L-023

スパム送信サーバの重篤度によるスパム送信パターン分類

Classification of spam mail transmission patterns based on their servers' severity levels

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

1 はじめに

本稿では，メール受信サーバにおいてスパムと判定されたメールのメールヘッダを元に収集したメール送信サーバの情報を観測し，その収集データを処理することで，メール送信サーバの行動や変化を明らかにする．前研究 [1] ではスパム送信サーバの行動を，その時間密度パターンによって解析し，スパム重篤度（ED 値）として定義した．ED 値は任意の期間中でのスパム送信サーバの悪質さ（スパム送信の期間，密度）を定量化している．

本稿では，ED 値よりスパム送信サーバの復元力を導出する．本稿におけるスパム送信サーバの復元力とは，スパムを送信する様になったサーバが健全な状態に戻るときにはたらく力である．復元力が高いサーバは，一時的に大量のスパムを送信しても短い期間でスパムを送信しない状態に戻すことができる．管理能力の高いサーバ管理者であれば，短い期間でスパム送信に気づき，それを止めることができると考えられるので，この復元力はサーバ管理者のスキルや，国の体制に依存することになる．したがって，このような能力を復元力として評価することができるならば，復元力の高いサーバをブラックリストフィルタから動的に除外したり，その国のカントリーガバナンスの程度を測ることに応用できると考える．

本稿では，復元力の各要素（最大送信能力，継続性，再発性）を導入し，それらによって，スパム送信サーバのスパム送信パターンがどのように分類されるのかを検討する．そしてこれらの分類結果から，各要素がどの様に復元力に影響しているのかを考察する．

2 Evolution Diagram

本章ではスパムの情報送信量特性を表す Evolution Diagram（以下，ED）について説明する．ED はメール送信を観測する時間間隔を変化させ，各時間間隔ごとのスパム送信頻度を求めることによってスパムの情報送信量特性を表す手法である．一般的に，ED の積分値である総 ED 値が大きいほど，より悪質な（スパム送信期間が広く，定期的に送信している）スパム送信サーバだと考えられる．ここでは山口ら [1] にあるアルゴリズムを簡潔にまとめる．まずスパム送信を観測する期間 T を任意の数 L の区間に等分する．その後，そのサーバが任意の区間中にスパム送信を行ったか否かを判定する．ここで任意の区間を i $(i = 1, 2, \cdots, L)$ とすると，

$$\begin{cases} a_i = 1 & \cdots \text{ スパムが区間 } i \text{ 中に存在している} \\ a_i = 0 & \cdots \text{ スパムが区間 } i \text{ 中に存在していない} \end{cases} \quad (1)$$

となる．a_i は時系列順に連続しているためベクトルとして表す．

$$\boldsymbol{a} = (a_1, a_2, \cdots, a_i, \cdots, a_L)$$

ここで，区間幅 t は，

$$t = \lfloor \frac{T}{L} \rfloor$$

となる．

ED は $\boldsymbol{a}$ より得られるスパム発生区間の頻度より求められる．まず，$\boldsymbol{a}$ は分割区間数 L を変化させることで，区間幅 t とスパム発生区間数が変化する．そして，$\boldsymbol{a}$ における分割区間数 L と，スパム発生区間数の比率，スパム発生頻度 $I(t)$ は，以下の式で与えられる．

$$I(t) = \frac{\sum_{i=1}^{L} a_i}{L}$$

この $I(t)$ を時間系列順に沿って並べることで，スパム送信頻度を表すベクトル $\boldsymbol{i}$ が得られる．

$$\boldsymbol{i} = (i_1, i_2, \cdots, i_L)$$

ED の一つの基準である総 ED 値は ED の積分値であり，分割区間数が L のとき，

$$\mathrm{ED}(L) = \sum_{n=1}^{L} i_n$$

と定義される．実際に計算する際の分割区間数を，N を 0 以上の整数として，

$$\boldsymbol{L} = (2^0, 2^1, \cdots, 2^N)$$

とすると，ED 値は

$$\mathrm{ED}(N) = \sum_{n=1}^{N} \frac{\sum_{m=1}^{L_n} a_m}{L_n}$$

で求めることができる．

下記の表 1 は，任意の条件でスパムを送信をシミュレートし，$T = 1$ 年 $(N = 24)$ の間隔で測定した ED 値を表したものである．表の列は任意の月に送信されるスパム数を表し，スパムはその月の中で等間隔に送信されるとして計算した．例えば 5 通であるなら，その月の中で等間隔になるように約 6 日置きに送信される．表の行はスパムが送信される月のパタンを表す．

† 長岡技術科学大学

これは等間隔と連続に分けられる．等間隔の項目は任意の間隔でスパム送信が行われる月が出現し，連続の項目ではスパムを送信する月が任意の期間連続で現れる．カッコ内は 1 年を通した総メール送信数を表している．表からわかるように，ED 値は必ずしも総メール数と比例しない．値は同じメール数であれば，より広範囲にスパムを送信するサーバをより危険だと判定している．

実際のスパム送信サーバはどの様な ED 値を取るのかを実測した結果を図 1 に示す．観測データは，筆者の大学のメールフィルタリングソフト（spam assasin）によってスパムと認定されたものを収集した．観測期間は 2013 年 3 月 1 日から 2014 年 2 月 28 日までであり，全体で 21,332,168 通のスパムと 1,733,929 ドメインを確認した．本研究では 1,733,929 ドメインの中からランダムに 10,000 ドメインを抽出し，これをスパム送信サーバの代表例として扱った．図 1 に，標本として抽出した 10000 ドメインにおける ED 値分布を示す．これが母集団と同等の振る舞いをすると考える．図からわかるように 5,703 ドメインは ED(1 年相当の N) ≈ 2 である．この $ED \fallingdotseq 2$ 場合のほとんどが 1 年を通してスパムが 1 通だったサーバの場合である．この場合，式 1 において $\boldsymbol{a}$ のスパムの存在する要素が 1 要素しかないため $\sum_{n=0}^{N} \frac{1}{2^n}$ となり 2 に近似する．そして ED 値が 2.5 以上のドメインは 2,109 ドメインであり，全体の 21% 程度を占める．表 1 から見えるように，1 ヶ月間毎日 1 通ずつスパムを送信するようなサーバ，あるいは 2 ヵ月連続でも月に 1 通しかスパムを送信しないサーバは $ED < 2.5$ である．そして 2 ヶ月以上，複数のスパムを送信する様になって $ED \geq 2.5$ を超えてくることがわかる．このことから $ED \geq 2.5$ はスパム送信の悪質さを示すうえでの ED 値の 1 つの区切りの基準であると考える．

表 1　$T = 1$ の場合の ED 値 (全スパム送信数)

メール数/1 ヵ月		1 通	5 通	10 通	30 通
1 ヵ月		2.00(1)	2.25(5)	2.34(10)	2.48(30)
連続	2 ヵ月	2.13(2)	2.99(10)	3,17(20)	3.45(60)
	3 ヵ月	2.63(3)	3.41(15)	3.70(30)	4.10(60)
	4 ヵ月	2.88(4)	3.72(20)	4.06(40)	5.57(120)
等間隔	6 ヵ月	3.00(2)	3.99(10)	4.16(20)	4.42(60)
	4 ヵ月	3.50(3)	4.37(15)	4.64(30)	5.06(90)
	3 ヵ月	4.00(4)	4.99(20)	5.57(40)	6.34(120)

3　復元力と ED 遷移図

本稿で導入する復元力とは，スパム送信サーバがスパムを送信している状態から，健全な状態に戻ることのできる力を示している．例えば，一時的に大量のスパムを送信してしまっても短期間で健全な状態に戻すことができるなら復元力は強いと考え，少数のスパムで長い期間送信するサーバを復元力が弱いと考える．1 年（N=24）での ED 値は長期的な期間でのスパム送信サーバの悪質さを調査できたが，表 1 が表すように，1 通ずつ 6 か月ごと来た場合（場合 1）と 2 ヵ月連続 5 通ずつ来た場合（場合 2）の様な違った状況で，その ED 値に大きな差がない場合がある．本稿では短期の ED 値の遷移を表す ED 遷

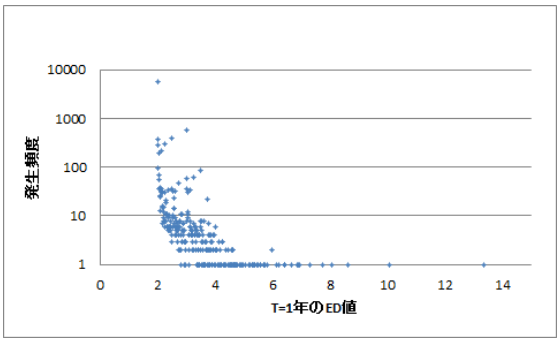

図 1　ED 値発生頻度

移図を定義した．ED 遷移図とは一定区間ごとに ED 値を計測し，その増減を遷移図として表したものである．y 座標は現在の ED 値を表し，x 座標は 1 つ前の区間の ED 値を表す．

$$\begin{cases} x_0 = 0 \\ x_n = y_{n-1} \end{cases} \tag{2}$$

図 3 は場合 1 と場合 2 の状況を ED 遷移図で表したものである．この例では 1 区間を 1 週間 ($N = 20$) としている．場合 1 は 6 ヵ月に 1 度 1 通だけスパムが来ているためスパムの来た区間は $ED \approx 2$ となり，そしてその両隣の区間は $ED = 0$ のため，座標は $(2,0) \rightarrow (0,2)$ と進み三角の図形ができる．場合 2 は毎週 1 通ずつスパムがくるため，座標が $(2,0) \rightarrow (2,2) \rightarrow (2,2) \rightarrow \cdots$ と進み四角い図形ができる．2 つの例からわかるように同じ 1 年の ED 値であれ，違った図形の形や大きさなどを表す．

ED 遷移図から復元力の 3 要素，スパム送信サーバの最大送信能力，継続性，再起性を定義する．最大送信能力は ED 遷移図に現れる図形の面積から，継続性は各要素の線積分から，再起性はスパム送信の再開回数から求める．

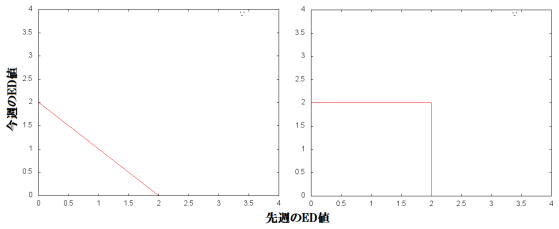

図 2　左：場合 1，右：場合 2

本稿では，復元力における最大送信能力を ED 遷移図の面積で定義する．面積の導出手順を以下に示す．

1. グラフを画像として任意の粒度でピクセル化する．
2. 画像の一辺の大きさは観測された最大の ED 値より高く設定する．
3. 右上の座標からベクトルを境界にラベリング処理を行い，図形に含まれていない面積を導出する．
4. 画像全体の面積から，導出した面積を引くことで，ED 遷移図に現れた図形の面積を導出する．

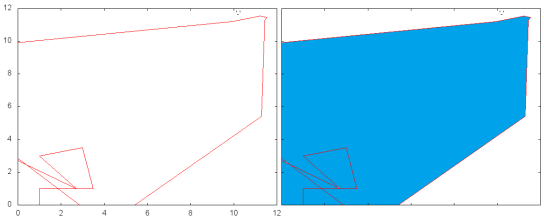

図 3　左:元データ 右:面積導出箇所

ED 遷移図を作成した期間の中で最も活発にスパム送信が行われた時期の ED 値の移り変わりが面積となって現れ，そのサーバの最も悪質だった時期のスパム送信の量，頻度によって面積が大きくなる．

遷移図の距離，ED 遷移図距離とは，グラフの連続する要素のユークリッド距離である．N を週数とすると，

$$d = \sum_{n=1}^{N} \sqrt{(y_n - y_{n-1})^2 + (x_n - x_{n-1})^2} \qquad (3)$$

連続する要素間の距離が大きいということは，長期間スパム送信を行っているということである．図 3 は実際のデータから ED 遷移図を作成し，その距離と面積を計測した際の 2 つの要素の関係図である．面積，距離共に 1 年で 1 通のみスパムが来た場合のデータを 1 として正規化してある．図 3 からわかるように面積と距離は比例していない．これは ED 遷移図の面積が最外殻のみから求められていることに対し，距離はすべてのベクトルを考慮しているからである．そしてこれらは，面積が大きく距離が短いクラスと面積が小さく距離が大きいクラスに分けられることができる．

一般的に面積が大きく距離が短いクラスには短期間 (1 ～ 3 週間) に大量のスパムを送信し，その後は健全な状態にサーバが移動した場合が多い．これは表 1 における 1 ヵ月に 30 通スパムを送信した場合などに当てはまる．1 区間での ED 値は大きく，それが連続するため面積の大きさは広くなるが，その後が続かないため 1 年全体で見た場合は局所的にスパムが来ているように見え，1 年での ED 値は比較的小さくなる．

次に面積が小さく距離が長いクラスは，長期的にスパムが送信されるが一つの区間で送信されるスパム量は少ない場合が多い．表 1 における 3 ヵ月ごとに 1 通スパムがくる場合などに当てはまる．1 区間でみるとスパム量が少なく ED 値も小さくなるが，1 年全体でみるとスパムがまばらに表れているため 1 年の ED 値は大きな値となる場合が多い．前者のクラスは短期的にスパムを大量に送信するが健全な状態にすぐに戻り再発しないという点から復元力が強いといえ，後者のクラスはスパム送信自体は少ないが，長期にわたりスパム送信を止めることができないことから復元力は弱いといえる．

4　スパム送信パターンと再起性

最大送信能力と継続性はスパム遷移図の形によって導き出されるものであった．スパム送信パターンとは ED 遷移図に現れる値の増減を記号でパターン化したものである．記号は以下のルールによって定められる．

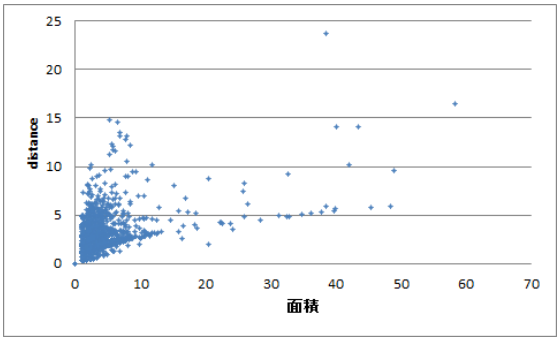

図 4　ED 遷移図における面積，距離の関係性

y_n は n 週目の ED 値

S: $y_{n-1} = 0,\ y_n = a$ (a は正数)

G: $y_{n-1} = a,\ y_n = 0$

_: $y_{n-1} = a,\ y_n = a$

E: $y_{n-1} = 0,\ y_n = 0$ (E が連続した場合 1 つにまとめる)

U: $y_{n-1} = a_1,\ y_n = a_2$ ($a_1 < a_2$)

D: $y_{n-1} = a_2,\ y_n = a_1$

この記法に従えば，再起性は S の個数で定義できる．表 4 はパターンの発生頻度を表したものである．本章では頻度が多い順に 5 つ紹介している．最も多いパターン ESGE は 1 年のうち 1 週間のみスパム送信をしたパターンを表している．またこのパターンの頻度は 4 に現れた ED ≈ 2 とも近似し，送信されるスパムの半数以上が 1 週間以内にスパム送信をやめることがわかる．次に頻度が多いパターンは ESGESGE である．このパターンは 1 週間スパム送信を行い，その後 1 週間以上を開けてスパム送信を行う場合である．パターン ESDGE，ESUGE は 2 週連続でスパム送信を行った場合である．間を開けてスパム送信を行うサーバより，2 週間続けてスパム送信を行うサーバ数が少ないことは，スパム送信者は連続してスパムを送信するより間を開けてスパムを送信するほうが効果的だと思っている．もしくはサーバ管理者が最初の 1 週間でスパム送信を発見し止めたが，後日またスパム送信を許してしまったと考えられる．どちらの場合であれ，パターン ESGESG は復元力が低いと言える．10,000 ドメインのサンプルの中で ESDGE と ESUGE がほとんど同じ頻度を示したことはスパム送信量の少なさに原因があると考える．2 週間の間に送信されたスパムの量が 2 ～ 3 通のため，区間の区切り方により ED 値が大きく変動し，結果としてスパムの短期間での増減の傾向が測れなかったと考える．

表 2　スパム送信パターン出現頻度

パターン	頻度
ESGE	5,904
ESGESGE	1,222
ESDGE	257
ESUGE	245
ESGESGESGE	202

この再起性を実測した結果，約 25% のスパムが 1 週間以上開けてスパム送信をすることがわかった．また，S の発生回数が 3 度以上だったサーバは全体の 7% だった．再起回数が多いということは，健全な状態から悪性な状態に何度も変化するということであり，復元力の強弱に最も影響のある要素といえる．

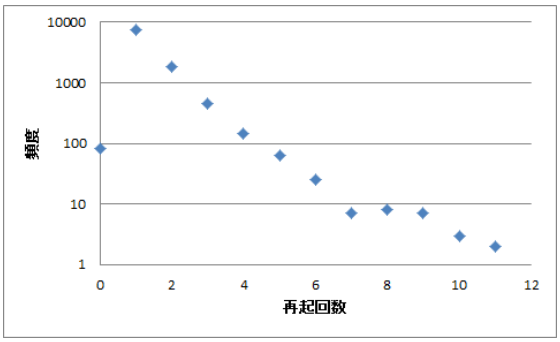

図 5　再起回数頻度

5　おわりに

復元力の 3 つの要素，最大送信能力，継続性，再発性を定義した．これらの要素を元にスパム送信サーバの復元力を考える．実際のスパム送信サーバからこれら要素を求め，各要素ごとに上位 75% の値を求めたものが表 5 である．

表 3　復元力各要素の悪性基準値

最大送信能力	継続性	再発性
1.65	1.99	2

いずれの要素も「1 年で 1 通のみスパムを送信したサーバ」の値が 1 となるように正規化されている．本稿では，この 75% を超えた値が，その要素における悪性と判断する．表 5 は基準を超えた要素の組み合わせのサーバの割合を示している．

表 4　復元力の各要素による分類（チェック項目は悪性）

クラス No	最大送信能力	継続性	再発性	割合
	P	C	R	
0				57.56%
1(R)			√	2.84%
2(C)		√		1.08%
3(CR)		√	√	12.74%
4(P)	√			12.47%
5(PR)	√		√	1.97%
6(PC)	√	√		2.37%
7(PCR)	√	√	√	8.77%

表からわかるように，No.0 の様な全要素健全である場合が最も多い．次に No.3(CR)，No.4(P)，No.7(PCR) の割合が高い．No.3(CR) は少数のスパムを長期にわたって送信するサーバであり，No.4(P) は短期的にスパムを大量に送るサーバ，No.7(PCR) は大量のスパムを送信し，なおかつ長期的な送信を続けているサーバだといえる．復元力の各要素はそれぞれ別の意味を持つ．復元力の弱さを直接示しているのが再発性 (R) である．スパム送信を一旦停止してから，一定区間以上の間隔をあけてスパムを送信を再開するサーバは，特に健全な状態への復元力が弱いといえる．次に継続性も復元力の弱さを示す．スパム送信における ED 値の変化の多さ，つまりスパム送信の長期短期を測る要素である．継続性は再発の多いサーバ，ED 値の変化が多い（スパム送信を大量に行う）サーバに高い値を示すため，この値が高いほど復元力は弱くなる．最大送信能力は継続性，再発性と違い，復元力の弱さではなく強さを示す場合に役立つ．一時的にスパムを大量送信しようと，その後管理者がしっかりと防いでいたならば継続性と再発性は低く抑えられるだろう．最大送信能力は他の 2 要素と組み合わせて意味を持つ要素と言える．

これらのことから表のクラスの中でも，単純にチェックの多さではなく組み合わせによって復元力の強弱を決める．そして特に復元力が特に低いと言えるのは，No.7(PCR)，No.5(PR)，No.3(CR)，No.1(R) だと考える．逆に復元力が高いと言えるクラスは No.0, No.4(P) だと考える．No.7(PCR), No.5(PR), No.3(CR)，No.1(R) は継続性 (C)，再発性 (R) が特に高く，スパム送信を長期的に行い，何度も再発するサーバである．これらのサーバには自力でスパム送信を止める力がないことを示している．逆に No 0，No.4(P) は一時的にスパムを送信するが，再発することなく (∼ R) 管理者がしっかりと管理していることを示していることから復元力が高いと判断できる．全体的に見ると 25% が復元力が弱いと言える．

本手法の利点として，$T = 1$ 年 の ED 値の様に 1 年観測を続けなくても随時各要素が計算でき，短期間のうちにそのサーバを評価できる点にある．これによって短期間でスパム送信サーバの評価を動的に行うことができ，より正確なスパムフィルタリングに近づいたと考える．今後の課題として (1) 復元力のより明確な指標化, (2) 最短観測期間を確定するための N 値サーチを目指す．

謝辞：本研究の一部は学術研究助成基金助成金 24500308 の助成を受けたものである

参考文献

[1] 山口 翔生, 中平 勝子, 北島 宗雄: メール送信サーバ情報送信量特性 , FIT 2013 L-025, 2013

[2] Joshua Goodman, Gordon V. Cormack, David Heckerman; Spam and the Ongoing Battle for the Inbox, COMMUNICATIONS OF THE ACM, Vol.50, No.2, 25, 2007.

[3] Zhengchuan Xu, Qing Hu, Chenghong Zhang: Why computer talents become computer hackers, COMMUNICATIONS OF THE ACM, vol.56, no.4, 64, 2013.

[4] 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, 一般社団法人情報処理学会 全国大会講演論文集, 2011(1), 499-501, 2011.

[5] 澤谷 雪子: メッセージ本文受信前でのスパムメール探知方式の制度向上に関する一検討, 信学技報 IEICE, Technical Report, ICSS2009-57, 19-24, 2009.