L-007

スパムメール発信源分析による TLD のガバナンス推定
Estimation of TLD governance in analyzing of spam mail source

竹下 峰弘† 中平 勝子† 三上 喜貴†
Minehiro Takeshita Katsuko T. Nakahira Yoshiki Mikami

1. はじめに

営利目的のメールを無差別に大量配信するスパムメールが社会的問題となっており，メール全体に対して約 80%ものメールがスパムと判定されている[1].

スパムメールを防御するために多くの技術が開発されている．主な発信元詐称防止技術としては，受信側で IP アドレスから判定し，スパムメールをブロックするブラックリスト，送信時に認証するなど送信者側から認証情報を送ることによって受信者側のフィルタリングを助ける技術として，OP25B(Outbound Port 25 Blocking)[2]，DNSSEC(Domain Name System Security Extensions) [3]，DKIM(DomainKeys Identified Mail) [4]などがある．

最近は，これら発信元詐称防止技術の導入が進み，Yahoo!は 2010 年 11 月 16 日に，google は 2011 年 1 月 7 日に DKIM を導入したと発表された[5]．また，DNSSEC が 2011 年 1 月 10 日に「.net」「.com」に導入される[3]など，スパム対策のために各プロバイダが発信元詐称防止技術の導入が進められている．

本稿では，最近におけるスパムメールの発信量の推移，及び発信元詐称防止技術導入による動向，スパムメール発信量の変化を調査した．また，トップレベルドメイン(TLD)別にみたスパムメールの比率などが各 TLD の管理状態をあらわすものと考え，これを「TLD のガバナンス」ととらえて推定した．

2. 使用データと分析手法

2.1 使用データ

本稿では，スパムメール発信源分析をするため，本学教員のメールアカウントで受信したスパムメールのうち，2010 年 12 月～1 月，5 月に送られたスパムメール約 150 万通をもとに，以下の 5 つの項目を利用して分析を行った．

A) 送信メールサーバの TLD 名
（TLD_{ms}．不明な場合は server unknown とする）
B) TLD_{ms} の IP アドレス（IP_{ad}）
C) メールヘッダの From:行に記載された TLD 名（TLD_{ad}）

表 1　各項目とスパムメール技術の関係

詐称困難な項目	スパムメール対策技術
TLD_{ms}	OP25B，DNSSEC
IP_{ad}	DNSSEC
TLD_{ad}	SPF 認証，DKIM
TLD_g	DNSSEC

D) IP_{ad} から GeoIP[6]を用いて割り出したサーバ所在国の TLD 名（TLD_g．IP が詐称されていた場合には GeoIP unknown とする）
E) メールの到着日付

2.2 分析手法

発信元詐称防止技術として，近年利用が広がりつつある OP25B, DNSSEC, DKIM, SPF 認証の 4 つの詐称防止技術をとりあげた．それぞれの詐称防止技術が利用されたことにより，メール発信を示す TLD_{ms}，IP_{ad}，TLD_{ad}，TLD_g の 4 つ項目のうち，どの項目について詐称を困難にするかを表 1 に示す．

OP25B はメール送信サーバにおいてスパムメール送信を防止していることから TLD_{ms} に，SPF 認証[7]は差出人のメールアドレスにする認定を行っていることから TLD_{ad} に，DNSSEC は送信時の DNS 応答からスパムかどうかを判定していることから TLD_{ms} と TLD_g に，DKIM はメール送信時に電子署名を要していることから TLD_{ad} にそれぞれかかるものと推定している．

これらの情報と合わせ TLD_{ms}，IP_{ad}，TLD_{ad}，TLD_g の定量／定性変化から TLD におけるガバナンス推定を行った．

3 分析結果

3.1 スパムメール発信量の推移

スパムメール変化量の月次変化を見るため，ここでは 2010 年 12 月と 2011 年 5 月の集計データについて分析する．その結果を表 2 に示す．表 2 は 2 つの要素から構成されており，上段 4 データは TLD_{ms}，IP_{ad}，TLD_{ad}，TLD_g 間の関係を，下段 2 データはメールサーバおよびサーバ設置国が不明なものを示している．表から，設置国や IP がはっきりしているサーバからのスパムメール送信量は激減しているが，対してこれらが不明なサーバからのスパムメール送信量は絶対数は減少しているもののスパムメール総数に対する割合は純増している．

表 2　スパムメール発信量比較

項目	2010 年 12 月	2011 年 5 月
TLD_{ms}＝TLD_{ad}	153,695(31.0%)	3,224(1.59%)
TLD_{ms}＝TLD_g	93,346(19.0%)	21,137(10.4%)
TLD_{ad}＝TLD_g	83,995(17.0%)	727(0.36%)
TLD_{ms}＝TLD_{ad}＝TLD_g	55,321(11.0%)	248(0.12%)
Server unknown	270,753(55.8%)	161,023(79.6%)
GeoIP unknown	64,085(13.2%)	57,147(28.2%)

†長岡技術科学大学　Nagaoka University of Technology

3.2 発信元詐称防止技術導入による TLD 別の変化

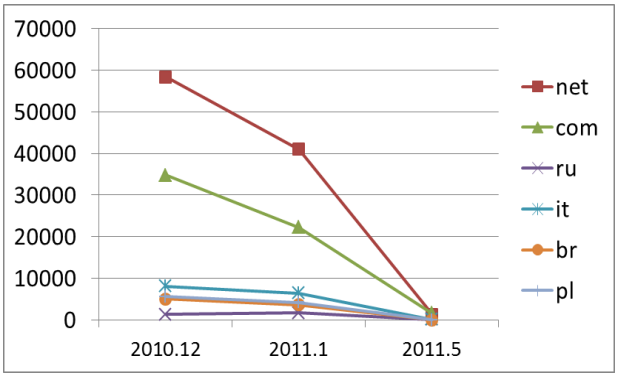

図 1　TLD_{ms}＝TLD_{ad}の一致状況推移

次に，図 1 に各 TLD 別のスパムメール発信量の変化を示す．ここでは，スパムメール発信量の上位 6 つの ccTLD, gTLD を対象とした．図 1 より，「.com」「.net」の割合が非常に高く，2010 年 12 月では，スパムメール総数の約 2 割を占めている．2010 年 12 月から 2011 年 1 月にかけての減少は，表 1 において，DNSSEC や DKIM が導入されたことによる効果であると推測できる．また，2011 年 1 月から 5 月にかけて「.com」「.net」からのスパムメール発信量は約 95%減少した．ボットネット閉鎖によるスパムメール総数の約 59%減少に比べ，減少率が多いことから，特に図 1 の gTLD／ccTLD からは，ボットネットからスパムメールが送信されたものであると推測できる．

3.3 IP アドレスから見たスパムメール発信量推移

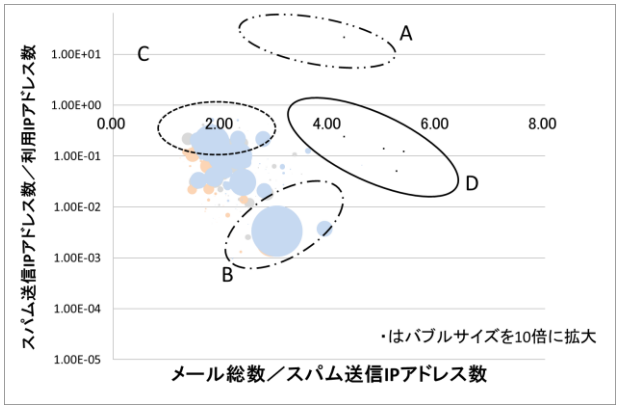

図 2　IP アドレスから見たスパムメール発信推移

メール送信情報のうち IP アドレスはもっとも詐称が難しいデータ項目と言えるが，逆にこれを詐称している TLD は IP アドレス管理が行き届いていないことを疑わせる．そこで，スパムメール発信に使われる IP アドレスを，実際に利用されている IP アドレスの数[9]で除してスパムメール発信ドメイン比率を求めた．これは本来 1 以下になるはずであるが，1 以上になった場合は，そのドメインには「IP アドレスを詐称して，スパムメールを発信しているものが多数存在することを意味する．図 2 において，この比率を縦軸に表す．横軸は ccTLD ごとのスパムメール総数を ccTLD ごとのスパム送信 IP アドレスの数で除したもので，1IP アドレスから送信されるスパムメール数の大小を示す．円の大きさは各 ccTLD のスパムメール総数を示している．

領域 A はスパムメール発信比率が 1 を超えたものを示す．ここには lc(セントルシア)が該当した．

領域 B は各国に割り当てられた IP アドレスからのスパムメール発信量は少ないが 1 つの IP アドレスからのスパムメール発信量が多い．特に us(アメリカ)が大きな割合を占めている．これは，特定の IP アドレスから大量のスパムメールが送信されていることを示している．

領域 C はスパムメールが大量に送信し，かつ IP アドレスの多くをスパムメール送信に使用していることを示す．

領域 D は，少ない IP アドレスで多くのスパムを送信していることから，IP アドレスのほとんどをスパム送信に使用されていることを示す．領域 C，D ともに sr(スリナム)や gu(グアム)などの島嶼国や sn(セネガル)や do(ドミニカ)など途上国がほとんどを占めている．スパムメール発信量は少ないものの，IP アドレスのほとんどがスパムメール発信に使われており，2010 年 12 月から 2011 年 5 月の間では大きな変化は確認出来ない．このことから，島嶼国の ccTLD が悪用され続けている現状を表している．

4.TLD におけるガバナンス推定

以上の分析から，TLD のガバナンスについて以下のように傾向を観察することができる．大規模な gTLD に対する発信元詐称防止技術の導入により，スパムメールは若干の減少が見られる．IP アドレスから見たスパムメール送信量推移は．IP あたりのスパムメール送信量及びスパムメール発信ドメイン比率から 4 分類される．島嶼国や途上国の ccTLD のほとんどではスパムメール発信状況に改善が見られず，十分な管理がなされていないと推定できる．

5.まとめ

本稿では，各 TLD の一致状況および IP アドレスから見たスパムメール発信推移から TLD におけるガバナンス推定を行った．今後も継続的に分析を行い，ガバナンスレベル評価に向けて各項目の詳細な分析を行っていく手法を考案する．

参考文献

[1]ｼﾏﾝﾃｯｸ ｾｷｭﾘﾃｨﾚｽﾎﾟﾝｽﾎﾜｲﾄﾍﾟｰﾊﾟｰ2011 年 4 月， 2011.5.28 http://www.symantec.com/ja/jp/business/security_response/whitepapers.jsp
[2]日本データ通信協会, 2011.4.29 http://www.dekyo.or.jp/soudan/taisaku/i1-3.html
[3]JPNIC, 2007.2.15; http://www.nic.ad.jp/ja/basics/terms/DNS-cp.html
[4]RFC5617, 2009.8; http://tools.ietf.org/html/rfc5617
[5]@IT; http://www.itmedia.co.jp/enterprise/articles/1101/07/news019.html
[6]GeOIP; http://www.maxmind.com/app/ip-location
[7]RFC4871, 2007.5; http://www.ietf.org/rfc/rfc4871.txt
[8]JPCERTCC,2006.7; http://www.jpcert.or.jp/research/2006/Botnet_summary_0720.pdf
[9]迷惑メール対策, 2011.6.25 ; http://www.quia.jp/spam/CountryList_ja.html