ハニーポットを用いたアドレスハーベスタと spam 送信者の spam 活動の調査

金髙一[†1]　松井一乃[†1]　加来麻友美[†2]　池部実[†2]　吉田和幸[†3]

spam 送信者は，ボットネットを用いて不特定多数のあて先に対して大量の spam を送信する．spam のうち，81.2%はボットに感染したエンドユーザコンピュータから送信される．spam 送信者は，spam のターゲットとなるあて先メールアドレスを入手すること，spam の送信手段であるボットネットを維持・管理することが重要となる．本研究は，アドレスハーベスティングと spam 送信に着目し，ボットネットの spam 活動を調査する．アドレスハーベスタ，spam 送信者がボットネットであるかの分類，ボットネットの活動拠点を調べるため IP アドレスの分布，どのような頻度で活動をするかアクセス回数の 3 つの観点から調査し spam 活動を考察した．ボットネットの分布に関しては，ハニーポットへアクセスしてきた IP アドレスをヒルベルト曲線にて 2 次元マトリックスへマッピングした. ヒルベルト曲線は，類似した IP アドレスを近傍に表示できる．ボットネットは集団感染をするという性質から，IP アドレスが近傍に表示された際，同じボットネットに所属している可能性がある．今回の観測から，ボットネットのアクセスのひとつの特徴として，連続する IP アドレスから同時に，短期間で大量のアクセスを試みることを確認した．

A Study for Spam Activities of E-mail Address Harvesters and Spammers Using HoneyPots

HAJIME KANETAKA[†1]　KAZUNO MATSUI[†1]　MAYUMI KAKU[†2]
MINORU IKEBE[†2]　KAZUYUKI YOSHIDA[†3]

Spammers send a large amount of spam to a large number of the e-mail addresses using botnet. End-user computers infected with a bot sent 81.2% of all spam. The following activities are important in spammers. (1)To obtain e-mail addresses for spam targets. (2)To manage the botnets for spam sending. Our research purpose is to study the activities of e-mail address harvesters and spam bots from access log of honeypots. In this paper, we investigated the botnets activity focused on spam activity with e-mail address harvesting and spam sending.We investigated botnet activities from the activity base in IPv4 address space of spammers and e-mail address harvesters, classification into botnets, and access frequency to the honeypots.We represent a mapping of the first and second octets of IPv4 addresses of botnet onto the Hilbert curve. The Hilbert curve visualizes the IPv4 address space in 2-dimensions. This curve keeps adjacent address physically near each other. If IP addresses appear in the vicinity, the IP addresses might belong to the same botnet. As a result, we verified one of features of botnets that amount of access from consecutive IP addresses have visited in short term.

1. はじめに

インターネットの急速な発展と普及に伴い, メールを始めとするネットワークを介したコミュニケーションは必要不可欠となっている．メールと通常の郵便を比べると，送信者は容易かつ安価に大量のメッセージを送信できる．これに伴いspamが大きな社会問題となっている．spamは受信者の意図を無視して無差別かつ大量に一括して送信されるメールを指し，UCE(Unsolicited Commercial E-mail)，UBE (Unsolicited Bulk E-mail)とも呼ばれる．現在，インターネットを流れるメールの約64%がspamである[1]．Symantec社の調査報告によるとspamのうち，ボットに感染したエンドユーザコンピュータから送信される割合は81.2%である[2]．メールが我々の生活にとって不可欠なサービスである現在において，添付によるウイルス感染，大量のspamに紛れた通常メールの紛失による受信者の精神的ストレスなどのspamによる被害は無視できない問題である．

アドレスハーベスティングは，spam送信の前段階として，ターゲットとなる受信者のメールアドレスを収集する行為である．一般的にspamは無差別かつ大量に送信されるため，多くのメールアドレスを収集する必要があり，spam送信とアドレスハーベスティングには密接な関係がある．

本研究は，ボットネットを用いたアドレスハーベスティングとspam送信それぞれのIPアドレスの近接関係，ハニーポットに対するアクセスの種類，アクセス回数を調査し，ボットネットによる被害を防ぐため，spam活動の特徴からボットネットの活動を考察する．

本論文の構成は以下の通りである．第 2 章で研究目的について述べ，第 3 章では関連研究について述べる．そして第 4 章で本研究の調査方法について述べ，第 5 章で調査結

†1 大分大学大学院工学研究科知能情報システム工学専攻
Course of Computer Science and Intelligent Systems, Graduate School of Engineering, Oita University
†2 大分大学工学部知能情報システム工学科
Department of Computer Science and Intelligent Systems, Faculty of Engineering, Oita University
†3 大分大学学術情報拠点情報基盤センター
Center for Academic Information and Library Services, Oita University

果について述べ，考察する．最後に第 6 章でまとめと今後の課題について述べる．

2. 調査対象の spam 活動

2.1 spam 活動とボットネット

インターネットを流れるメールの約 64%が spam であり，そのうち 81.2%が，ボットに感染したエンドユーザコンピュータから送信される．ボットとは，悪意を持った第三者がインターネットを通じてコンピュータを外部から遠隔操作することを目的とした悪性プログラムのことである．アンチウイルスソフトによるウイルス対策がなされていないコンピュータが多数存在するネットワークにおいてボットが 1 台でも感染すると，感染したコンピュータが属するネットワークのコンピュータがボットに集団感染することがある．インターネット上には複数のボットネットが存在しており，ボットネットはメールアドレスやソフトウェアのシリアルナンバーなどの重要な情報の収集などにも用いられる[3].

本研究ではボットネットの活動を調べるために，ボットネットを使った活動の中でも，特にボットネット使用率が高いとされる spam 活動に着目した．spam 活動には，spam を送信するだけでなく, spam を送信する前段階であるアドレスハーベスティングも含まれる．ボットネットを使用した spam 活動として，アドレスハーベスティング，spam 送信について調査した．

2.2 アドレスハーベスティング

spam 活動のうち spam 送信は，ターゲットとなる受信者の意図を無視して無差別かつ大量に一括して送信する行為である．本節では，spam 活動の残りのアドレスハーベスティングについて説明する．

アドレスハーベスティングは, spam のターゲットとなるメールアドレスを収集する行為を意味し，以下の 3 つの方法に分類される[4].

(a) 公開情報(Web ページ)からのメールアドレス収集
Web サイトを巡回し，Web ページからメールアドレスを収集する．

(b) 非公開情報からのメールアドレス収集
社外などへ公開していない情報システムに不正侵入し，顧客名簿などからメールアドレスを入手する．

(c) プログラムによるメールアドレスの生成
辞書ファイルを用いてローカルパート(@マークよりも前の部分)をプログラムにより生成し, あて先ドメインと組み合わせて，メールアドレスを生成する．有効なメールアドレスの情報を取得する目的で大量のメールを送りつけるディレクトリハーベスティング攻撃 (Directory Harvesting Attack; DHA)はこれに分類される．

アドレスハーベスタがメールアドレスを収集するうえで効率が良い方法は，ボットネットやプログラムを用いて自動でメールアドレスを収集する方法である．(b)の方法は確実に使用されているメールアドレスを入手することができるが，他の方法と比べてメールアドレス入手のためにかかる手間が大きい．(b)に比べ，(a)，(c)はプログラムを実行するだけで人の手をかけずに，メールアドレスを入手できる．さらに，それらのプログラムをボットネット上で実行すれば大量のメールアドレスを短期間で大量に入手することができる．アドレスハーベスティングは spam 活動の準備段階として非常に重要である．

2.3 ボットの分類

ボットは，悪意を持った第三者がインターネットを通じて感染したコンピュータを外部から遠隔操作して, spam 送信や，Web ページからの情報を収集するための悪性プログラムである．

本論文では, spam 活動のうちメールアドレスを収集するボットを”アドレスハーベスタ”，spam を送信するボットを”spam ボット”と定義する．

また，検索エンジンのインデックス作成のために，Web ページの内容を収集する良性のボットが存在する(例：Googlebot，Baiduspider)．本論文では，これらのボットを”クローラ”と呼ぶ．

3. 関連研究

本研究ではボットネットによるアドレスハーベスティングと spam 送信の spam 活動を調査する．そこで，関連研究としてハニーポットを用いて spam 活動を調査した研究について述べる．

Oliver ら[5]は，Web 上に存在するメールアドレスを収集するハーベスティングボットの収集プロセスとそれらを操作するハーベスタを 3 年にわたり調査した. その結果, 様々な spam の原因はハーベスティング活動を行う特定のわずかな人物・組織であること，また Web 上でメールアドレスを公開するにあたり, JavaScript を用いた難読化技法がメールアドレスを収集されることを防ぎ, spam を受信しないために有効であることなどを示した．

Wilfried ら[6]は，運用開始時期や mailto 掲載数の異なる複数のハニーポットを用いて spam を 8 ヶ月間収集し，どのハニーポットに spam が送られてきたかを調査し，spam トラフィックとしていくつかの種類に分類した. その結果, メールアドレスを公開して初めて spam を受信するまでの時間が最速で 13 日，最遅で 64 日であり，メールアドレス公開から spam 受信まで最低でも 10 日以上の間隔があること．また，多くの spam は公開しているメールアドレス宛

てに送信してくるが，一部の spam は存在しないアカウントへ送信してくることなどを観測し，greylisting などの spam 対策の排除率を検証した．

本研究では，アドレスハーベスタ，spam 送信者のハニーポットへのアクセスログから，アクセス元を S25R や whois によって分類することにより，ボットネットの発見，spam 活動を調査する．

4. アドレスハーベスタと spam ボットの調査

本研究は，アドレスハーベスティングと spam 送信に着目し，ボットネットの spam 活動を調査する．アドレスハーベスタ，spam 送信者がボットネットであるかの分類，ボットネットの活動拠点を調べるために IP アドレスの分布，どのような頻度で活動をするかアクセス回数の 3 つの観点から調査し spam 活動を考察した．

4.1 調査環境

本研究における調査環境の構成を図 1 に示す．アドレスハーベスタと spam 送信者の情報を収集する手順を以下に示す．

(1) Web サーバは，CGI(Common Gateway Interface)によってページ要求ごとに異なるメールアドレスを含む Web ページを返す．メールアドレスは，クライアントがページを要求したアクセス時刻をもとにして，自動生成する．メールアドレスにアクセス時刻を含むことにより，メールを受信した際，どのタイミングで収集されたかを判別できる．

(2) メールサーバは，Web サーバで公開したメールアドレスの他に，自身のドメイン宛で存在しないアカウントに送信されたメール，オープンリレーによる第 3 者宛へのメールを含むすべてのメールを spam として受信する．

本研究では(1)のアクセスログと(2)のメールログを調査することにより，アドレスハーベスタと spam 送信者の活動状況を調査する．

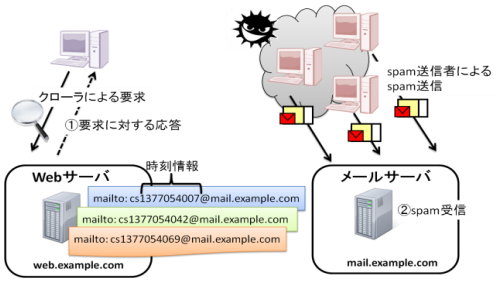

図 1　アクセス情報の収集

4.2 IP アドレス空間の表現方法

多次元空間の情報を一次元に写像する手法として，空間充填曲線が用いられる．空間充填曲線の代表的なものとして，ヒルベルト曲線[7][8]，ルベーグ曲線(Z-Ordering)，シェルピンスキー曲線，ゴスパー曲線などがある．二次元空間充填曲線は平面上において近隣のノードが一次元線上でも近くに配置されるように空間を充填するため，平面上のノードの近接性と変換後の一次元線上での近接性に相関をもたせることができる[9][10].

空間充填曲線を用いる利点は，図 2 に示す空間充填曲線を用いない場合と比べ，連続する IP アドレスを 1 つの塊として描写できるため，アクセス元の IP アドレスの近接関係の把握が容易になる．またアドレスハーベスタと spam 送信者に関する分布を比較する際に，IP アドレスがどの IP アドレスブロックに存在しているか分布の差分を調べる際に直感的に違いが分かる点にある．

図 3 に示すルベーグ曲線は 7-8 間，15-16 間，31-32 間，63-64 間，127-128 間では IP アドレスブロックが大きく離れている部分がある．シェルピンスキー曲線など他の曲線も基本図形の性質から同様に IP アドレスブロックが大きく離れる部分ができる．これに対してヒルベルト曲線は，IP アドレスブロックが大きく離れることなく近接性を保持したまま描写できる．このことから，本研究では他の空間充填曲線と比較して IP アドレスの近接関係の描写に優れているヒルベルト曲線を用いる．

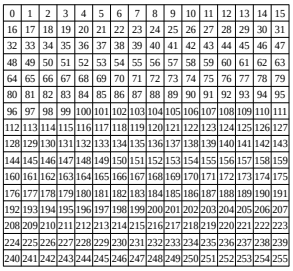

図 2　空間充填曲線を用いない 0~255 の分布の描写の一例

図 3　ルベーグ曲線を用いた 0~255 の分布の描写

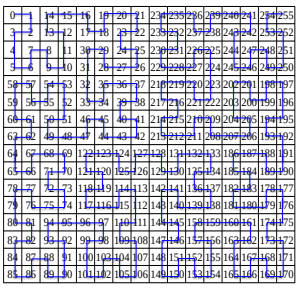

図 4　ヒルベルト曲線を用いた 0~255 の分布の描写

ヒルベルト曲線には図 5 に示す 4 つの基本図形がある．基本図形を ULD，DRU，RDL，LUR とする．この基本図形を 4 つのルールに従って再帰的に呼び出すことでヒルベルト曲線ができる．4 つのルールを以下に示す．↑，↓，→，←は各方向への描画を示す．

ルール 1 を例にルールを説明する．まず，DRU(n)が呼び出されると，RDL(n-1)を呼び出す．RDL(n-1)の描画が終了すると，RDL(n-1)の描画が終了した場所から↓方向に線を描画し，DRU(n-1)を呼び出す．以後同様に DRU(n-1)，→，DRU(n-1)，↑，LUR(n-1)の順で呼び出し，n=0 になるまで繰り返す．

・ルール 1 (Down-Right-Up)
DRU(n)=RDL(n-1), ↓, DRU(n-1), →, DRU(n-1), ↑, LUR(n-1)
・ルール 2 (Up—Left-Down)
ULD(n)=LUR(n-1), ↑, ULD(n-1), ←, ULD(n-1), ↓, RLD(n-1)
・ルール 3 (Right-Down-Left)
RDL(n)=DRU(n-1), →, RDL(n-1), ↓, RDL(n-1), ←, ULD(n-1)
・ルール 4 (Left-Up-Right)
LUR(n)=ULD(n-1), ←, LUR(n-1), ↑, LUR(n-1), →, DRU(n-1)

IP アドレスの第 1 オクテットをヒルベルト曲線で表現すると, クラス A~E のクラスフル IP アドレスは図 6 となる．

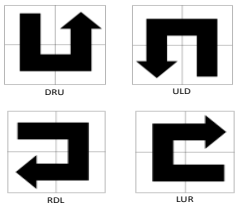

図 5　ヒルベルト曲線の基本図形

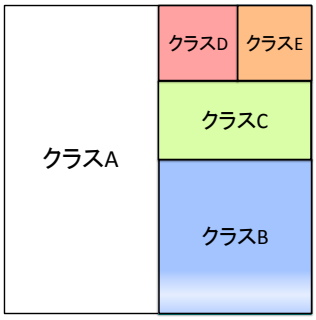

図 6　ヒルベルト曲線を用いたクラス A~E の
クラスフル IP アドレスの表現

本論文ではハニーポットへアクセスしてきた IP アドレスの上位 16 ビット(第 1 オクテット，第 2 オクテット)を 1 ブロックとして扱い，ヒルベルト曲線を用いてアドレスハーベスタ，spam 送信者の分布を表現する．その結果から，ボットネットとして使用されているアドレスハーベスタと spam 送信者の IP アドレス空間における活動拠点を知るために，分布と近接関係を調査する．

4.3 エンドユーザコンピュータの推定

Symantec 社の調査報告[2]によると，spam の 81.2%はボットに感染したエンドユーザコンピュータから送信されている．ボットに感染したエンドユーザコンピュータからの spam を排除する対策に S25R(Selective SMTP Rejection)がある[11]．S25R はメール送信元の IP アドレスを逆引きして得た FQDN(Fully Qualified Domain Name)を規則と照合し，エンドユーザコンピュータを推定し，SMTP(Simple Mail Transfer Protocol)アクセスを拒否する. 企業や学術機関, ISP が管理するメール送信サーバの多くは FQDN を設定している．しかし，エンドユーザコンピュータの多くは FQDN に IP アドレスの数字を含む場合や，FQDN を設定していないことが多い[12]. S25R で検知できる FQDN の例を図 7 に示す．

```
m1-110-0-0-0.example.jp
adsl-0-0-0-0.adsDRU.net
1.1.1.1.f.sta.codetel.net.do
```

図 7　S25R で検知できる FQDN の例

本研究ではボットネットの spam 活動を調査するため，ハニーポットで収集したアドレスハーベスタと spam 送信者をエンドユーザコンピュータとその他に分類した．以下に分類の手順を示す．

① S25R による分類
IP アドレスを逆引きして得た FQDN を S25R の規則と照合し，一致したらエンドユーザコンピュータに一致しなければその他に分類する．

② RBL(Real-time Blocking List)[13]による分類
S25R でその他に分類したホストが RBL に登録されていたならばエンドユーザコンピュータに，登録されていなければその他に分類する．

エンドユーザコンピュータからのアクセスであれば，ボットに感染している可能性が考えられる．また本研究におけるエンドユーザコンピュータは，一般家庭で使用されているコンピュータだけでなく企業や組織でも使用されているコンピュータを含む．

5. 調査結果

5.1 Web サーバ，メールサーバへのアクセス分類

調査期間は 2013 年 8 月 11 日から 2013 年 10 月 28 日までの約 3 カ月間である．ハニーポットへのアクセスには，ポートスキャンが含まれている．そこで，我々が開発・運用している不正通信検知システム[14]を用いて，アクセスしてきた IP アドレスをスキャン攻撃かどうか判定した. ただし HTTP(HyperText Transfer Protocol)と SMTP(Simple Mail Transfer Protocol)は, 誤検知が多いため不正通信検知システムの検知対象ポートから除外している．表 1 にハニーポッ

トで観測したアクセスのうち不正通信検知システムで検知した IP アドレスの数を示す.

表 1　ハニーポットへのアクセスのうち不正通信検知システムで検知された IP アドレスの数

ハニーポット	IP アドレスの総数	不正通信検知システムで検知した IP アドレス数
Web サーバ	414	48
メールサーバ	61	4

(集計期間：2013/08/11~2013/10/28)

不正通信検知システムでは，ハニーポットだけでなく他のホストや 80 番，25 番のポート以外に対して水平スキャンや垂直スキャンを仕掛けていた IP アドレスを攻撃者として検出している．今回の調査で収集したハニーポットのアクセスログと検知システムのログを照合した結果，ハニーポットと不正通信検知システムに同時刻にアクセスがあったものをポートスキャンと判断した.

ポートスキャンを除外したハニーポットのアクセスのうち，検索エンジンのインデックス作成のためのクローラからのアクセスをユーザエージェントから判断した．本論文では，“Googlebot”，“bingbot”，“Baiduspider”などをクローラとした．これらが正しいユーザエージェントか確かめるために S25R，RBL の結果を用いて判別した.
正規の Googlebot であれば, crawl-A.B.C.D-googlebot.com などからアクセスする(A.B.C.D は数字)．これらは S25R の規則から除外する．ユーザエージェントを Googlebot と偽装した IP アドレスからのアクセスの場合，FQDN は S25R のルールセットに該当すると推測できる.

これらのホストが RBL に登録されているか調査した結果，RBL では該当する IP アドレスはなかった．表 2 に分類したクローラ，アドレスハーベスタ，spam 送信者の IP アドレスを逆引きし，エンドユーザコンピュータかどうか判別した結果を示す.

表 2　クローラ・アドレスハーベスタ・spam 送信者の S25R によるホスト分類

	エンドユーザコンピュータ	その他	総数
クローラ	53	96	149
アドレスハーベスタ	169	48	217
spam 送信者	46	11	57

(集計期間：2013/08/11~2013/10/28)

表 2 の結果より，アドレスハーベスタや spam 送信者に関しては，エンドユーザコンピュータと思われるアクセスが 8 割ほどを占めており，ボットに感染したコンピュータからのアクセスの疑いがある．また表 2 に示すクローラに関して, エンドユーザコンピュータとして分類した 53 件は, ユーザエージェントを偽装したアドレスハーベスタの疑いがあるため，これ以降はアドレスハーベスタとして扱う.

表 3　Web サーバへのアクセス回数

	アクセス回数
クローラ	764 回
アドレスハーベスタ	768 回
アクセス総数	1532 回

(集計期間：2013/08/11~2013/10/28)

表 2，表 3 より観測したアドレスハーベスタの IP アドレスの数がクローラと比べて多いにもかかわらず，アクセス回数にあまり差がみられなかった．この理由として，クローラは新しい Web ページや，更新された Web ページを発見し，収集結果をデータベースに反映させるためにアクセスを繰り返す．それに対して，アドレスハーベスタは，メールアドレスの収集が目的のため，1 度アクセスしてメールアドレスを収集した Web ページに再訪することは少ないためであると推定される.

5.2 IP アドレスの分布

これ以降, Web サーバへのアクセスはクローラを除外し，アドレスハーベスタのみを調査対象とする.

図 8，図 9 はハニーポットへアクセスしてきたアドレスハーベスタと spam 送信者の分布をヒルベルト曲線にて表現し，IP アドレスブロックの色でアクセス数を表現している．論文中の図では，1 ドットを IP アドレス上位 8 ビットで表現している．実際には，1 ドットを IP アドレス上位 16 ビットとして描画し，調査した．1 ブロックを 8 ビットと 16 ビットで描写した場合では, 32 ビットで表現した場合と比べて IP アドレスの近接性が曖昧になるが, アクセス数を色で表現することで近接性の問題点を補完する.

図 8 (1)は，Web サーバへの単一の IP アドレスからの大量のアクセスがあったブロックである. これらの IP アドレスは全て逆引きが設定されていなかった．図 8(2)は ISP に所属する 47 の IP アドレスからのアクセスであり，逆引きが設定されていなかった．このアクセスは，ISP を介して身元を隠し，様々な情報を収集あるいは spam を発信するアクセスが多く，アドレスハーベスタからのアクセスの可能性がある.

図 9 (3)は, 1 つの IP アドレスから大量にアクセスがあったブロックである. アクセスは 1 秒間隔で 510 回であった.

この IP アドレスは逆引きが設定されており，PC 販売業者ということがわかった．図 9(4)は 13 個の連続した IP アドレスからのアクセスであり，1 つの IP アドレスから 2 秒間隔で 181 回のアクセスがあった．これらはフィッシングサイトとしてブラックリストに報告されている IP アドレスからのアクセスであり，すべて逆引きが設定されていなかった．

ハニーポットへアクセスしてきた IP アドレスは重複していなかった．今回の調査期間中では，アドレスハーベスタ，spam 送信者の関係性を確認できなかった．

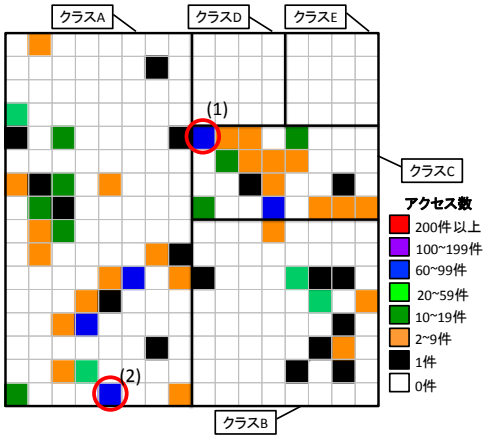

図 8　Web サーバへのアクセス数と IP アドレスの分布

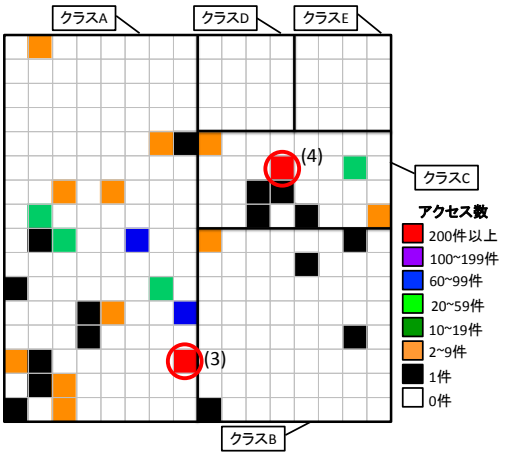

図 9 メールサーバへのアクセス数と IP アドレスの分布

5.3 IP アドレス別リクエスト分類

本調査では，アドレスハーベスタの Web ページへのアクセスを追跡するため，同じ内容の Web ページを複数用意した．用意したページは，以下のとおりである．

- index.html
- index-english.html
- kadai[1-20].cgi
- eng[1-20].cgi
- kadai[11-20].html
- eng[11-20].html

本調査では，index.html，index-english.html は，それぞれ日本語と英語で記述したトップページ，index.html からリンクしたメールアドレスを含む kadai[1-20].cgi と kadai[11-20].html，index-english.html からリンクしたメールアドレスを含む eng[1-20].cgi と eng[11-20].html の計 62 の Web ページを用いた．kadai[11-20]と eng[11-20]について同じ内容を CGI と HTML で別々に提供している．そのため今回の集計期間には kadai[11-20]と eng[11-20]について CGI のページと HTML のページ 2 つが混在している．図 10 にそれぞれのページへのアクセス頻度，表 4 にアクセスをページ毎に集計した結果を示す．表 4 では，アクセス回数の上位 5 件の IP アドレスを示している．

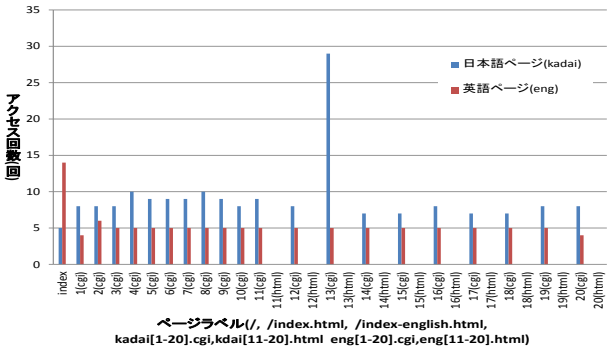

図 10　各ページへのアクセス頻度
(集計期間：2013/08/11~2013/10/28)

表 4　Web サーバへのアクセス上位 5 件

IP アドレス	アクセス回数	アクセスしたページ
A	69 回	phpMyAdmin
B	43 回	すべてのページ
C	40 回	index-english.html, kadai[1|2|7|8|9|10|12|14|15|17].cgi
D	38 回	kadai[3|4|5|6|11|13|16|18|19|20].cgi
E	23 回	index.html, index-english.html eng[1-19].cgi

(集計期間：2013/08/11~2013/10/28)

図 10 より，調査期間内に html ファイルのページへのアクセスはみられなかった．しかしアドレスハーベスタとして分類したホストの多くは偏りなく CGI ファイルのページへアクセスしていることが分かる．図 10 の kadai13.cgi だけが他のページと比べアクセスが突出していた．このときのアクセス傾向は，同じネットワークセグメントに所属する複数のホストからそれぞれ 1 回ずつのアクセスを確認した．観測した 2 つのネットワークセグメントからのアクセスのうち，1 つは逆引きを設定してあるホスティングサービスが運用しているネットワークからのアクセス，残りの 1 つは逆引きを設定していない ISP からのアクセスである．同じネットワーク上に存在し，同じ Web サーバの異なるページをそれぞれ参照していることから，これらはボットに

感染したエンドユーザコンピュータの可能性があると考えられる．アクセス元を調査したが，なぜこのページだけが増えているのか明確な理由は不明である．

表 4 に示す最もアクセスを試みた IP アドレス A は，phpMyAdmin に対するページ要求であり，脆弱性を狙った攻撃であると考えられる．IP アドレス C，D についてはクラス C に属する近隣のホストからのアクセスであり，互いに Web ページのファイル名に対して，重複がないようにアクセスしてきた．この 2 つのホストには逆引きが設定されていなかった. IP アドレス A 以外のホストについては，トップページから始まりほぼすべてのページを順に参照していたことから，ページに埋め込まれたメールアドレスなどの目的の情報を探す，ボットネットだと推察した．

5.4 spam メールに関する分類

メール送信者のメールサーバへ対する挙動を目的別に集計した結果を表 5 に示す．メール送信者の挙動は「自ドメイン宛へのメール」,「オープンリレーを目的としたメール」,「SMTP セッションを張るのみで何も行わない」の 3 つに分類した．

表 5　目的別にみた spam 送信者の分類

目的	サーバ数
自ドメイン宛	0
オープンリレー	9
セッションのみ	48

(集計期間：2013/08/11~2013/10/28)

表 5 より，spam 送信者は，メールアドレスを入手したターゲットへ直接 spam を送るのではなく，オープンリレーを許可している MTA を介して身元がばれないように spam を送る傾向がある．次に表 5 の”セッションのみ”においてどのタイミングでセッションを切断したかの割合を表 6 に示す．1 つの IP アドレスから複数のセッションを受ける場合もあるため，ここでは割合として示す．全セッションの数は 3401 回である．

表 6　SMTP コマンド内訳

SMTP コマンド	回数	割合	サーバ数
lost connection after CONNECT from ****	5 回	0.15%	3
possible SMTP attack	3 回	0.09%	2
did not issue MAIL/EXPN/VRFY/ETRN during connection to MTA	3393 回	99.7%	47

(集計期間：2013/08/11~2013/10/28)

表 6 に示すコマンド内訳のうち，セッションを張るのみで何もせずにタイムアウトになったことを意味する” did not issue MAIL/EXPN/VRFY/ETRN during connection to MTA” が 99%を占めていた．このログは，外部ネットワークからメールサーバを使用してメールを送信できるシステムの有無を確認しようとした場合に出され，不正中継を試みるためのサーバチェックの可能性がある．そのため，メールを送信することが目的ではなく MTA のサービスの生存確認，25 番ポートを狙った不正通信検知システムで検知されなかったホストからのポートスキャンの可能性がある[15].

表 7　メールサーバへのアクセス上位 5 件

IP アドレス	アクセス時刻	アクセス間隔	アクセス回数
F	10/8　07:47-7:55	1 秒	510 回
G	9/20　14:57-15:10	2 秒	465 回
H	9/8　13:18-13:18, 9/16　2:55-2:58, 10/19　15:45-15:47	2 秒	181 回
I	10/9　11:11-11:11	1 秒間に 10 回以上	65 回
J	9/18　2:35-2:35	1 秒間に 10 回以上	65 回

(集計期間：2013/08/11~2013/10/28)

表 7 は Web サーバへ最もアクセスを試みた上位 5 件を示す．IP アドレス F~J のアクセスの内容はすべて” did not issue MAIL/EXPN/VRFY/ETRN during connection to MTA”であった. それぞれの IP アドレスを学内ファイアウォールの外側のスイッチで取得したミラーパケットで確認した結果，IP アドレス I は, 学内ホストの 25 番ポートに対して水平スキャンをしており，スキャン攻撃後にメールサーバのサービスの稼働を確認して大量にアクセスをしてきたと推測する．IP アドレス G，H についてはメールサーバの 25 番ポートを狙ってアクセスしてきていた．また，IP アドレス H については，H の他に同じネットワークに所属する連続する 11 のホストから同じく 181 回のアクセスを観測しており，更にこれらのホストはフィッシング詐欺サイトとして報告されていた．IP アドレス I，J についてもメールサーバの所属するサブドメイン宛にアクセスをしていた．また，IP アドレス I 以外には逆引きが設定されていなかった．

一般ユーザが 1 秒おきに何度もアクセスを繰り返すということは，ほとんど無いと考えられる．そのため，今回のメールサーバへのアクセスは，自前の MTA をもつ spam 送信業者か，ボットに感染したエンドユーザコンピュータを使ったプログラムからのアクセスだと推測できる．

5.5 考察

ハニーポットを用いてアドレスハーベスティングと spam 送信について，spam 活動を調査した．その結果，ハニーポットへのアクセスにおいてボットネットの疑いのあるアクセスを発見した. Web サーバに関しては Web ページを探索してメールアドレスを収集する動きがみられたが，メールサーバに関しては SMTP システムの悪用を目的としたサービス確認のためのアクセスが大半であり, spam を受信できなかった．今回 spam 送信者に関しては十分なデータを集められなかった．Wilfried ら[6]の調査によると，メールアドレスを公開してから最遅でも 64 日で spam を受信したが，今回の 3 ヶ月間の調査では spam を 1 通も受信することはなかった．しかし，これから調査期間を増やし，時間が経つにつれてメールアドレスやドメインがアドレスハーベスタや spam 送信者に拡散し spam 送信や DHA のターゲットにされ，データを収集できると考えられる．

6. まとめ

ボットネットの活動をアドレスハーベスタと spam ボットの spam 活動に注目して調査した．ハニーポットとして仕掛けた Web サーバとメールサーバのアクセスログから，アドレスハーベスタや spam ボットがどのような IP アドレスブロックを拠点に活動しているか，また，それらのアクセス時刻やアクセスの規則性について調査し，特徴を発見した．今回の観測からひとつの特徴として，ボットネットは短期間に大量のアクセスを試み, 連続する複数の IP アドレスからのアクセスである場合，Web ページへのアクセスは，互いに Web ページのファイル名に対して，重複がないようにアクセスをする．メールサーバへのアクセスに関しては，アクセス間隔やアクセス回数が等しくなることを確認した．

また, 大学や研究所などの歴史的 PI(Provider Independent) アドレスを所有するクラス B よりも，CIDR(Classless Inter-Domain Routing)を用いて企業や ISP に IP アドレスを割り振られているクラス A，クラス C を拠点としたボットネットからのアクセスを多く観測したことから，組織や ISP などによる対策により，ボットネットの spam 活動を抑制する試みがなされているが，依然として greylisting や S25R などによるボットネットを排除するための spam 対策は有効であると考察する．

今回の調査でアドレスハーベスタと思われるホストを発見できたが，メールに関しては spam を送信してきたホストはなかった．

今後の課題として，観測期間を数カ月単位ではなく，数年単位で継続して調査し，アドレスハーベスタと spam 送信者の spam 活動の調査する．

参考文献

[1] シマンテックインテリジェンス月次レポート 2013 年 1 月号, 入手先 <http://www.symantec.com/content/ja/jp/enterprise/white_papers/sr_wp_spam_report_1301.pdf>(参照 2013-07-13)
[2] シマンテック インターネットセキュリティ脅威レポート第 17 号, 2012 年 4 月，入手先 <http://www.symantec.com/content/ja/jp/enterprise/white_papers/istr17_wp_201207.pdf >(参照 2013-09-12)
[3] D.Plohmann, E.Gerhards-Padilla and F.Leder, “Botnets:Measurement, Detection, Disinfection and Defence”,European Union Agency for Network and Information Security, CIIP and Resilience, Jul.2011,入手先 <http://www.enisa.europa.eu/activities/Resilience-and-CIIP/critical-applications/botnets/botnets-measurement-detection-disinfection-and-defence/at_download/fullReport>(参照 2013-11-05)
[4] 渡部稜太，愛甲健二, “スパムメールの教科書”，データハウス, 2006 年
[5] O.Hohlfeld, T.Graf and F.Ciucu, “Longtime Behavior of Harvesting Spam Bots”, proceedings of the 2012 ACM conference (ICM’12), pp.453-460,Nov.2012
[6] W.N.Gansterer and M.Ilger,”Analyzing UCE/UBE Traffic”,proceedings of the ninth international conference on Electronic commerce(ICEC’07), pp195-204,Aug.2007
[7] B.Irwin, and N.Pilkington, “High Level Internet Scale Traffic Visualization Using Hilbert Curve Mapping”, proceedings of the VizSEC 2007, Mathematics and Visualization, Springer Berlin Heidelberg, pp147-158,Oct.2007
[8] R.Munroe, Map of the Internet,入手先, < http://www.xkcd.com/195/>(参照 2013-11-05)
[9] 木谷友哉,中村嘉隆,”遅延状況を考慮した構造型 P2P オーバレイネットワーク構築法”,情報処理学会研究報告,マルチメディア通信と分散処理,2009-DPS-139,Vol3,pp.1-6(2009 年 6 月)
[10] C.Muelder and KL.Ma,”Rapd Graph Layout Using Space Filling Curves”,Proceedings of the IEEE Transactions Visualization and Computer Graphics 2008,Vol.14,No.6,pp.1301-1308,Nov.2008
[11] 阻止率 99%のスパム対策方式の研究報告,入手先 <http://www.gabacho-net.jp/anti-spam/anti-spam-system.html>(参照 2013-09-10)
[12] 松井一乃,小刀稱知哉,金髙一,池部実,吉田和幸,”大分大学宛のメール送信サーバの分布”，情報処理学会 第 61 回 CSEC・第 21 回 IOT 合同研究発表会, Vol.2013-IOT-21,No19,pp.1-7, May.2013
[13] RBL.JP,入手先<http://www.rbl.jp/>(参照 2013-11-18)
[14] 小刀祢知哉，天本大地，池部実，吉田和幸,” scan 攻撃検知システムを用いた被検知ホストの挙動についての調査”，電気関係学会九州支部第 65 回連合大会, pp.278-278, 2012 年 9 月
[15] 長谷川明生,鈴木常彦,山口栄作,”たかが spam されど spam:spam を呼び込んだら”,情報処理学会研究報告,分散システム/インターネット運用技術,2005-DMS37,pp75-78,May.2005