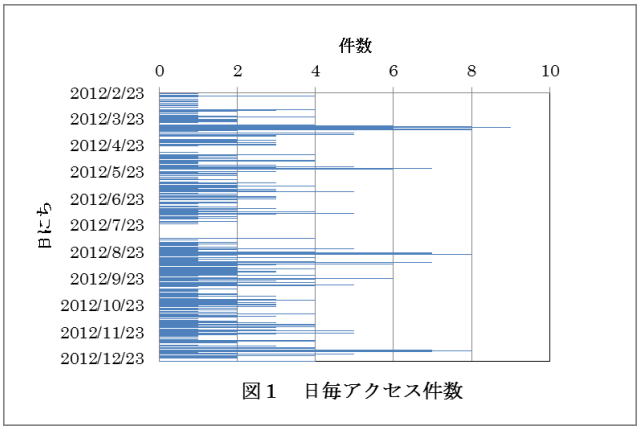

図 1　日毎アクセス件数

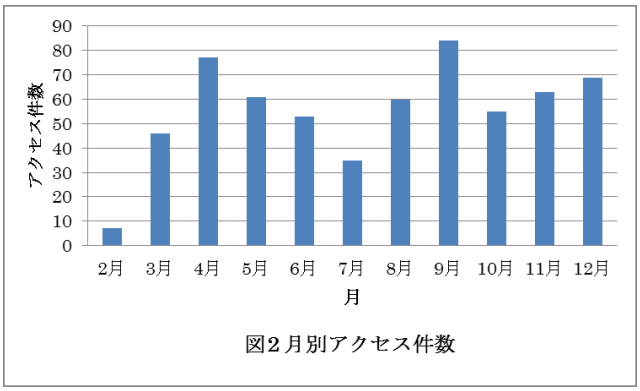

図2 月別アクセス件数

図 1 に日毎のアクセス件数を示す．平均のアクセス件数は 2.55 件で，1 日あたりの最高アクセス件数は 9 件である．月単位のグラフを図 2 に示す．2 月の計測開始時期および 7 月末から 8 月当初の欠測を考慮すると，ほぼ毎月一定のアクセス件数があると考えてもよいであろう．月平均のアクセス件数は約 55 件であった．

3.3 IP アドレスの国別件数

Perl の IP::Country モジュールにより IP アドレスの割り当て国等を検索し，アクセス件数の多い順に上位 5 個を表 1 に示す．IP 割り当て国の未解決が多いことが目につく．

表 1 国別アクセス件数

ドメイン	アクセス件数
BR	148
CN	123
US	47
KR	45
Amazon	26
未解決	44

この傾向は予備的調査や日常運用での経験とあまりかけ離れたものではない．相変わらずクラウドからのアクセスが無視できない数存在することも読み取れる．

3.4 ソースの特性

個々のアクセス元 IP アドレスについて，アドレスあたりのアクセス回数と IP アドレスの個数の関係を示したものを表 2 に示す．ツールからの集中アクセスを集約化した結果の全 610 件のうちで，1 回のみアクセスしたソースの数は 293 個，2 件が 33 個である．128 回アクセスしたものが 1 個存在する．この表から，ツールによって数分の間に集中的にスキャンするものとは別に，期間中に複数回アクセスしてくるものが存在することがわかる．

表 2 ソース毎のアクセス回数

アクセス回数/IP	IP 個数
128	1
35	1
13	1
10	1
9	1
7	1
6	1
5	1
4	2
3	10
2	33
1	293

3.5 HTTP コマンドとアクセス対象およびブラウザ

アクセスに使われた HTTP コマンドとアクセス対象をまとめて表 3 および表 4 に示す．この表では，スキャン・ツールからのアクセスも集約せずに集計している．表 3 で，「phpmyadmin」として示したものは，php で実装された DBMS ツールへの脆弱性スキャンを代表させものである．この種のアクセスの大半は ZmEu と呼ばれる脆弱性スキャナを使って行われている．このツールはアクセスにほとんど GET を使っており，表 3 の GET の数とおおよそ対応している．ただし，少ないながらも POST を使うものもある．ドキュメント・ルートへのアクセスには HEAD コマンドが使われることが多い．

「manager/status」へのアクセスは，Apache Tomcat に対するアクセスと考えられる．表には少数のためにリストしていないが，soapcaller.bs に対するアクセスも観察された．これは Drupal と呼ばれるコンテンツ管理システムへのアクセスのようである．さらに，サーバ外への URL へのアクセスを試みた形跡が 48 件見られた．

表 3 HTTP コマンド内訳

HTTP コマンド	件数
GET	1124
HEAD	394
POST	28
OPTIONS	10
CONNECT	2
空白	1

表 4 アクセス対象

目標	件数
phpmyadmin	900
/ or //	388
manager/status	90
translators.html	40

表 5 エージェント情報

エージェント	アクセス数
ZmEu	596
Mozilla/4	106
Mozilla/5	79
Java	33
その他	55
空白	688

表 5 にアクセスに使われたエージェントの主なものの一覧を示す．ただし，脆弱性スキャナが使用された場合には正しい値がセットされているとは限らない．

ZmEu と呼ばれるツールの使用が目立つ．Mozilla 系のブラウザとみられるものが ZmEu に続く．

Java の Web フレームワークへのスキャンが表 4 および表 5 から一定数存在していることがわかる．

3.6 データのクロス集計

3.1 節から 3.4 節のデータについてクロス集計を試みる．発行された HTTP コマンドと IP アドレスの割り当て国情報をクロス集計したものを表 6 に示す．この表の件数は表 1 とは異なり ZmEu のようなツールによるアクセスをすべて含んでいる．表 1 からは BR ドメインからのアクセス件数が最多であるが，表 6 からは BR ドメインからのアクセスは，ほぼ HEAD コマンドで占められていることがわかる．逆に BR ドメイン以外のドメインからのアクセスでは GET が卓越していることが読み取れる．この表から，単純に特定のドメインからのアクセス数だけでは脅威の大きさを評価できないことがわかる．この表で GET の件数は， ZmEu スキャナによるアクセスに起因するものが多い．

表 6 HTTP コマンドとソースのドメイン

	CN	KR	BR	US	EU	unk
GET	217	207	7	120	123	73
HEAD	79	6	144	26	2	34
OPTIONS	9		1			
POST			2	2		1
空白	1					

国別のアクセス元の数の月別推移をアクセス数上位について図 3 に示した．2012 年の後半に BR からのアクセスが増加していることが読み取れる．BR のアクセス件数の増加

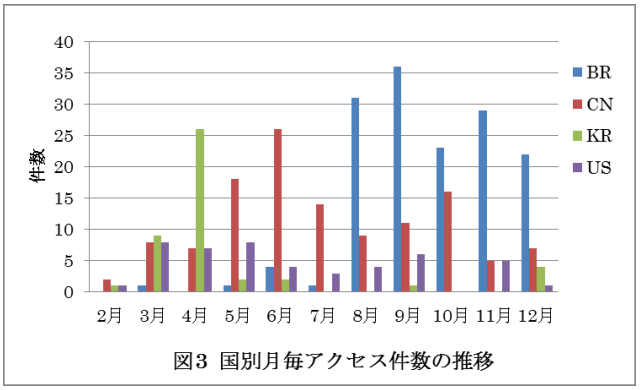

図3 国別月毎アクセス件数の推移

については次節の 1 個の IP からのアクセスの寄与が顕著であると考えられる．近隣国からのアクセスについては前半と後半であきらかに変動があるが，時事問題と関連しているかどうかについては検討していない．

POST コマンドについては，ほとんどが ZmEu もしくは類似のツールによると考えられる．

3.7 ステルスもしくは特異なアクセス

表 2 に示したソースについて，複数回のアクセスがあるものを選び出し，アクセスの時間間隔について調査した．IP アドレスあたりで，最もアクセス回数が多い 128 件の例については，アクセスは 2012 年 8 月に始まり，最短間隔 13 分，最長間隔 7 日，平均 1 日間隔でドキュメント・ルートに対して HEAD コマンドを発行している。この活動は健在も継続している．

同一のアドレスから 5 回以上のアクセスのあるものについて調べたところ規則性が認められるものが見つかっている．ひとつの例では，脆弱性スキャナ ZmEu を用いて正確に 4 時間 22 分の間隔で，php により実装された DBMS ツールの置かれているだろうと推測される場所を順次スキャンしていた．これは，いわゆるステルス型と呼ばれるものである．このような種類のスキャンでは，特定の脆弱性を 7 時間 40 分間隔でスキャンするものも存在した．

また，2 個の隣接するアドレスから約 17 日の間隔でアクセスするものも発見されており，このアドレスは予備調査にも現われ，現在も継続している．このアクセスの場合は，設置している疑似サーバをオープン proxy とみなして不特定の画像にアクセスを試みているように見える．

現在のところ，規則性の見られるものでアクセス間隔の最も長いものは 22 日であった．このサイトは，特定の URL に対して常に同じパラメータでアクセスを試みている．このサイトからのアクセスも継続している．

これらの IP アドレスについては，いずれも Project Honey Pot に記録されており，その多くが「害なし」かつ「意図不明」とされている．

3.8 その他のデータから見えること

収集した記録を観察していると，複数回アクセスに来るものについて，多くが疑似サーバからの否定応答にも関わらずアクセスを継続しているものが多い．これらのアクセ

スでは，一切プログラムの返す応答を考慮していないと考えられる．疑似サーバのドキュメント・ルート以下の検索ではなく，外部 URL へのアクセスの試みは 48 件あった．これらのアクセスの詳細を調べてみると．ランダムに画像を集めようとしているようにみえるもの，実在サイトからオートバイ画像を収集しようと試みるものも見られた．
予備調査時に 2 件見られた google.com へのアクセスは，今回の調査期間では 1 件だけであった．

4. まとめ

予備調査を含めると約 1 年間にわたり簡易なツールを使用して Web サーバへのアクセスを記録して解析した．

一切公開していない疑似サーバに対するアクセスは不正と言わないまでも正常なものとは言えない．

これらのアクセスのうち，複数回アクセスを試みているものについてはプログラムによるアクセスが大半のように推定される．なぜなら，疑似サーバの応答コードによらず動作を変えないこと，規則的にアクセスにくるものについて，その間隔が人が介在していると考えると中途半端な値であることによる．

「phpmyadmin」や Drupal CMS，Java フレームワークへの脆弱性スキャンが GET のものについては, 攻撃前の偵察行動と考えられなくもないが観測したデータからでは明確な意図は読みとれない．

外部 URL を指定したものについて, サーバの応答コードを確認していないようなので機械的アクセスと思えるが，その対象はランダムな画像と思われるもの等，奇妙な対象へのアクセスが目立つ．

これらの不正と考えられるアクセスについては，Web サーバの保守を適正に行う以外にないであろう．

5. おわりに

疑似 Web サーバにより Web へのアクセスの約 1 年間のデータの分析から，アクセスの規則性等を見出すことができた．ただし，プログラムの考慮漏れによる欠測もあり，アクセス傾向や長期の規則性等について分析を行うにはデータの個数や観測期間をさらに増やしたいと考え，観測を継続している．

この疑似サーバは非常にコンパクトかつ手軽なので，ボードコンピュータ等で実装すれば気軽に配置できるセンサーとして使えそうである．デーモン化およびネットワークを利用したログ機能の実装も今後の課題である．

6. 謝辞

本研究の一部は，文部科学省科学研究費補助金（基盤研究 (B) 22300288）の援助を受けている．ここに記して感謝の意を表する．

参考文献

1) 長谷川明生:”単純なハニーポットによるウエブアクセス動向の予備的調査”, 情報処理学会インターネット運用技術研究会研究報告，Vol.2012-IOT-16,No.51
2) IP::Country モジュール
http://search.cpan.org/~nwetters/IP-Country-2.21/
3) Project Honey Pot
http://www.projecthoneypot.org/list_of_ips.php