2Y-8

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

竹下 峰弘　　中平 勝子　　三上 喜貴
長岡技術科学大学

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 87%ものメールがスパムと判定されている[1]．ここ最近，スパムは減少傾向であるといわれているが，体感的には多いのが実情である．

スパムメールはサーバ負荷の増大や通信量の増加を誘発する一因となりうる．これを上嶋ら[2]が提唱したデジタルデバイドの 3 階層と照らし合わせると，次のように考えられる．デジタルデバイドの 3 階層とは，情報通信技術の利用を前提となる投資や政策の充実を「基盤」，情報通信技術の利用度合を示す「利用」，情報通信技術の利用によって生み出される経済的な価値や社会・文化的な価値からなる「効用」がある．ドメイン名は，IP アドレス，各種プロトコルなどと並んでインターネットを構成する重要資源の一つであり，ccTLD は世界中のすべての国・地域に平等に割り当てられた資源である．基盤レイヤーの運用の良しあしが利用したレイヤーの実態に左右されることを意味する．この観点から，安全・安心かつ地球規模でユビキタスなネットワークの実現のためには，すべての ccTLD 管理者が一定の水準を満足するカントリードメイン管理（CDG[3]）を実現する必要がある．CDG では実態を多面的に評価する指標として CDG Index がある．これには 2 つのレイヤー（リソース，コンテンツ）と 3 つの基準（アクセス，言語多様性，安全・信頼）を評価の枠組みとしている．

本稿では，スパムメールに着目し，発信源分析をすることで各国のサーバ・ドメインの管理実態を推定する．CDG Index の中ではリソースレイヤと安全・信頼基準の中にある，「セキュリティ，安全性，回復性（SSR）」に属す．

なお，本稿において詐称とは GeOIPTLD とメールアドレス TLD が一致しないことを定義する．

Assumption of server and domain governance by Analyzing source of SPAM mail
TAKESHITA Minehiro, NAKAHIRA T. Katsuko, MIKAMI Yoshiki
Nagaoka University of Technology

2. スパムメール発信源分析

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 月に送られたスパムメール約 38 万通を元に，以下の 5 つの項目を利用して分析を行った．なお，スパムメールの判定には SpamAssassin を使用している．

A) 送信メールサーバの TLD 名（以下，TLD_{ms}．不明な場合は unknown とする）
B) 送信メールサーバの IP アドレス（以下，IP_{ad}）
C) 送信アドレスに記載された TLD 名（以下，TLD_{ad}）
D) IP_{ad} から GeoIP を用いて割り出したサーバ所在国の TLD 名（以下，TLD_g．IP が詐称されていた場合には IP_{un} とする）
E) メールの到着日付

なお，GeoIP[4]とは MaxMind 社が提供している IP アドレスから地理情報を逆引きするサービスである．

2.1 各 TLD の一致状況

スパムメールにおける詐称には様々なレベルある．これを表 1 にまとめた．この表に従い，分析する．

① TLD_g と TLD_{ms} の一致状況

分析結果を表 2 に示す．対象データについて，TLD_{ms} と TLD_g との一致関係をまとめたのがこの表である．一致状況は日によって変動するが，全体の 3 分の 1 から半分は IP が詐称されていることがわかった．また，国によって GeoIP と一致するもの，gTLD，不明なものとそれぞれの割合が大きく異なっていた．そこで，TLD_{ms} が GeoIP と一致する割合が多い国をグループ A，gTLD の割合が高いものをグループ B，TLD_{ms} が不明なものが多い国をグループ C とし，さらに詳しく分析を行った．なお，us についてはスパム発信件数が他の国よりも高いため，どのグループにも属させず，単独の扱いとした．

表 1　TLD_{ms} と TLD_g から見る詐称レベル

TLD_{ms}＝TLD_g	公然とスパムメール発信
TLD_{ms}≠TLD_g	どちらかが虚偽の情報で発信
TLD_{ms}＝不明	セキュリティ管理が不十分

表 2　TLD_g に基づく
スパムメール発信源上位 10 カ国

Gr.	TLD_g	発信件数	TLD_{ms} 分布			
			GeoIPと一致	他のccTLD	gTLD	不明
A	it	13362	76.40%	0.07%	0.26%	23.26%
B	fr	16249	22.35%	0.06%	63.42%	14.17%
	ua	16494	5.34%	0.00%	50.24%	44.42%
C	br	31495	32.26%	0.00%	0.23%	67.51%
	ru	30133	23.67%	0.49%	1.72%	74.13%
	ro	14271	3.80%	0.00%	1.42%	94.78%
	in	30491	3.39%	0.00%	0.28%	96.33%
	kr	14651	0.14%	0.00%	0.00%	99.86%
	vn	16005	0.04%	0.00%	0.00%	99.96%
-	us	60908	0.04%	0.09%	82.42%	17.46%

② グループ B の分析

フランス（fr），ウクライナ（ua）では TLD_{ms} が gTLD になっている割合が半分を占めている．フランスの場合はホスティングサービスなどで gTLD が使われているが，gTLD を悪用し，詐称をして使われる可能性も十分に考えられる．ウクライナについては現在調査中であるが，TLD_{ms} が不明の割合が多いので，悪用の可能性を否定できない．

③ グループ C の分析

調査対象のスパムメールの中で TLD_{ms} が不明になって届いているメールは全体の 70%を占めている．不明の場合でも二つのパターンがあり，メールサーバのみが不明で IP_{ad}・TLD_{ad}・TLD_g が示されているもの，到着日付以外のすべての項目が不明なものがある．

TLD_{ms} が unknown である上位 10 カ国について表 3 に示す．到着日付以外がすべて不明なものに関しては，TLD_g を割り出すことが不可能であるため，IPunknown として載せている．

表 3 より TLD_{ms} を unknown と詐称して送る国については主に BRICs を中心とする成長地域であることがわかる．

また，表 2 の分析結果から TLD_{ms} が不明なものが多い国をグループ C としたが，グループ C に属する TLD_g がすべて入っているため，TLD_{ms} が不明である割合が高いことが改めてわかった．

表 3
TLD_{ms} が unknown の場合の TLD_g 割合

TLDg	TLD_g割合	TLDg	TLD_g割合
in(C)	9.38%	IPunknown	3.72%
ru(C)	6.87%	us(-)	3.56%
br(C)	6.65%	sa(-)	2.45%
vn(C)	5.18%	cn(-)	2.33%
kr(C)	4.78%	ua(B)	2.30%
ro(C)	4.34%		

(　)内は表 2 のグループを示す．

3. サーバ・ドメイン管理実態の推定

以上の分析から，サーバ・ドメインの管理実態について，大きく 3 つのパターンに分けられることか推定できる．

グループ A のように GeoIP と一致していることについては，ドメイン管理が把握しやすい環境であると推定される．

グループ B においては，ホスティングサービスの利用など，サーバの管理を他国に委託している場合と，詐称など悪用している場合の 2 つの使われ方が推定できる．

グループ C の国では，サーバ詐称の防止などスパムに対しての対応をしていないことが推定されることから，サーバ管理が不十分になっているもとの推定できる．

4. まとめ

本稿では，スパムメールについて TLD の観点から分析を行い，サーバ・ドメイン管理実態の推定を行った．今後も継続的に分析を行い，月別のスパムの発信の変化など情報量を増やし，管理実態をより適切に推定できる手法を考案する．

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 http://www.symantec.com/ja/jp/business/security_response/whitepapers.jsp

[2] 上嶋智大,中平勝子,三上喜貴:”デジタルデバイドの評価指標についての一提案”, 第 6 回情報科学技術フォーラム, pp.481-484(2007)

[3] Katsuko T. Nakahira, Hiroyuki Namba, Minehiro Takeshita, Shigeaki Kodama and Yoshiki Mikami:”Country Domain Governance: An analysis by Datamining of Country Domains”, The proceeding of Artificial Life and Robotics, in press(2011)

[4] GeOIP; http://www.maxmind.com/app/ip-location