Vol.2013-CSEC-61 No.19
Vol.2013-IOT-21 No.19
2013/5/10

大分大学宛のメール送信サーバの分布

松井一乃[†1]　小刀稱知哉[†1]　金髙一[†1]　池部実[†2]　吉田和幸[†3]

本論文では, 大分大学宛のメール送信サーバの分布を調査する. メール送信サーバの IP アドレスの第 1 オクテット, 第 2 オクテットを基本として，通常メールの送信メールサーバと spam の送信メールサーバの分布をヒルベルト空間曲線にて 2 次元マトリックスへマッピングする．ヒルベルト空間曲線を使用することで，類似した IP アドレスを近傍に表示できるため，どの IP アドレスブロックからメールが送信されているのか明確になる．大分大学のメールサーバのログを調査した結果，メール送信サーバの IP アドレスを出力した画像から，ボットに感染している可能性のあるアドレスブロックを発見することができた. また多くの spam 送信サーバはそれぞれの IP アドレスから 20 通以下の spam を送信していることが確認できた.

A distribution of MTAs for mail to the Oita University

KAZUNO MATSUI[†1]　TOMOYA KOTONE[†1]　HAZIME KANETAKA[†1]
MINORU IKEBE[†2]　KAZUYUKI YOSHIDA[†3]

In this paper, we research a distribution of MTAs for mail to the Oita University. We use a Hilbert curve to map the first and second octets of MTA's IP addresses to a two-dimensional matrix while maintaining the similarity of IP address. We are able to check whether mail has been sent to Oita University from the IP address blocks. As a result, we discover the IP address blocks that may have been infected with bots from outputted images. And, we confirmed that spam MTA's IP address sent spam less than 20.

1. はじめに

インターネットの急速な普及と発展に伴い，電子メールをはじめとしたネットワークを介したコミュニケーションは不可欠になっている．これに伴い，spam が大きな社会問題となってきている．spam とは，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電子メールを指す．現在，ネットワークを流れる電子メールの約 64%が spam である[1]．spam による被害としては，フィッシング詐欺やメールに添付されたウイルスに受信者が感染, 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が挙げられる[2]．2012 年 10 月から 2012 年 12 月までの大分大学宛のメールを調査したところ，大分大学においても受信したメールの約 61%が spam であった．spam の送信元としては，企業が提供するフリーメール，ボットに感染したエンドユーザコンピュータなどがある．フリーメールからは通常メールも送信されるため，メールを送信する際の MTA の挙動だけでは通常メールと spam か判断することが困難である．しかし，エンドユーザコンピュータは通常，SMTP サーバを経由してメールを送信するため，エンドユーザコンピュータから直接送信先の MTA に SMTP コネクションを張り，通常メールを送信することはない．また，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．

そこで本研究では，大分大学宛のメール送信サーバの分布を調査する．調査方法は，メール送信サーバの IP アドレスの第 1 オクテットと第 2 オクテットを基本としてヒルベルト空間曲線を用いて 2 次元マトリックスにマッピングし, 大分大学宛のメール送信サーバの分布を出力した．視覚化することで, テキスト形式のログから送信元 IP アドレスを抜粋するよりも簡単に, 受信したメール全体の送信元 IP アドレスの分布を把握することが可能となる. 送信元 IP アドレスの分布を把握することで，ボットに感染している恐れのあるネットワークを発見できる可能性がある．発見したネットワークから送信されたメールは Blacklist へ登録する, 他のネットワークから送信されたメールよりも多くの spam 対策をするなどの処理を適用することで, ボットから送信される spam を受信する可能性が低くなる.

本論文の構成は以下の通りである．第 2 章で大分大学のメールシステムの構成について述べ，第 3 章では関連研究について述べる．そして第 4 章で大分大学宛のメール送信サーバの分布の調査について述べ，最後に第 5 章でまとめと今後の課題について述べる．

†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

2. 大分大学のメールシステムの構成

2.1 メールシステム構成

大分大学のメールシステムの構成を図 2.1，各プロセスでの spam 対策の構成を図 2.2 に示す．

大分大学のメールシステムではまず whitelist を用いてメールをプロセス 1 とプロセス 2 へと振り分ける．whitelist とは，信頼できる MTA（Mail Transfer Agent）の IP アドレスを列挙したリストである．実際は，iptables を用いて whitelist を参照し，登録されている MTA からのメールならプロセス 2，登録されていない MTA からのメールはプロセス 1 へ振り分ける． 図 2.2 に示すようにプロセス 1 では greylisting をはじめ様々な spam 対策を実施する．一方，プロセス 2 では信頼できる MTA からのメールであるため簡単な spam 対策を実施するにとどめている．プロセス 1 では 2012 年 2 月から milter manager を導入して greylisting の再送要求に伴う配送遅延がかかる通常メール数を減らしている[3].

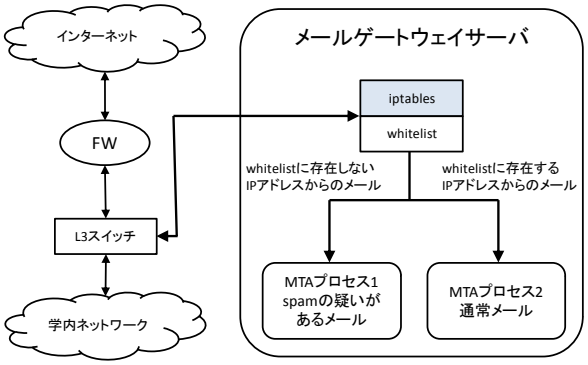

図 2.1 大分大学のメールシステムの構成

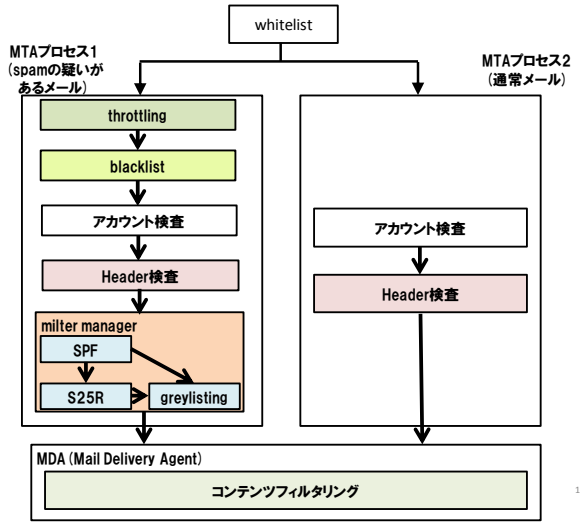

図 2.2 プロセスでの spam 対策

2.2 milter manager

milter manager は複数の milter を管理する milter である[4]. 通常 MTA では各 milter の適用の有無しか設定できない．しかし milter manager は各 milter の処理結果を他の milter の適用条件にすることが可能なため，動的に milter の適用の有無を決定できる．大分大学では milter manager を用いて図 2.3 に示す構成で SPF, S25R の判定によって greylisting の適用の有無を決定している．SPF，S25R を用いて通常メールと spam を分別することで，SPF，S25R による検査のみで受信できるメールが増え，greylisting による再送遅延を減らすことができている．また, SPF, S25R によって spam と誤検知した通常メールは greylisting によって救済することが可能である．

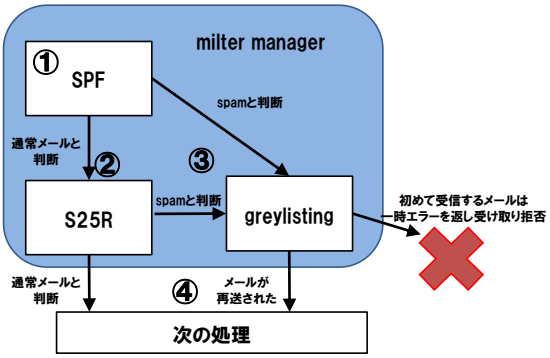

図 2.3 milter manager を用いたシステム構成

2.2.1 SPF

SPF(Sender Policy Framework)[5]とは SMTP によるメールの送受信において，送信者の正当性を検証し送信者のドメインの偽称を防ぐ送信ドメイン認証方式である．SPF はメールを受信時に，送信者であるメールアドレス（エンベロープ送信者）のドメインから送信されたものかどうかを検証することでメール送信者の正当性を確認する．

送信側はあらかじめ自ドメインの権威 DNS サーバ上に，自ドメインでメール送信を許可する MTA を特定する SPF レコードを登録する. 同時にメール送信を許可しない MTA からのメールの送信があった場合の判定を記述する．SPF レコードの記述例を図 2.4 に示す. 図 2.5 を例に認証の流れを示す．まずメールを受信すると(①)，受信者は送信者のメールアドレスのドメインの SPF レコードを送信元のドメインの権威 DNS サーバへ問い合わせ(②)，SMTP 接続元の IP アドレスと取得した SPF レコード(③)が一致するか確認することで，送信ドメインを認証する(④).

```
example.jp.  IN TXT " v=spf1 +ip4:192.168.100.0/24 ~all"
```

図 2.4 SPF レコードの記述例

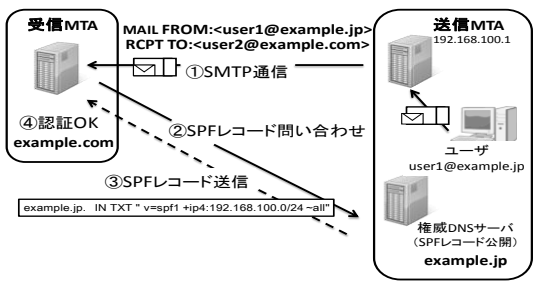

図 2.5 SPF の認証の流れ

Vol.2013-CSEC-61 No.19
Vol.2013-IOT-21 No.19
2013/5/10

2.2.2 S25R

シマンテックの調査報告[6]によると，spam メールの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[7]．S25R は SMTP アクセスを行ってきたメール送信サーバの 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 アクセスを拒否する．企業や学術機関などに管理されているメール送信サーバのほとんどは FQDN を設定しているが，エンドユーザコンピュータの多くは FQDN を設定していない．もしくは，エンドユーザコンピュータに対して FQDN を設定していることがあるが，メール送信サーバと比較すると，IP アドレスの下位 16 ビットなど多くの数字を含むことが多い．図 2.6 に S25R で検知した FQDN の例を示す．

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

図 2.6　S25R で検知した FQDN の例

2.2.3 greylisting

greylisting は「spam 発信 MTA（Mail Transfer Agent）は再送をしない」という仮説に基づいた対策手法である[8][9]．一時的に受信を拒否し，再送処理が行われたメールのみを受信する．ほとんどの spam 発信 MTA は仮説通りに動作するため，高い効果がある．しかし，MTA に再送を要求するため配送遅延が多く，1 時間以上の遅延が発生する可能性がある．

3. 関連研究

Wanrooij ら[10]は既存の blacklist を複数組み合わせ，独自のルールと閾値を設定し，メールに含まれる送信元 IP アドレスとメール本文に含まれる URI（Uniform Resource Identifier）を用いて，spam を判定する対策手法を考案した．この対策手法では，ネットワーク管理者が視覚的にわかりやすいようにするため，ヒルベルト空間曲線を用いて spam の送信元の IP アドレスを画像に出力している．画像にすることで，spam 送信者の多いブロックがわかり，ボットに感染している可能性のあるブロックを容易に発見することができる．Ward らのシステムでは spam のみを可視化するため，通常メールの送信 IP アドレスを把握することができない．spam を送信する IP アドレスの中にも通常メールを送信する IP アドレスは存在するため，spam 送信者のみの可視化では，通常メール送信者を spam 送信者と誤認識してしまう可能性がある．そこで本研究では，通常メールのみを送信する IP アドレス，spam のみを送信する IP アドレス，spam と通常メールの両方を送信する IP アドレスの 3 つを表示することで，より正確な spam 送信者の情報を得ることができる．

4. 大分大学宛のメール送信サーバの分布の調査

4.1 調査方法

調査期間は 2012 年 11 月 25 日から 2013 年 2 月 24 日である．この期間のメールサーバのログから，whitelist に記載がなく，プロセス 1 において処理したメール送信サーバの IP アドレスを抜粋し，その IP アドレスをヒルベルト空間曲線にて 2 次元マトリックスへマッピングする．ヒルベルト空間曲線を用いることで，IP アドレスの類似性を 2 次元空間上で近隣に配置できるので，ボットに集団で感染している可能性のある IP アドレスブロックを容易に発見できる[10]．今回の調査では，greylisting の再送要求に対して再送しなかったメール送信サーバのみを spam 送信サーバとして取り扱った．

4.2 ヒルベルト空間曲線

ヒルベルト空間曲線[11][12]は空間充填曲線の一つである．空間充填曲線とは多次元空間を 1 次元空間に写像する手法である．充填曲線を用いることで，多次元空間上で距離の近いデータを 1 次元空間へ写像後も比較的近い位置で表現できる．

ヒルベルト空間曲線はカタカナのコの文字の形状を基本図形とした自己相似図形であり，再帰的に呼び出す．

ヒルベルト空間曲線には図 4.1 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空間曲線ができる．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，ULD(n)が呼び出されると，LUR(n-1)を呼び出す．LUR(n-1)の描画が終了すると，LUR(n-1)の描画が終了した場所から上方向に線を描画し，ULD(n-1)を呼び出す．以後同様に,ULD(n-1),←,ULD(n-1),↓,RDL(n-1)の順で呼び出し実行する．n=0 になった場合は，なにもしない．

ルール 1
ULD(n)=LUR(n-1),↑,ULD(n-1),←,ULD(n-1),↓,RDL(n-1)

ルール 2
DRU(n)=RDL(n-1),↓,DRU(n-1),→,DRU(n-1),↑,LUR(n-1)

ルール 3
RDL(n)=DRU(n-1),→,RDL(n-1),↓,RDL(n-1),←,ULD(n-1)

ルール 4
LUR(n)=ULD(n-1),←,LUR(n-1),↑,LUR(n-1),→,DRU(n-1)

このルールに従ってヒルベルト空間曲線を描くと 2 の n 乗の正方形になる. n=3 の場合のヒルベルト空間曲線を図 4.2 に示す．また，ヒルベルト空間曲線を用いて図 4.3 にクラスフル IP アドレス（クラス A~E）を表現した.

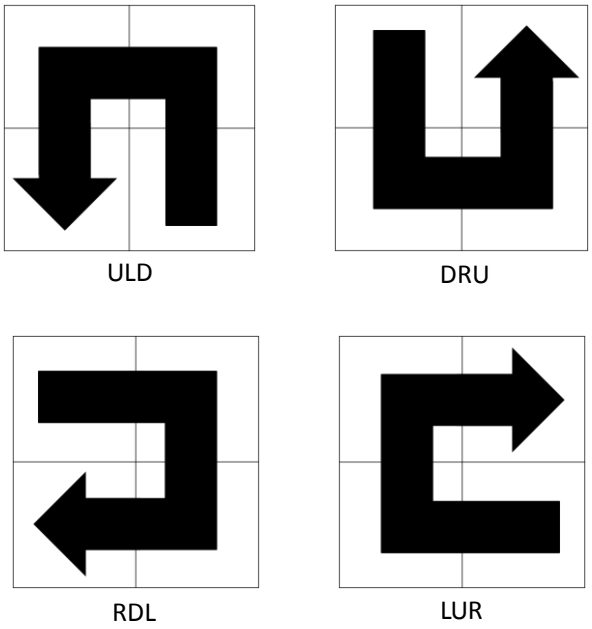

図 4.1　基本図形

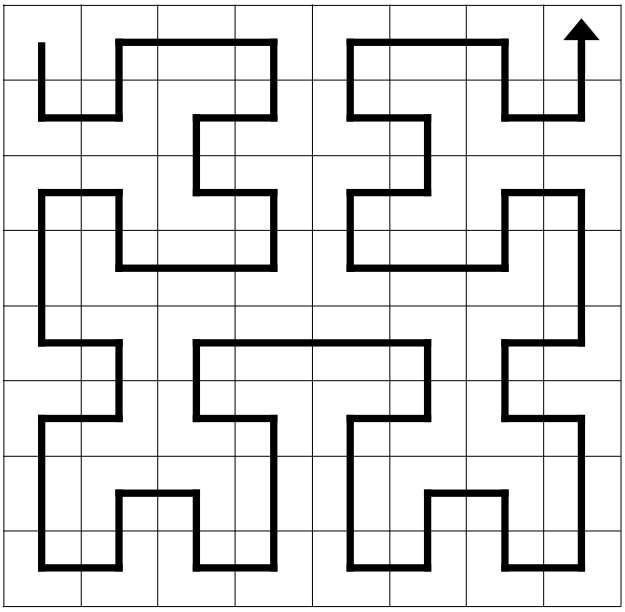

図 4.2　n=3 の場合のヒルベルト空間曲線

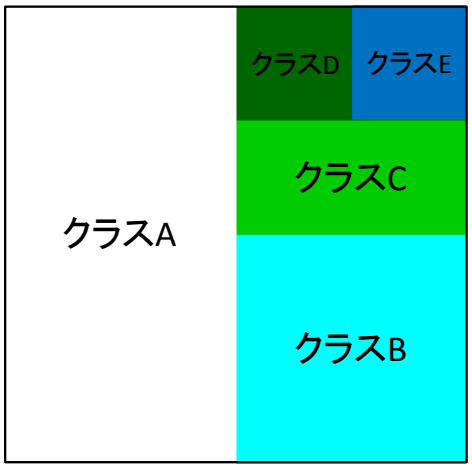

図 4.3　ヒルベルト空間曲線を用いたクラス A~E のクラスフル IP アドレスの表現

4.3 調査結果

2012 年 11 月 25 日から 2013 年 2 月 24 日までの約 3 ヵ月間の大分大学宛の通常メール送信サーバの分布を図 4.4, 大分大学宛の spam 送信サーバの分布を図 4.5, 通常メールのみを送信する MTA と spam のみを送信する MTA，通常メールと spam の両方を送信する MTA の分布を図 4.6 に示す．また，同期間に受信した通常メール数と spam 数を表 1 に示す． この期間の 1 つの IP アドレスあたりの平均メール送信数は 118 通，最大メール送信数は 111,044 通，最少送信メール数は 1 通である．今回の調査では，greylisting の再送要求に対して再送しなかったメール送信サーバのみを spam 送信サーバとして取り扱った．その他の spam 対策で spam と判定されたメールサーバの分布の例として blacklist に記載のあったメール送信サーバの分布を図 4.7 に示す.

図 4.4 に示した通常メールはクラス A，クラス C の IP アドレスからのメールが多く，特に黄色の丸で囲んだクラス C の特定のブロックから多く送信されていた．これらの IP アドレスを whois によって登録者の情報を調査したところ，該当する IP アドレスブロックには企業の MTA が設置されていた．通常メール上位 3 つの MTA を表 2 に示す．通常メールのメール受信数上位を占めているのは医師・医療従事者向けの学習サイトからのメールであった．これは大分大学に所属する会員がおり，その会員宛に企業が定期的にメールを送信したからだと考えられる.

一方，図 4.5 に示した spam の分布もクラス A，クラス C の IP アドレスからのメールが多い．また，通常メールと比較すると黄色の枠で囲んだ 176.0.0.1~191.255.255.255 のブロックから spam が多く送信されており， spam 送信数が 20 通以下の IP アドレスが多く存在することがわかった．spam 上位 3 つの MTA を表 3 に示す. spam の多くは多数のユーザへメールを一斉配信するサービスを提供するメール配送サイトからのメールであった．メール配送サイトから

Vol.2013-CSEC-61 No.19
Vol.2013-IOT-21 No.19
2013/5/10

のメールには通常メールは確認されなかった．通信販売サイトからのメールには通常メールも混在しており，spam と通常メールではドメイン名が異なっていた．

図 4.6 に示した通常メールと spam メール，両方を送信する MTA の分布をみると，通常メールのみを送信する MTA が水色の枠で囲んだクラス B の前半とクラス C に多く分布している．クラス B やクラス C の前半の IP アドレスは，大学や研究所のような歴史的 PI（Provider Independent）アドレスを所有する機関の MTA が多く存在した．歴史的 PI アドレスを所有する大学や研究所が spam を送信してくる可能性は低いと考えられる．一方ピンクの枠で囲んだ部分には spam のみを送信する MTA が多く存在した．クラス A，クラス B の後半は，CIDR(Classless Inter-Domain Routing)を用いて，企業や ISP に割り振られていると考えられ，特に黄色の丸で囲んだ場所では spam のみを送信する MTA が密集していた．IP アドレスの FQDN や登録者の情報を調査したところ，黄色の丸で囲まれた部分のうち 12%は，逆引きの PTR レコードが設定されてなかった．また，87%は S25R の規則にあてはまる FQDN が設定されており，残りの 1%は通常のメール送信 MTA であった．S25R の規則にあてはまる FQDN が設定されている IP アドレスはエンドユーザコンピュータである可能性が高いと考えられる．通常，エンドユーザコンピュータが SMTP サーバを経由せずに，直接メールを送信する可能性は低いため，黄色の丸で囲まれた部分の IP アドレスは，ボットに集団感染しているのではないかと推測できる．該当する IP アドレスはアルゼンチン，チリ，ウルグアイ，ブラジル，コロンビアなどの南アメリカの地域に割り当てられていた．

図 4.7 に示した blacklist に記載のあったメール送信サーバは通常メール送信サーバや spam 送信サーバよりも広く分布していた．図 4.7 に示した blacklist に記載のあったメール送信サーバは通常メール送信サーバや spam 送信サーバよりも広く分布していた．

表 1　2012 年 11 月 25 日から 2013 年 2 月 24 日までのメール受信数

総受信数	通常メール数 (A)	greylisting 排除数(B)	その他 spam 対策排除数
1,493,938 通	635,778 通	433,705 通	424,455 通

表 2　通常メール受信数上位の MTA

送信 MTA	受信数	表 1(A)に占める割合
医学学習サイト MTA(1)	12,855 通	2.0%
プロバイダ	12,812 通	2.0%
医学学習サイト MTA(2)	12,669 通	1.9%

表 3　spam 受信数上位の MTA

送信 MTA	受信数	表 1 (B)に占める割合
通信販売サイト	2,717 通	0.6%
メール配送サイト MTA (1)	2,484 通	0.5%
メール配送サイト MTA (2)	2,464 通	0.5%

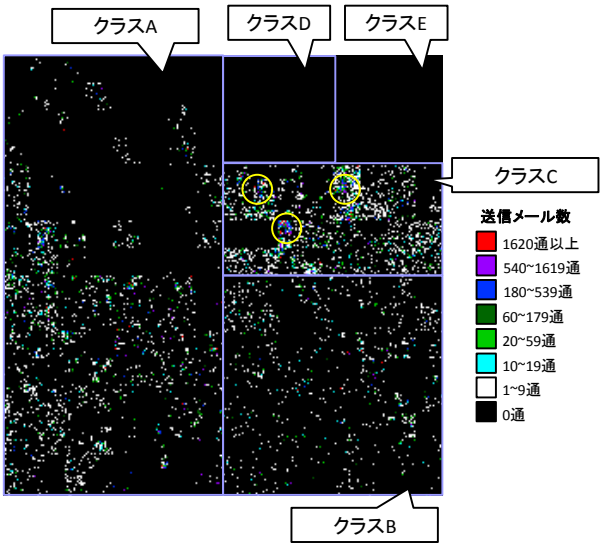

図 4.4　大分大学宛ての通常メール送信サーバの分布とメール数

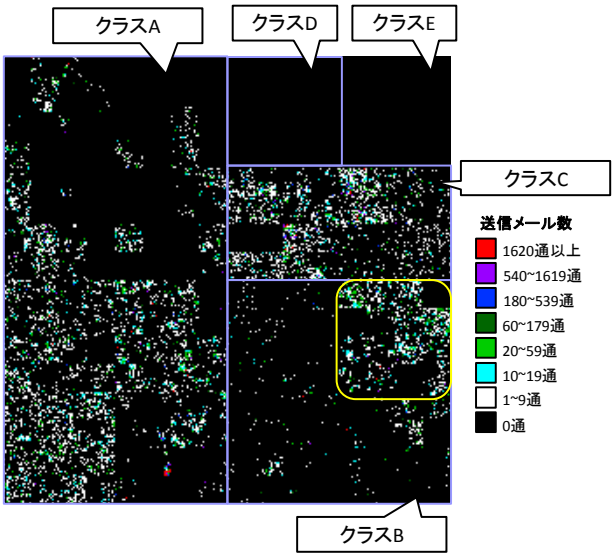

図 4.5　大分大学宛ての spam 送信サーバの分布とメール数

Vol.2013-CSEC-61 No.19
Vol.2013-IOT-21 No.19
2013/5/10

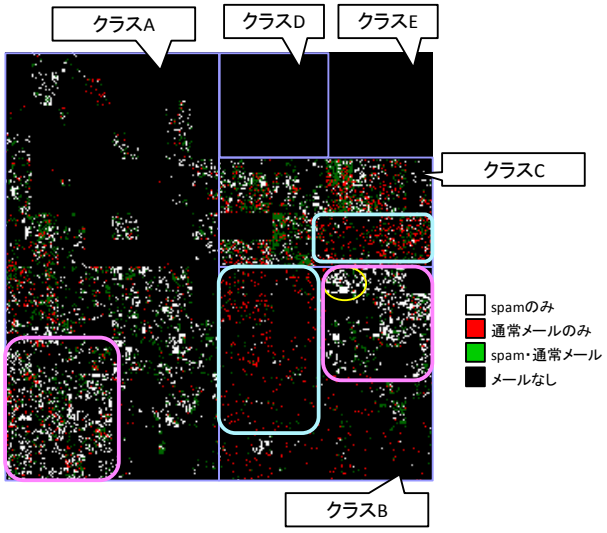

図 4.6　大分大学宛てのメールの送信サーバの分布

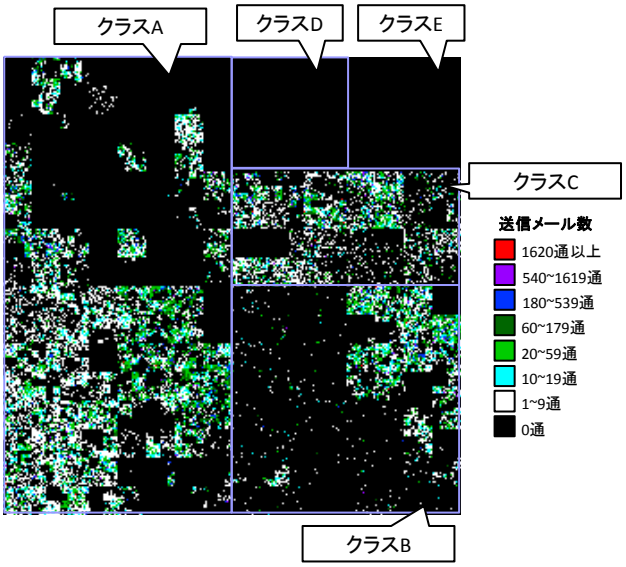

図 4.7　blacklist に記載のあったメール送信サーバの分布とメール数

5. おわりに

5.1 まとめ

本論文では，大分大学宛のメール送信サーバの分布について調査した．ヒルベルト空間曲線を用いてメール送信サーバの IP アドレスを視覚化した．ヒルベルト空間曲線を用いることで，類似した IP アドレスを近傍に表示できる．調査した結果，spam を送信するエンドユーザコンピュータの集団を発見した．エンドユーザコンピュータから SMTP サーバを経由せず，直接通常メールが送信される可能性は低いので，ボットに感染している可能性があると推測できる．また，spam 送信サーバの多くは 20 通以下の少量の spam を送信していることがわかった．ボットは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所属するエンドユーザコンピュータに一台でも感染すると，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に集団感染することがある．そのため，S25R の規則にあてはまる FQDN をもつ類似した複数の IP アドレスから spam を受信した場合は，その IP アドレスの近傍もボットに感染している可能性があると考えられる．S25R の規則にあてはまる FQDN をもつ IP アドレスが複数存在するネットワークからのメールは spam であると推測できるため，該当するネットワークからのメールの受け取り拒否をすることで，spam の受信数が減少する可能性がある．

5.2 今後の課題

現在のシステムでは，S25R で spam と判定されたメールは greylisting を適用している．greylisting は再送されたメールかどうか確認するために送信元 MTA と受信 MTA の IP アドレス，送信者と受信者のメールアドレスと受信時刻を一定時間保持している．greylisting が保持するデータが多ければ，MTA に多くの負担をかけてしまうため，ボットに感染している IP アドレスブロックからのメールは FQDN の規則への照合のみで受信拒否する S25R を利用するのが望ましい．しかし，S25R の判定では誤検知が発生する可能性があるため，全てのメールを S25R の判定だけで受信拒否すると通常メールを受信拒否する恐れがある．そこで，今後は以前 S25R で spam と判定され，greylisting の再送要求に応答のなかった IP アドレスブロックからのメールのみ S25R で受信拒否し，その他のメールが S25R の規則に当てはまった場合は，greylisting を適用する．このようにすることで，ボットから送信された spam による greylisting のデータ保持に伴う MTA への負荷をかけず，S25R による通常メールを誤検知する恐れも減少すると考えられる．

また，現在は IP アドレスの上位 16 ビットを用いて画像にしているため，より細かいメール送信元 MTA の把握をすることができない．よって今後の課題として，下位 16 ビットを用いた視覚化の方法を検討することが挙げられる．また，ヒルベルト空間曲線を用いて画像にした場合，画像の特定のポイントの IP アドレスを即座に知ることができないため，今後は画像上の IP アドレスを知りたいポイントをクリックすると IP アドレスを表示するなど，IP アドレスの表示方法を工夫していくことも課題である．

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, http://www.symantec.com/content/ja/jp/enterprise/white_papers/sr_wp_spam_report_1301.pdf
2)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
3)金髙一，松井一乃，池部実，吉田和幸: milter manager による低配送遅延を目指した spam 対策メールサーバの設計とその運用結果，情報処理学会第 5 回インターネットと運用技術シンポジウム，pp.8-15, 2012 年 12 月

4)milter を使った効果的な迷惑メール対策,
http://milter-manager.sourceforge.net/
5)W. Schlitt: Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1, RFC4408, ApLUR006
http://tools.ietf.org/rfc/rfc4408.txt
6)インターネットセキュリティ脅威レポート第 17 号,
http://www.symantec.com/content/ja/jp/enterprise/white_papers/istr17_wp_201207.pdf
7)阻止率 99%のスパム対策方式の研究報告,
http://www.gabacho-net.jp/anti-spam/anti-spam-system.html
8)Greylisting.org - a great weapon against spammers,
http://www.greylisting.org/
9)吉田和幸: greylisting による spam メールの抑制について,
情報処理学会研究報告, 分散システム/インターネット技術,
2004-DSM-35,pp.19-24,2004 年 9 月
10)W. V. Wanrooij, A.Pras: Filtering spam from bad neighborhoods, International Journal of Network Management, Vol 20 No.6, pp.433-444 , Nov. 2010
11)B. Irwin, N. Pilkington: High level Internet scale traffic visualization using Hilbert curve mapping, VizSEC, pp147-158, Oct. 2007
12)R. Munroe: Map of internet, http://www.xkcd.com/195/