中継ホストの地理的視覚化によるスパムメール対策の検討

向坂 真一† 小池 英樹†

†電気通信大学大学院情報システム学研究科
182-8585 東京都調布市調布ヶ丘 1-5-1
muko@vogue.is.uec.ac.jp, koike@acm.org

あらまし 迷惑メールの対策はフィルタリングやスコアリング等，様々な方法が使用されている．昨今の迷惑メールは日本国外から送られる場合が多く，また botnet 等を利用して中継している場合もある．本研究ではメールの中継ホストの地理的情報に着目し，通常のメールと迷惑メールの中継ホストの位置を世界地図に視覚化した．この結果を比較し，迷惑メールの対策方法を検討した．

Analysis of Anti Spam E-mail using Geographical Visualization of Relayed Hosts

Shinichi Mukosaka† Hideki Koike†

†Graduate School of Information Systems, University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 Japan
muko@vogue.is.uec.ac.jp, koike@acm.org

Abstract In order to block spam mails, filtering and scoring are widely used. These spam mails usually come from foreign countries and sometimes botnet relays the spam emails. In this paper, we focused on email relay hosts. By visualizing trafic routes of emails on the world map, we found the different attributes between normal mails and spam mails. We discuss how this visualization is used to analyze spam mails.

1 はじめに

従来よりスパムメールはインターネットにおける問題となっている．現在では botnet を利用したフィッシング詐欺との連携など，スパムメールの手口はますます巧妙になってきている．スパムメールは大量の無駄なトラフィックを発生させるためエンドユーザだけの問題ではなく，ISP 等のメールを受信する事業者にとっても解決したい課題である．ISP はスパムフィルタをサーバに設置したり，他の ISP の動的 IP アドレスからの接続を禁止する IP25B などの対策を組み合わせてスパムメールの削減を行っている．メールは確実な異常検知が難しいのでスコアリングしてスパムメールの可能性の高さを算出する．新しいスパムメールが出現しても大部分のルールをすり抜けなければスパムメールとして判定されることになるため，ルールの種類を増やす事はスコアリングをより確実なものとするためにとても効果的な方法である．キーワードによるスコアリングなどは学習が必要であるため，利用者が多いほど判定が正確になる．学習を必要としないルールであれば誰もが容易に効果を出せるスコアリングとなると考えられる．

ウィルスやワームは一般的に IP アドレスを利用して伝播していく．ランダムな IP アドレスを選択して伝播する方法が多いので，感染さ

れたホストが操作されてスパムメールを送信したとすると全世界から送られてくると考えられる．また，botnet 等を利用してスパムメールを送信する場合には別のホストを経由して送信する場合も考えられる．一方，正常なメールは個人利用では特定の相手から送信されるため，特定の地域からのみ送られてくると考えられる．

本論文では，スパムメールのスコアリングの一つとして地理情報を利用することを提案する．このことを確認するために，正常なメールとスパムメールを世界地図にマッピングした視覚化システムを構築し，その結果からスパムメールのスコアリングの一つとしての利用可能性を検証する．

2　関連研究

スパムメールの送信元地域については以前から統計的に調べられている [1][2]．どちらも一定期間のスパムメールの送信元の国の割合を調べている．スパムメールの送信元の国の比率は毎年変化しており，1 位の国も変わっていることがわかる．また，中継ホストがどの国を経由しているかは調べられてはいない．

大量のマルウェアの攻撃を地理情報と関係づけて視覚化する方法が提案されている [3][4]．送信元・送信先 IP アドレスを地理情報と関連づけて視覚化しているが，その経路は示されていない．メールは中継して送信されるため，この視覚化手法をそのまま利用して解析することは難しい．

3　メールと地理情報

電子メールはヘッダ情報と本文から成る．ヘッダ情報にはヘッダフィールドがあり，その中にはリレーしたホストが記録する Received ヘッダが存在する．この Received ヘッダには，ホスト名， IP アドレス，時間等が記録される．

IP アドレスをベースとしたマルウェア解析手法は多い．しかしながら SMTP では DNS の MX レコードを利用してメールを送信するため， IP アドレスをベースとした一般的なマルウェアとはアクセスの方法が異なる．そのため IP アドレスを使った解析方法は難しいので，ドメインを用いて解析する必要がある．また単純にドメインを解析した場合には， gTLD を使用していたときにどの国か判別できない．IP アドレスが割り当てられている国の情報を得るか，IP アドレスに対応する地理情報 (緯度・経度) に変換するサービスを使用することでどの場所から来ているのか判断することができる．

4　視覚化システム

情報量が多いため全体的な傾向を把握することが難しいと考えられるので地図上に視覚化した．実装したシステムの処理の流れを図 1 にアクティビティ図で示す．

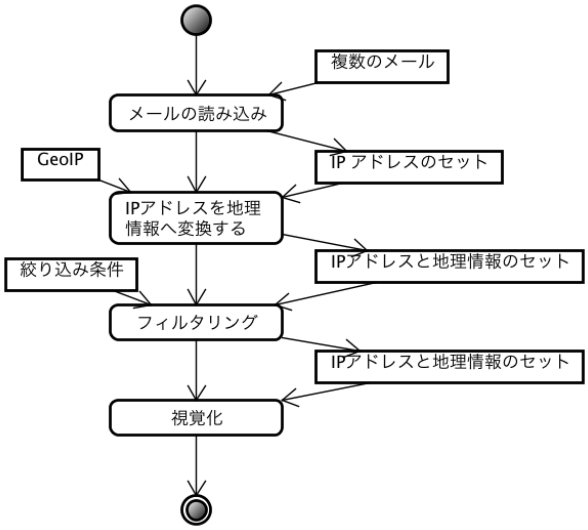

図 1: 処理の流れ

IP アドレスに割り当てられている国の情報ではなく，緯度経度の情報を使用して直感的なわかりやすさを優先した．メールの Received ヘッダに記載された IP アドレスを地図上にマップした．IP アドレスが世界地図上のどの位置に対応するかは GeoIP [5] を使用して求めた．視覚化には OpenGL を使用し，地球儀のような視覚化 (図 2)，正距円筒図法による二次元へ展開した視覚化の 2 種類を用意した．

地理的な位置に対応する IP アドレスを示すために IP Matrix [6] を使用した．IP Matrix は 2 次元マップで，縦軸が IP アドレスの 1st octet，横軸が 2nd octet を示し，左上が 0 となる．また，地図上の位置と IP アドレスの関係を線で示した．地図情報のみを見たいことが多いので， IP アドレスと地図上の座標との関係線および IP Matrix は表示・非表示を切り替えできるようにした．

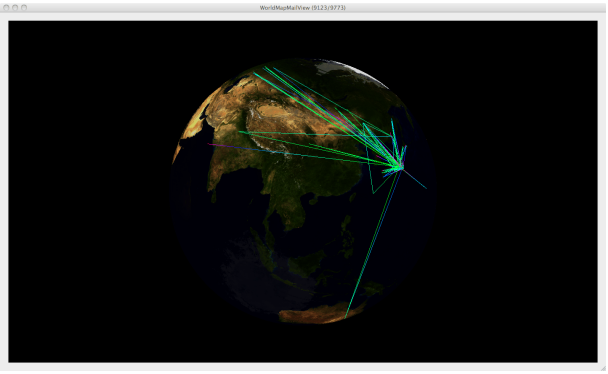

図 2: 地球儀のような表示

メールの経路の線は色をグラデーションで変化させるようにし，どの程度中継したかを認識しやすくした．送信元から送信先へ，red → magenta → blue → cyan → green に色を変化させた [7].

地球儀と正距円筒図法の切り替えは画面を瞬間的に切り替えてしまうと着目していた位置がどこに行ったか分からなくなるため，地図，経路の線，IP アドレスを示す線は関係を維持したままアニメーションで切り替えるようにした．また，地図の中心となる経度をなめらかに変更できるようにした他，拡大縮小もできるようにした．大量情報の拡大縮小・アニメーションを高速化するため shader を使用して実装した．

IP アドレスを示す面は，地球儀のように示す場合は地球儀の中に表示して線が重なりにくくなるようにした．

4.1 絞り込み

Honeynet Project によると，スパマーはセキュリティの低い国のホストや botnet を経由させるとされている [8]．botnet は様々な方法で bot を増やすが，基本的には IP アドレスを利用して広まるため，地理的に様々な場所に bot ができる．そのため地理的に離れた中継をしているものを抽出した．日本国内の中継を回避するために 20 度を閾値としてそれ以内の距離で中継したメールを除外した．また，単純に往復しているだけのものも除外した．

単純に中継している数が何回以上あるかでもフィルタリングできるようにした．

5 結果と考察

上記のシステムを用いて解析を行った．

5.1 結果

使用したデータは，筆者宛の 2001 年 1 月から 2010 年 6 月までのメールで，スパムメールの判定は手動で行った．正常なメールは 9773 件，スパムメールは 750 件あった．それぞれのデータを上記のシステムで視覚化した．

5.1.1 全てのメールを視覚化

通常のメールとスパムメールをそれぞれ全て視覚化した．スパムメールの中には異常な IP アドレスも存在した．図 3, 5 は通常のメール，図 4, 6 はスパムメールを視覚化したものである．

Received ヘッダに地理情報を持つ IP アドレスを持っていたメールを対象としているので，通常のメールは，9773 件中 9123 件，スパムメールは 750 件中 746 件を視覚化した．

IP アドレスで比較すると，明らかに使用している IP アドレスの数とばらつきが異なることがわかる．通常のメールは同じホストを経由していることがわかる．

地図情報で比較すると図のように明らかにその分布・ばらつきが異なっている．通常のメールはスパムメールの 13.0 倍も件数があるにもかかわらず送信元の位置は局所的である．

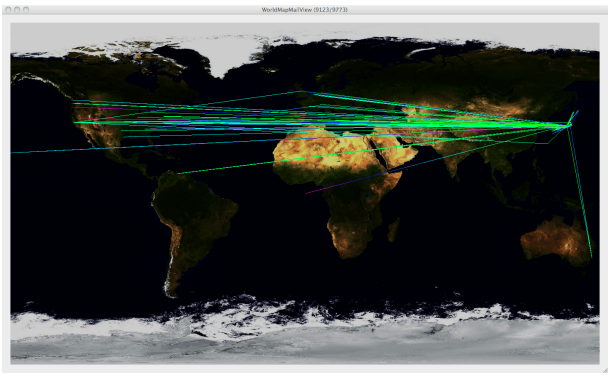

図 3: 全ての通常のメールの経路

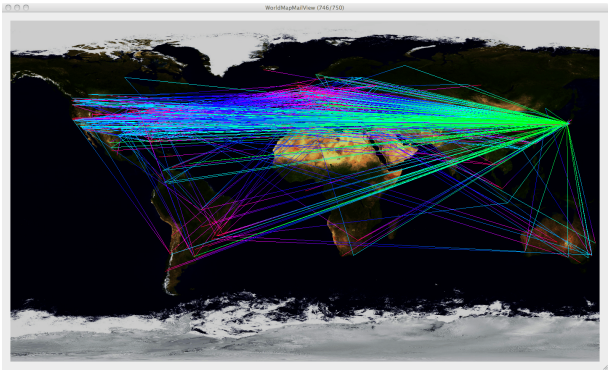

図 4: 全てのスパムメールの経路

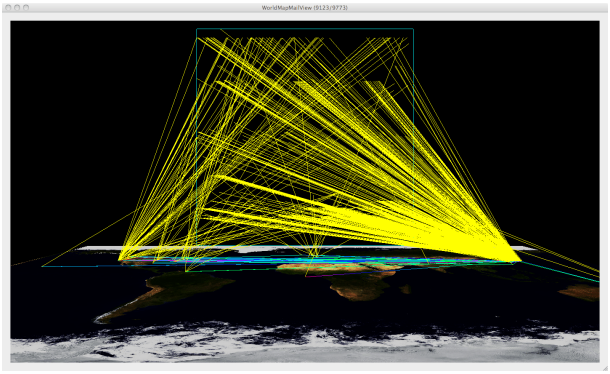

図 5: 全ての正常なメールの経路と IP アドレス

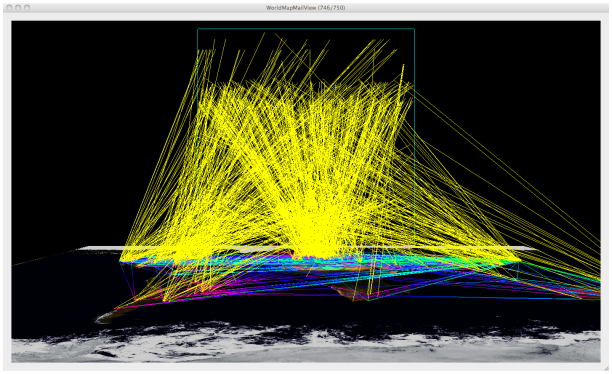

図 6: 全てのスパムメールの経路と IP アドレス

5.1.2　中継しているものを抽出して視覚化

通常のメールとスパムメールをそれぞれ地理的に離れた場所を中継しているもののみを抽出して視覚化した．図 7, 9 は通常のメール，図 8, 10 はスパムメールを視覚化したものである．

絞り込みの結果，通常のメールは，9773 件中 29 件，スパムメールは 750 件中 252 件を視覚化した．

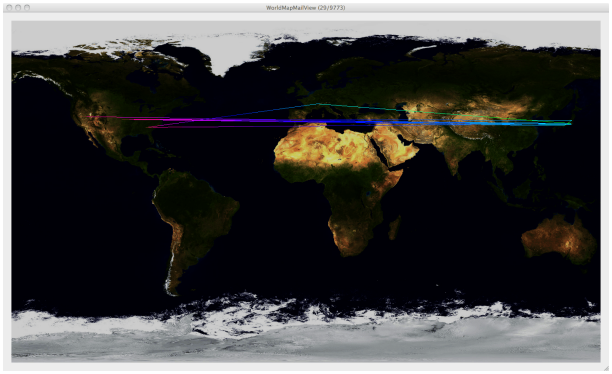

図 7: 抽出された通常のメールの経路

IP アドレスで比較すると，通常のメールはわずかな IP アドレスしか経由していないことがわかる．一方で，スパムメールは多くの IP アドレスを経由していることがわかる．

通常のメールは絞り込みをした方が地理的なばらつきが小さくなっているが，スパムメールはばらつきが小さくなりにくいことがわかる．

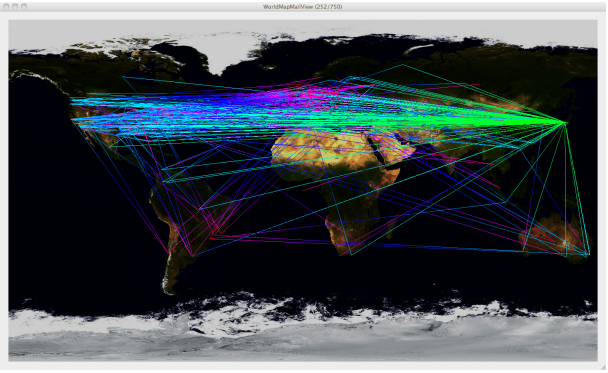

図 8: 抽出されたスパムメールの経路

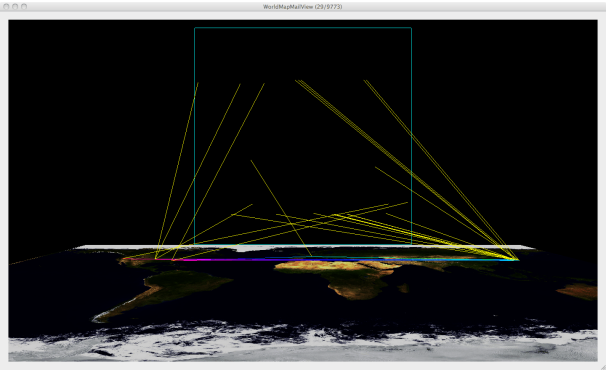

図 9: 抽出された通常のメールの経路と IP アドレス

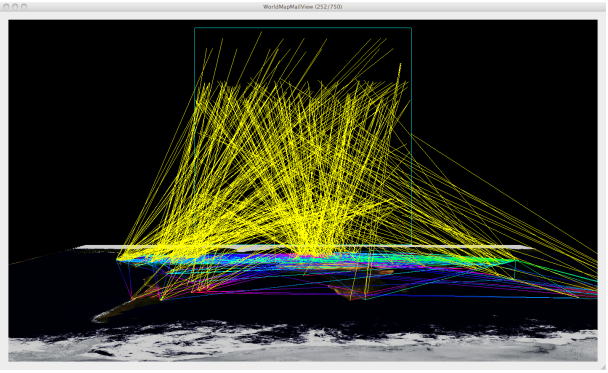

図 10: 抽出されたスパムメールの経路と IP アドレス

5.1.3 中継回数の多いものを抽出して視覚化

図 11 はスパムメールのうち，地理的情報を持つグローバル IP を持つ Received ヘッダが 3 つ以上持つもののみを視覚化したものである．ドイツ→ブラジル→アメリカ→日本，タイ→ベラルーシ→イタリア→日本 と送られてきたメールもあった．一方，通常のメールではこのように 3ヶ国以上を経由して送られるメールはなかった．

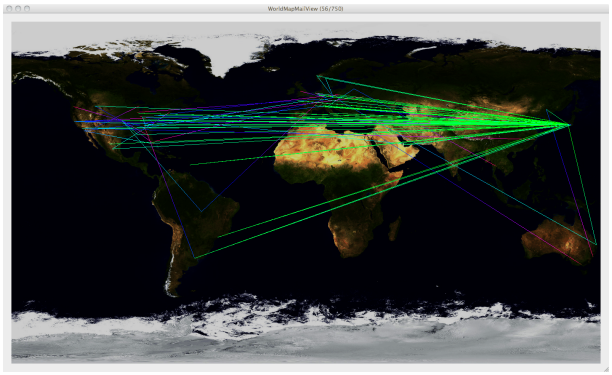

図 11: 中継点の多いスパムメールの経路

5.2 考察

視覚化の結果より，スパムメールは地理的にちらばっている事，地理的に離れた場所を中継しているメールが多い事がわかる．あらためて絞り込みの結果をまとめ，その割合を表 1 に示す．地理的に離れた場所を中継したスパムメールの数は正常なメールの 113 倍存在した．また，複数の国を経由して送られるメールもスパムメールの方が多かった．これらの結果より，スパムメールの可能性を示す指標の一つとしてメールの中継ホストの地理情報が使用できる可能性があることがわかった．

表 1: 絞り込み結果の比較

	通常のメール	スパムメール
割合	0.30% (29/9773 件)	33.6% (252/750 件)

これをスパムメールのフィルタリングのルールとして利用した場合には特に学習を必要とし

ないルールとなると考えられる．学習を必要としないルールであれば利用者の数に関係なく使用できる対策方法となり，企業等の団体でも使えるルールとなる．

正常なメールは地理的なばらつきが少なく，また地理的に離れた場所を中継する事が少ないことも分かった．一方，スパムメールは地理的なばらつきが大きく，地理的に離れた場所を経由する事が多い．多くの国を経由して送信されるスパムメールもあるが，これは数が少ないのでこの情報だけでスパムメールを除去するのは難しい事がわかった．

6　今後の課題

複数のメールアドレス宛のメールで地理情報の比較を行い，今回の結果が他の環境でも同一であるか確認する必要がある．日本では IP25B 等が行われているため，他の国の環境と異なっている可能性がある．そのため特に日本以外の環境でどのような結果になるのか確認する．

また，Received ヘッダはリレーするホストが記録する情報であるため，botnet では改竄される可能性がある．これらを検討し実験をした上で，地理情報がスパムメールのスコア付けに使えるか，それとも多人数で報告するなどして中継ホストのブラックリストの作成ができるのか，再度検討する．

7　まとめ

本論文では，マルウェアの伝播の方法とメールの通信方法の違いと，正常なメールとスパムメールの地理的な特徴の違いに着目し，メールの経路について視覚化し検討した．視覚化システムには絞り込みの機能を持たせ，著者宛のメールでいくつかの事例について報告した．その結果，空間的に離れた場所を中継しているメールが正常なメールよりスパムメールに多い事がわかった．この結果は学習を必要としないスコアリングなど，スパムメールの対策方法の一つとして利用できると考えられる．

参考文献

[1] G.Hulten, J.Goodman and R.Rounthwaite, Filtering Spam E-mail on a Global Scale. *Proceedings of the 13th international World Wide Web conference (WWW2004)*, 2004.

[2] IIJ, Internet Infrastructure Review vol.007, `http://www.iij.ad.jp/development/iir/pdf/iir_vol07.pdf`

[3] 向坂真一, 小池英樹, 内部ネットワーク監視を目的とした時間・論理・地理情報の統合的視覚化システム情報処理学会論文誌, Vol.49, No.11, pp.503-512, 2008.

[4] Y.Hideshima and H.Koike. STARMINE : A Visualization System for Cyber Attacks. *Asia Pacific Symposium on Information Visualization (APVIS2006)*, pp. 131-138, 2006.

[5] MaxMind Inc. GeoIP, `http://www.maxmind.com/`

[6] 大野 一広, 小池英樹, 小泉芳, IP Matrix: 広域ネットワーク監視のための視覚化手法, 情報処理学会論文誌, Vol. 47, No.4, pp. 1077 - 1086, 2006.

[7] Bailey, Mike and Cunningham, Steve, Introduction to CG shaders, *ACM SIGGRAPH ASIA 2008 courses*, 2008

[8] Honeynet Project, Know your Enemy: Phishing, `http://www.honeynet.org/papers/phishing/`