L-002

スパムメールサーバの発信行動変容の推定

Estimation of SPAM servers evolution in transmission behavior

山口　翔生 †　　中平　勝子 †　　北島　宗雄 †
Kakeru Yamaguchi　　Katsuko T. Nakahira　　Muneo Kitajima

1　はじめに

近年のスパムメール増加によって, ユーザの必要情報取得の負荷が増大している.

必要情報所得の負荷を軽減する手法には, ユーザ側でメールフィルタを行うベイジアンフィルタ方式と, 受信サーバ側で送信元フィルタを行うブラックリスト方式がある. ベイジアンフィルタ方式は現在の主たるスパムメールの防御法である. これはメールコンテンツからベイズ統計に従って禁止ワードを見つけ防御する. ブラックリスト方式はスパムを送信したと判断されたスパムメールサーバ (以降, スパムサーバ) をリストに追加する形で防御する.

ベイジアンフィルタ方式はメールコンテンツの禁止ワードを学習することで, スパム排除の精度が高くなる. 一方, メールコンテンツの言語や文章の構成によって精度が変化する欠点を持つ [1, 5]. ブラックリスト方式はコンテンツ内容を必要としない点で優位であるが, 誤ったサーバの登録やスパムサーバからハムサーバへの回復などを検知しないため, 登録された情報が自動的に削除される事がなく柔軟性に乏しい.

ブラックリスト方式の問題点を解決するため, メール受信側がスパム・ハムサーバを分類する必要があり, そのためにはユーザにメールを発信するサーバの行動を継続して観測する事が必須となる. スパムメールを発信したサーバを継続して観測する事で, 該当サーバをモニタすることができると考える.

本稿ではスパムサーバの行動と変容を観測する事で現在のブラックリスト方式の問題点を解決し, コンテンツ内容に左右されないスパム防御法式の作成を目標とする. スパムサーバの行動とは, ある期間におけるスパムメールの送信パターンを示し, 変容とはその時間変化を示す. ブラックリスト方式によるフィルタリングの研究には [4] がある. この研究ではスパムサーバの地理的, 論理的位置や固有の特徴量を利用して確率的なブラックリストフィルタを作成している. 本稿ではこれに個々のスパムサーバおよび, 国ドメインのスパム発信行動変容の 2 点を総合的に見ることでブラックリストフィルタを作成しようと試みる. 本稿では検討項目の 1 つである国別のスパムサーバ発信行動変容の傾向をその国のスパムメール数, スパムサーバ数, 地理的な情報などから推定する.

2　メール送受信とスパム

メールの送受信という行為 $\boldsymbol{f}$ は 3 つのオブジェクト, メール利用者 $\boldsymbol{u}$, メールサーバ $\boldsymbol{S}$, メール本体 $\boldsymbol{Ct}$ を用いて

$$\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{S}, \boldsymbol{Ct}) \tag{1}$$

と記述する.

$\boldsymbol{u}$ は, $u(madd)$ で表す. $madd$ はその利用者のメールアドレスを表す. $\boldsymbol{S}$ は, $S(time, domain, IP, protocol())$ で表す. $time$ はメール送受信時刻, $domain$ はドメイン名, IP はサーバ IP を表し, $protocol()$ はそのサーバのメールプロトコルの処理を表すメソッドである. $\boldsymbol{Ct}$ は, $Ct(text, time)$ で表す. $text$ はメールの内容を, $time$ はメールが作成された時間を表す.

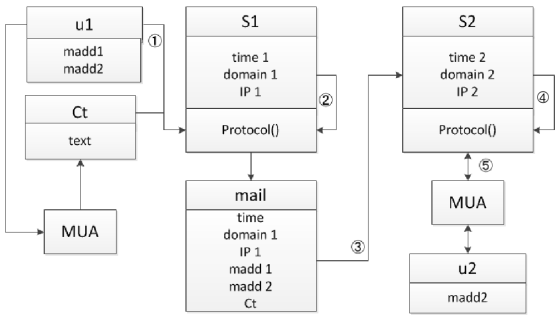

図 1　メール送受信図

図 1 はメール送受信の様子を表したものである. **1.** メール送信者 $u1$ は MUA を用いて Ct を作成し, それに送信元メールアドレス $madd1$, 送信先メールアドレス $madd2$ および, 作成時刻 t をメールヘッダとして付与し, メールを作成する. その後, サーバ $S1$ に送信する. **2.** $S1$ のメソッド $protocol()$ は, $u1$ から送信されてきメールに $S1$ の情報である $time, domain1, IP1$ をメールヘッダとして追加する. **3.** $S1$ の $protocol()$ は $madd2$ から $domain2$ を取り出し, $IP2$ に変換して, 送信先のサーバ $S2$ を検索しメールを送信する. **4.** $S2$ の $protocol()$ により $S1$ から送信されたメールを $S2$ 内に保持する. **5.** $u2$ は MUA を介して $S2$ に保持されているメールを受信する. メールヘッダには $madd1$, $madd2$, C_t および $S1$ の情報, メール発信時刻 $time$, ドメイン名 $domain1$, サーバ IP $IP1$, および, メールが通過した経路情報が含まれる.

通常のスパム判定の場合, ベイジアンフィルタ方式の場合には Ct を, ブラックリスト方式の場合には $IP1$ を, それぞれメール受信サーバ, もしくは MUA においてチェックすることで行われる. 本稿では, メール受信サーバにおいてスパム判定を行う. ある特定のスパムサーバから明らかに異常な量のメールが送信される状態が継続的に観察されれば, そのサーバをスパムサーバであると判定する. 従って, 計測すべき量は $S1$ からある時刻, もしくはある時間間隔に送信されるメール総量をもって決定する.

次にスパムサーバの行動, および変容を定義する.

スパムサーバの行動は, 時間の関数であり, ある時刻 t においてサーバから発信されるメール送信量 $m(t)$ の時間変化と定義する. スパムサーバの行動は, その地理的分布や所属する国の状況に影響される.

スパムサーバの変容は, スパムサーバの行動の時間変化で定義する.

次に国ドメインのスパム発信行動変容について考える. 国ごとの IT に関する方針, 法令により, その国独自のスパムサーバの行動があることは様々な論文, レポートで言われており [2, 3, 4], この考察は重要である. 本稿では国ドメインの行動変容を次の方法で分析する. 行動変容を考えるにあたり, 観測範囲という概念を定義する. 観測範囲は, 国の緯度経度を元にいくつかのセルに分割した際, スパムメールを送信するサーバを持つセルを指す. サーバの物理的位置や所属国の検索

† 長岡技術科学大学

には GeoIP*を用いた．ある国 i における観測範囲は，t において観測範囲内に存在するスパムサーバ数 $s(t)$, およびそこから送信されるメール送信総量 $M_t = \sum_k^{s(t)} m_k(t)$ と，その国固有な特徴量 A_i で考察できる．A_i には，国土面積，インターネット利用者数，割り当て IP 数，違法行為に対する取り締まりの強さなどが含まれる．

3　国ドメインでのスパム特性

ここで, スパムの国特性をみるため, 国 i に存在する，t における各観測範囲の $s(t)$ と M_t の分布特性を考察する．

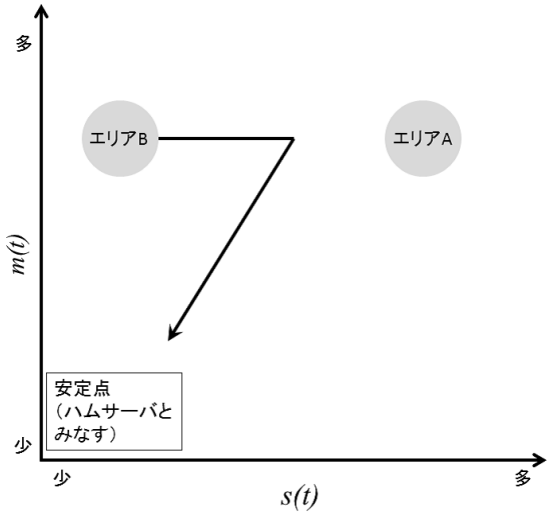

図 2　各観測範囲の $s(t)$ と M_t の分布特性

図 3 における観測範囲の分布と, それによる国の傾向は, スパムサーバ 1 台あたり同数のスパムを送信するとした場合, 例えば次の様な傾向が考えられる．

case1. 特定観測範囲からスパムを大量に送信する国は, 観測範囲ごとのメールの送信量の差が開いていると考えられる．図中ではエリア A の位置に特定観測範囲が孤立して存在し, 残りの観測範囲は左下に固まっている場合が多いと予想される．

case2. 国中にスパムを発信しているサーバが分散する場合は, 図中での観測範囲間が開かないと予想される．同じ理由で 1 つ 1 つの観測範囲の $m(t)$ は, 一部の地域にスパムメールが集中する国と比較して, 低い値になる．そのため各点の $m(t)$ の値は小さいが, $s(t)$ の値は高く, 観測範囲の分散が少ないのでエリア B の位置に観測範囲が集中しやすい形になると予測できる．

こうした傾向を時間軸に沿って観測する事で, その国の行動変容を観測できる．この国の観測範囲分布は時間が進むごとに図中の様々な方向に移動していく．**case2.** を例に国 i の発信行動変容を予測する．先に述べたように, この様な傾向がある国はエリア B の位置に観測範囲が固まっている．それがスパムサーバの摘発などで, 観測範囲数が減少したとする．摘発されたサーバを利用していたスパムメール生成者が, 別の既存のスパムサーバを利用し始めるため, 1 観測範囲あたりのスパムメール量は増加する．すると観測範囲分布はエリア B から右に移動する．その後, スパムサーバの摘発が続くのなら観測範囲数は減少し, 一部の地域が突出する事になり, 目立つ観測範囲から摘発の対象となる．それが続くことで国全体のスパムメール, スパムサーバが減少し, 左下の安定点に落ち着くといったストーリーが考えられる．

図 3 は中東, アジア, 欧州から代表的な国を選択しスナップショットを作成した．アジアの国は横軸に伸び, 観測範囲同士の距離が中東

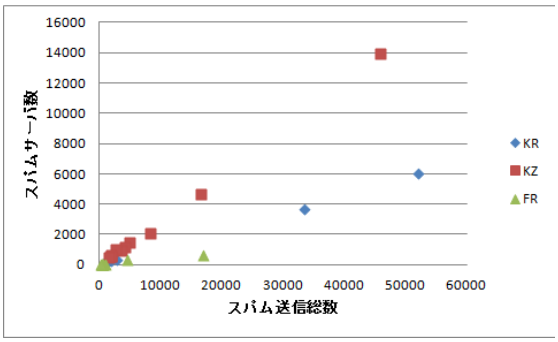

図 3　地域別のスパム特性 (韓国, サウジアラビア, フランス)

と比較して狭い事から, メールを大量に送信する地域が複数ある傾向があるといえる．欧州の国は, 横軸への伸びは少ないが, 観測範囲同士の距離が 3 つの地域の中で最も狭いことから, 観測範囲数が多く, スパムを発信する地域が分散する傾向があるといえる．中東の国は観測範囲が右上に突出する国が多いことから, 一部の地域に大量のスパムサーバがある傾向があるといえる．

図 2 や図 3 では各観測範囲の分布の意味について考察したが, サーバの集中度の別の基準として, 各スパムサーバから距離 r にあるサーバ数を計算することが考えられる．それによって個々のスパムサーバから見たサーバの集中度を測れると考える．

4　まとめ

本稿では国ドメインのスパムサーバの発信行動変容の推定可能性について検討した．その結果, 国特有の特性が存在する事が予想でき, その特性の変化を観察することで, 国別のスパムサーバの行動変容を表す事ができることが示唆された．国別の行動変容は個々のサーバがスパムサーバであるか否かを判断する際に, どの要素に注目するべきかの判断の手助けになると考えられる．今後の課題として, 今回予想できた特性をどの様に解析するかを考える．またもう一つのスパムサーバ行動変容の軸である個々のサーバの行動変容を調査する．

参考文献

[1] 王戦, 堀良彰, 櫻井幸一 「中国語迷惑メールにおけるベイジアンフィルタの適応と評価」2006-CSES-33 (9) pp. 45-50 2006
[2] IIJ メールテクニカルレポート
[3] 竹下峰弘, 中平勝子, 三上喜貴 「スパムメール発信源分析によるサーバ・ドメイン管理実態の推定」 第 73 回全国大会講演論文集 pp. 3-499 - 3-500 2011
[4] 森達哉 「RrBL:スパムメールのための確率的なブラックリストの提案」信学技法 NS2008-146(2009-03) pp. 15-19 2009
[5] 小川健司, 稲葉宏幸 「記号と未知語の分布を用いたベイジアンスパムフィルタの提案」信学技報 SITE2008-79, IA2008-102(2009-03) pp. 209-212 2008
[6] Symantec Intelligence 2012-05

* http://www.maxmind.com/app/ip-location