一般口演

一般口演28
機械学習・アルゴリズム・解析モデル

2017年11月23日(木) 12:45 〜 14:15 E会場 (10F 会議室1003)

[4-E-2-OP28-6] 雑誌投稿・学会参加勧誘を目的とした迷惑メールおよびマルウエア添付メールのエンベロープヘッダ特徴解析

渡辺 淳, 仲野 俊成 （関西医科大学 大学情報センター）

【背景・目的】医療・研究機関に向けた営利目的の雑誌・学会勧誘メールおよびマルウエア添付メールが増加している。 SMTPエンベロープヘッダの特徴解析によって、それらの脅威を回避するための方略を検討した。【材料と方法】メール受信中継サーバにおける直近2年間の電子メールの約20%を占める迷惑メール（約200万通）のSMTPエンベロープヘッダをアソシエーション分析とクラスタ分析に供した。【結果と考察】 Predatory雑誌および bogus学会への勧誘メールは全迷惑メールの約20%を占めた。その大部分は南アジアと米国のホスティング業者、および最大手のソフトウェア開発・販売会社のメールサービスから送信されていた。送信者ドメインは直近2年以内に登録されたものが多かった。南アジアから送信されたものでは送信組織を隠蔽する傾向がみられ、米国からのものは南アジアの企業・組織が登録したドメインから送信される傾向があった。ここ数年の predatory雑誌の増加は著しく、送信組織の多くが DNS-BL回避策を講じていると推定されることから、今後、それらの迷惑メール受信を引き金としたトラブルの増加が懸念される。マルウエア添付メールは迷惑メールの1%程度を占め、おもにロシア、ベトナム等の spam bot感染 PCおよび特定のサーバ（欧州に多い）からの送信が多かった。最近、国内から送信されたものが増加しつつあり、詐称送信者アドレスに国内の大学・医療機関のドメインや、研究者、医師が用いている個人アドレスが使われるようになってきた。医療期間に属するスタッフの電子メールアドレスの多くは、論文検索サイト等で公開されている。それらの公開アドレスを送受信に用いることで、受信者の警戒心を解いて送信者の企図した行動をとらせやすくする攻撃が今後も増加すると予測され、それらの脅威に対抗するための方略策定が急がれる。

©一般社団法人 日本医療情報学会

学会参加勧誘を目的とした迷惑メールおよびマルウエア添付メールのエンベロープヘッダ特徴解析

渡辺 淳[*1]、仲野俊成[*1]、
*1 関西医科大学 大学情報センター、

Signature analysis of SMTP envelope headers on spam mails from predatory publishers and bogus conference organizers and those with malicious attachments.

Jun Watanabe[*1] , Toshiaki Nakano [*1]
*1 University Information Center, Kansai Medical University

Spam messages from predatory publishers, and those with malicious attachments are increasing in medical institutions. We thus examined strategies to avoid those threats by characterizing their SMTP envelope headers using the association and cluster analyses. Messages from predatory journals and bogus conference consists of about 20% of all spam messages. Most of them were sent from hosting sites in South Asia or the United States (US), and well-known international e-mail services. The spam messages from the US tended to be transmitted using their envelope sender domains registered by South Asian organizations, and those sent from South Asia tended to conceal their owner information to avoid filtering with DNS-blacklists. Thus, there is concern about an increase in troubles triggered by receiving those spam mails. Malware attached mail accounted for a few percent of all spam messages. They were transmitted primarily from spam bot-infected PCs in Russia, Vietnam, China and East Europe. Recently, personal or official addresses of researchers and doctors become used as spoofed sender addresses. The r-mail addresses of staff in medical institutions are exposed to the public via paper search sites. By using these exposed addresses of researchers, it will make an attack easier. It is thus urgent to formulate strategies against the threats.

Keywords: Spam messages, data mining, malware-attachment, predatory publishers, bogus conferences

1. はじめに

1.1 背景

医療機関・研究機関向けに送信される情報漏洩等を企図したマルウエア添付メールやドライブ・バイ・ダウンロード(Drive-by download) によるマルウエア感染を企図したメールが増加しつつある [1-3] 。加えて、「捕食出版社 (predatory publishers)」からの投稿料奪取を主目的とした投稿依頼メールや、営利目的の学会・偽学会(bogus/fake conference)への参加費奪取・詐取を企図した迷惑メール(CFP spam, call for paper spam; Scientific scam とも呼ばれる)も、また、増えつつある [4,5]。さらに、国内外において、ランサムウェアをはじめとするマルウエアを用いた医療機関への攻撃事例も報道されている。

これらの迷惑メールの増加が、医療機関における診療情報システムの可用性、保存姓、引いては真正姓を損なう事故の発生や医療施設間のネットワークを介した情報授受の阻害要因となることが懸念される。そこで、これらの危険性に対する理解、およびそれらを用いた攻撃に対する適切な防御方略の策定が重要と考えられる。

我々は、2007 年 3 月から、我々の所属する医科大学に送信される迷惑メールの SMTP エンベロープヘッダ情報を解析し[6] 、それらの特徴抽出を試みてきた。そして、それらの結果に基づいて作成した阻止ルールを用いてフィルタを構成し[6] 、抽出した特徴の感度・特異度を算出してルール・フィルタの精度を検証してきた [2,7,8]。これらの試みによって、SMTP ヘッダの特徴を用いて迷惑メールの大部分(95%以上)を検出可能なこと、および、SMTP ヘッダの特徴から迷惑メールの種類を相当程度、推定可能なことが明らかとなりつつある。

1.2 目的

本研究では、各種迷惑メールのうち、近年、増加が著しく、検出・阻止が困難になりつつある医療・研究機関に向けた営利目的の雑誌・学会勧誘メール(scientific scam)およびマルウエア添付メール(ドライブ・バイ・ダウンロードを含む)について、SMTP エンベロープヘッダ情報の特徴分析によって特徴を抽出し、抽出された特徴に基づいた阻止ルールを策定することによって迷どの程度排除(受信拒否)可能かを検討することで、それらの脅威を回避するための方略を検討した。

2. 材料と方法

2.1 材料

直近 2 年間に調査対象機関の主メール受信中継サーバ群に送信された電子メール 約 850 万通(正規メールと迷惑メールの総計)の SMTP エンベロープヘッダ情報を用いた。

2.2 方法

電子メール受信中継サーバ群の MTA(実装は Postfix)の設定ファイルに検証・阻止のための迷惑メールフィルタをセットした[6]。

データマイニングはアソシエーション分析を主要な手法とし 、SMTP エンベロープの記載項目のうち送信者アカウント(@ マークの左辺) 、送信者ドメイン(@マークの右辺)、宛先アカウント、送信サーバ (SMTP ではクライアント)のドメイン名(サブドメイン名を含む)、送信サーバの IP アドレス、送信サーバが接続の際に greeting message で名乗る helo/ehlo の各項目の相互関係を多言語対応版 KH coder[9]を用いて調べた。マイニングの結果に基づいて検出ルールを策定し、受信中継サーバのフィルタに実装して感度、特異度、尤度比、事後確率を指標として検出精度を測定した。

送信元の検証にはおもに Hurricane Electric の BGPtool kit1 [10] および各レジストラの Whois 情報を用い、送信元のレピュテーションについては Sender Base/TALOS[11]を用いた。

Predatory journals の同定には 2016 年末まではおもに Beall's List[12]を用い、当初はリストアップされている出版社・雑誌が所有するドメイン名をプローブとして検出を試みた。営利目的の学会・偽学会の同定には囮アカウントによる捕捉およびメーリングリスト等のアドレスに誤送されて公開されているものの情報、およびそれらを監視している団体のブログ等を、必要に応じて参照した。Beall's List の公開が停止 [13]した 2017 年初頭からは、2017 年1月 10 日付けの Beall's List をもとに、それまでに収集した SMTP ヘッダのプロファイル情報を加えて判定を行った。

また、偽陽性がほとんどない DNS-BL（spamhaus.org）[14]および、ホワイトリストによって偽陽性発生の低減措置を講じた DNS-BL (BarracudaCentral)[15]を併用し、本研究で構成したフィルタと DNS-BL での検出結果の差異についても検討した。

解析対象とした電子メールが迷惑メールであるか正規メール（legitimate messages）であるかどうかの判定は、各項目の記載事項を検索して web site や blog で公開されている情報から判定する手法、および囮アカウントを用いて収集したメールの目視確認を併用した。なお、受信後 24 時間以内では迷惑メールか正規メールかを判別できないものが、電子メール1万通あたり 20~50 通の割合で含まれていた。それらについては、その後の解析で約 85%が正規メール、5%が迷惑メール、10%弱が1ヶ月を経ても判別不能という結果が得られたが、本研究での解析に当たっては、それらのすべてを判別不能（受信したものは偽陰性、ドロップしたものは偽陽性）として扱った。また、他機関、ISP、フリーメール等から転送されたメールは、解析の対象から除外した（迷惑メールであることが明確なものは、実際にはフィルタリングしている）。 フィルタでの阻止に際しては、判定不能および精度の検証が十分でないものについては、ログに標識 (reject_warning) を付したうえで、メール HUB の各ユーザのメール BOX 宛に配信した。 なお、マイニングデータから迷惑メールと判別されたものの、その種別を特定できなかったメールが少数存在した。それらは迷惑メールの総数には算入したが、迷惑メールの種別毎の解析には用いなかった。

3. 結果

3.1 迷惑メール全数の推移

調査対象としたサイトでは 2007 年からメール受信中継サーバにおいてメールフィルタリング機能を用いた迷惑メール対策が施されている [6,7] 。本研究で作成したフィルタ群は、この迷惑メールフィルタに組み込んで検証した。フィルタで阻止したメールについては、原則、応答コード 450(再送要求) または 550（受信拒否）とともに阻止事由を送信サーバに返した。

フィルタを通過したメールはメール HUB に送信され、ウイルス・マルウエアのチェック後にユーザのメール BOX に配信される。 2013 年以降、外部からの SMTP 受信接続数は週あたり平均約 10 万コネクションで、それらの 80～85%は正規メール（メールマガジンや企業等のオプトイン・バルクメール等を含む）であった。迷惑メールの総数は約 200 万通（約 2 万通/週、送信されてきた全メールの約 20%）と、10 年前の対策開始当初（迷惑メールの総数約 16 万通／週、送信されてきたメールの約 80%）と比べて減少しているが、調査期間中に漸増する傾向を示した（図1）。この増加は、おもに predatory journals/publishers からの迷惑メールを主とする scientific scam の増加に起因していた （図2）。

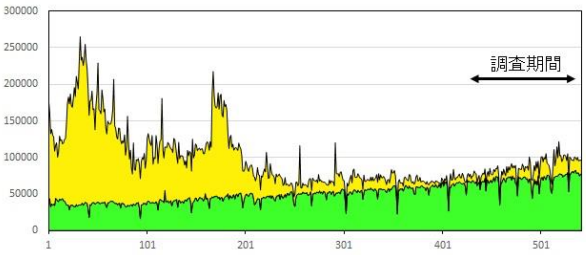

図1 送信された電子メール数の推移

調査開始から約 10 年間の電子メール数の推移。黄は迷惑メール、緑は正規メールの総数。横軸は週。今回の調査期間を図中に示す。

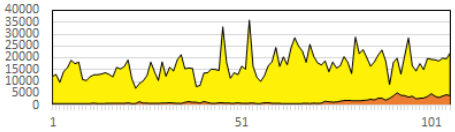

図 2

図2 送信された迷惑メール数の推移

調査期間（直近2年間）の迷惑メール数の推移。橙は scientific scam、黄はその他の迷惑メール、横軸は週。

3.2 マイニングによる特徴抽出と阻止

迷惑メールのエンベロープ情報のうち、マイニングによって共起関係が認められたのは、送信者アドレス（アカウント部、ドメイン部とも）、送信に用いられた SMTP サーバの FQDN または IP アドレス、および送信サーバが名乗る HELO/EHLO 名で、宛先アドレス名との共起関係が検出される場合もあった。共起関係を示す項目は、迷惑メールの種類によって違いが観察された。

以下、predatory journals/publishers からの迷惑メールを「scientific scam（科学詐欺メール）」、ドライブ・バイ・ダウンロード攻撃を含むマルウエア添付メール攻撃を「マルウエア攻撃」と称し、その SMTP ヘッダの特徴、送信特性、阻止状況について述べる。

3.2.1 Scientific scam

Predatory journals/publishers からの迷惑メールでは、送信者アドレスのドメイン部と送信サーバの FQDN/IP アドレスの間に特に強い共起関係が観察された。また、一部のケースでは送信者アドレスのアカウント部にも editor、cfp、journal 等の特徴語が比較的高頻度で検出され、それらと送信者アドレスのドメイン部との間に共起関係が認められた（図3）。

送信に用いられたサーバの大半は DNS に登録された正規サーバであり、その多くはホスティング業者、メール送信サービス業者のものであったが、一部、世界最大のソフトウエア会社のメールサービスを利用しているケースもあった。送信サーバの FQDN には著しい特徴が観察され、predatory journals/publishers および営利目的の学会運営業者は、特定の業者を利用している傾向が明らかとなった。送信に用いられたサーバの FQDN は以下のようであった（一部伏せ字）。セカンドレベルドメインとトップレベルドメインの組みで特定される scientific scam 送信サーバ群の例には:

registerexxxte.com
conferencenxxxfy.net
pcsconfexxxce.net
avenspublixxxrs.com
omicspublishingxxxup.com
oap-jouxxxls.net
sbjsxxxer.com
webhoxxxox.ne
yourjourxxx4u.org
mxxxnd.org
ikkxxxss.org
researchplatxxxm4.com

等が存在した。その他、サードレベルドメインまでの組み合わせで特定される scientific scam 送信サーバ群としては

[sin2, phx3, iad2 のいずれか].securesxxxer.net

の例があり、第 5 レベルドメインの一部に特徴のある scientific scam 送信サーバ群の例としては

[mail-bo1indxxxx/mail-ma1indxxxx].outbound.protection.ouxxxok.com

の例があった。なお、本研究での調査開始前に頻用されていた yaxxo.com からの、scientific scam の送信は、調査開始直後に、ほとんど収束した。なお、 また、上述の送信サーバ群を用いて送信されるメールが、大手フリーメール、ISP の正規サーバ、データセンターのサーバ間を移動してゆく傾向が見られたが、移動先の多くが上述のドメイン間を相互に移動している傾向を認めた。

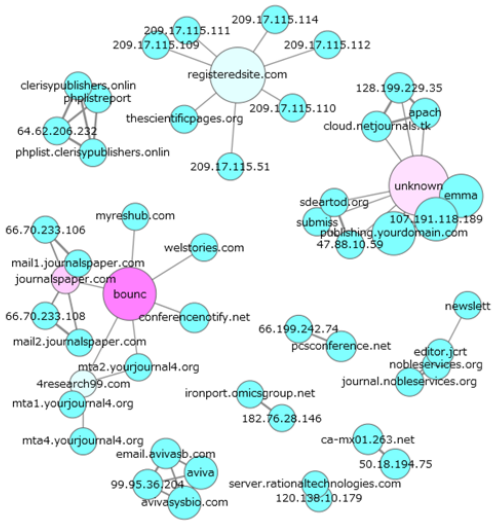

図3 Scientofic scam の共起ネットワーク分析例

SMTP エンベロープヘッダから scientific scam の送信者アドレスドメイン部、送信サーバ FQDN および IP アドレスを抽出して解析した例。

これらの FQDN を持つサーバ群から送信された scientific scam は、全 scientific scam の 70%以上を占めた。また、これらの送信サーバ(サイト)ドメインから調査対象の機関に送信されたメールの 99.8%以上が scientific scam であり、残りも論文の校閲サービスなどに関する迷惑メールであった。

それらの送信者アドレスのドメイン部は、出版社名や雑誌名の一部または略称がセカンドレベルドメインに記載され、トップレベルドメインを net、com、info、org や、最近の新汎用ドメインである xyz 等を変化させて付与されていることが多かった。送信者アドレスのドメイン部は、DNS の送信者ドメイン情報 (sender policy framework SPF など)において、上述の送信サーバ群のいずれかに紐付けて記述されているケースがほとんどで、scientific scam の大半が、RFC に則った正しい SMTP ヘッダを有していることが判明した。

このことから、Whois 情報を用いた送信組織の特定および送信業者や送信者アドレスドメイン部の相互関係を描出することが可能なことが判明した。

送信者ドメインの多くは南アジアの2カ国で登録されており、大半は直近 2 年以内に登録されていた。南アジアの送信業者から送信されたものは Domains by Proxy や Privacy Protect を介して登録することで送信組織を隠蔽する傾向がみられたが、その多くは web ページのデザインや記載されたリンクから、南アジアを本拠地とするオープンアクセス出版社との関係が推定された。米国から送信されたものについても、南アジアの企業・組織が登録したドメインから送信される傾向があった。

調査期間半ばの 2016 年 7 月の時点では、Beall's List で悪質と報告されている predatory 出版社と、web サイトの記述から明らかに営利目的と判明した業者からのメールを中心に、上述の特徴に基づいて構成したフィルタを用いて阻止する方法によって、70%前後を阻止した。その後、Beall's List の公開は停止した [13]が、本研究で抽出した特徴を用い、上述の FQDN フィルタに加えて、送信者アドレスフィルタ、および DNS の逆引きができないサーバを用いて送信されるものに対する IP アドレスフィルタを組み合わせることで、2017 年 4 月以降、ほぼ 95%を超える阻止率を記録している。なお、DNS-BL のみでの阻止率は 20%以下であった。

3.2.2 マルウエア攻撃

マルウエア添付メールおよび悪意のある web サイトに誘導してマルウエアのダウンロードを企図するドライブ・バイ・ダウンロード攻撃メールでは、当初は特定の国の比較的少数のサーバ群から送信される傾向が観察されたが、徐々に送信国が増加する傾向を示した。調査期間の前半では送信元 IP アドレスと helo/ehlo との間に強い共起関係が見られたほか、それらの情報と SMTP エンベロープ情報を元に Whois の情報から得られる送信プロバイダ、サーバ群の IP アドレスおよびドメインの所有組織との間にも共起関係を認めるケースが多かった。その後、送信者アドレスのアカウント部とドメイン部、送信者アカウント部またはドメイン部と送信元サーバが接続に際して名のってくる helo/ehlo との間に共起関係が観察されるようになり、それらと送信元 IP アドレスとの関係が検出されなくなった。これらの特徴と、SenderBase/TALOS[11]を用いた送信 IP アドレス領域接続機器の FQDN の精査や囮で捕捉したメール本文のヘッダ情報などから、当初は snowshoe 攻撃(複数の送信 IP アドレス領域に送信サーバ群を用意し、送信領域を変えながら送信する)[8]と Botnet を用いた攻撃が併用されていたが、次第に Botnet からの攻撃が主体となり、Botnet に移行が進むにつれて、送信国が増加したことが推定された。

Snowshoe 型の攻撃が比較的多かった調査期間の前半には、その送信特性から送信元 IP アドレスと helo/ehlo との間に強い共起関係が見られたほか、それらの情報と SMTP エンベロープ情報を元に Whois の情報から得られる送信プロバイダ、サーバ群の IP アドレスおよびドメインの所有組織との間にも共起関係を認めるケースが多かった。これらは、詐欺メールの一種である 419 scam と類似 [2]していた。Botnet から送信されたマルウエア攻撃メールの SMTP ヘッダにおける送信者ア

ドレスのドメイン部には、フリーメールのアドレスが多用されていた。アカウント部については、特定のアカウントを継続して使用しているケースと、任意の文字列を用いて常に異なったアカウントを用いているケースに大別された。これらのことから、送信組織が複数あることが推定された。

Botnet を介して送信されたマルウエア攻撃メールの最大の特徴は helo/ehlo にあった。2015 年には送信サーバ（実際はほとんどがサブスクライバーラインに接続した感染 PC）の IP アドレスを直書きまたはブランケットで囲った IP アドレスを helo/ehlo に用いたもの（後者は RFC 違反ではないが、正規のメールサーバは FQDN を用いる）が多かった。また、それまでの迷惑メールと同様、受信サーバの IP アドレスを helo/ehlo に用いたものも多く存在し、その一部にマルウエア攻撃が含まれていた。2016 年になると受信サーバの IP アドレスを helo/ehlo に用いたものは激減し、送信サーバの IP アドレスを直書きまたはブランケットで囲った helo/ehlo も大幅に減少した。それらの代わりにユーラシア大陸北部の大国で登録された送信者ドメインと helo/ehlo を持つマルウエア攻撃メールが急増した。当初、それらのメールの多くは当該国の大手フリーメールサービス経由で配信されたが、その後、送信国は東南アジアの某国を経て東アジアの大国に拡がり、現在は東欧、南欧、南米等からも送信されている。東南アジアと東アジアを中心に送信されているものは、送信者アドレスと helo/ehlo に特徴があり、前者は 15 種程度の特定のアカウントに国内大手のフリーメールサービスのドメイン部を詐称して付しており、helo/ehlo にはユーラシア大陸北部の某大国で登録された特定の FQDN（サードレベルドメインは任意の文字列、セカンドレベルとトップレベルドメインには詐称された特定のドメイン名が用いられている（pxxem.net、sacxxil.com、jxxii.com、ixxox.com 等 x の一部は伏せ字）。また、あるものは、送信者アドレスのドメイン部に、その多くが ad で始まるセカンドレベルドメインを付し、その helo/ehlo は outlook.com の正規サーバを模倣しているが、正規サーバの helo/ehlo がサーバの FQDN を名乗るのに対して、サーバ FQDN ではなく DNS に登録されている MX レコードが使われている。これらの例のように、マルウエア攻撃メールは、SMTP ヘッダ情報のうち、helo/ehlo に加えて送信者アカウントまたは送信者ドメインに特徴があるため、それらの検出を中心としてフィルタを構成することで 9 割以上の検出・阻止が可能となり、阻止については迷惑メールに対する対策および DNS-BL を併用することで 95%以上を阻止できた。なお、DNSBL のみでの阻止率は 80%前後であり、特に特徴が変化した直後のすり抜けが頻発する傾向が見られた。

4 考察

学術雑誌への投稿記事および学術会議への投稿論文および発表申し込みを促すためのメールで、以前は募集している論文、発表のテーマと関わりが深い医師、研究者、医療技術者宛に大手の出版社や大規模学会の国際会議等から送信されるものが大半であった。ところが近年、先駆的な Open Access (OA) 雑誌の成功を受けて、非常に多くの OA 雑誌が刊行されつつあり、その出版元は数百、雑誌タイトルは数千にのぼる[12]。それらの大半は、新興国の科学振興のために雑誌掲載・発表の場を提供することを謳っているが、実際の目的は研究者・医師等をターゲットとした投稿料収入が目的とされている[4,5]。 近年、これも急激に増加した多分野横断型の国際会議も参加料収入が目的であり、プロシーディングを営利目的の出版社から出版するなど、両者の協業も行われている。それらの本拠地は主に南アジア（中東、東南アジアの一部も）であるが、それらからの迷惑メールは、主に米国等の先進国のデータセンターや大手フリーメール業者から送信されている。特徴抽出は容易であるが、正規の出版社や国際会議からのメールとの区別がつきにくいものがあること、および、何らかの事由によって営利目的の雑誌への投稿や営利目的の国際会議への参加を希望するターゲットの存在も考えられ、阻止戦略の策定にはこういった点についても注意を要することが明らかとなった。

最近、国内の医療施設や医学部・医科大学に実在する医師、研究者のメールアドレスを詐称して送信されるマルウエア攻撃が増加している。例えば:

xxxxx@xxxxx.med.xxxx.ac.jp 、
xxxxx@xxhp.xxxxx-u.ac.jp、
xxxx@xxxxxxxx.ac.jp、
xxxxxx-xxxx@xxxx-jxc.org、
xxxx@xxx.ac.jp

のような実在アドレス（x は伏せ字）が用いられていた。最初の例のドメインでは5つのアカウントが使われており、これらは、詐称したアドレスのクレデンシャルがクラックされたものではなく、採取されたアドレスを詐称に用いられた可能性が示唆される。医師、研究者の中にはフリーメールのサーバから所属機関の公式アドレスを用いて送信するもの、および、送信組織によっては、その機関の正規サーバからフリーメールアドレスを送信者アドレスに用いて送信する例が少なからず存在する。そのため、通常、それらのアドレスをホワイトリストに登録することが行われている。送信者が送信に用いるサーバを意識して送信者アドレスを適正に用いてくれない限り、実在アカウントの詐称は有効なフィルタリングの回避策として拡がってゆく危険性が危惧される。

本研究によって、迷惑メールの種類を問わず、標的への到達率を向上させることを目的として、迷惑メールの受信側サイトにおけるフィルタリングを回避するために、共通あるいは類似した送信戦略が採られるようになっている傾向が示された。データマイニングによって、ふたつの方略の存在が推定された。 ひとつは送信者アドレス名によるフィルタリングの回避で、英数字がランダムに並んだアカウントや、非常にポピュラーな字句を持つアカウントの使用であり、これは、送信者アドレスの web 検索を困難にし、かつ、汎用フィルタの構築を困難にすることを企図していると推測される。ただ、現時点では、それらの大半は送信者ドメイン、送信サーバのドメインまたは IP アドレス、helo/ehlo の特徴の組み合わせによって、ほぼ、阻止可能なことが明らかとなった

もうひとつの方略は、受信側で安易に阻止することが困難なサーバから、検索困難なアカウントで送信することである。Gmail、Yahoo、Outlook、Aol 等のポピュラーなメールサービスについて送信ドメイン・送信サーバで規制をかけることは、迷惑メールと同時に多数送信されている正規メールへの影響が極めて大きいことから、それらから送信すれば、受信側でのフィルタリングを相当程度回避できる可能性が高いことが予測される[1]。 これまで、偽装送信者の送信ドメインとして「騙られる」ことが多かった大手フリーメール[1]または ISP のうち、送信元対策が講じられている Gmail や Aol からの迷惑メール送信は容易でなく、事実、それらから送信されて来る迷惑メールは少数であった。しかしながら、大手フリーメールまたは ISP の正規 SMTP サーバが、詐欺メール、フィッシングメールに加えて scientific scam の送信元となってきている。 現在、

それらの大手フリーメールまたは ISP を利用している医療関係者は多く、送受信をそれらにアウトソースしている医療機関、製薬会社、医療機器メーカー、治験関連企業、医科大学等も少なくない。そのため、それらの送信元から送出される迷惑メールについては、今後、効果的に阻止することが困難となる可能性も、また、予測される。

5 結語

医療・研究機関に向けた営利目的の雑誌・学会勧誘メールおよびマルウエア添付メールの SMTP エンベロープヘッダの特徴解析によって、それらの脅威を回避するための方略をマイニング（アソシエーション分析・クラスタ分析等）によって検討した。営利目的の雑誌・学会への勧誘メールの大半は南アジアや欧米のホスティング業者および大手メールサービスから送信されていた。送信者ドメインは直近 2 年以内に登録されたものが多く送信組織を隠蔽する傾向がみられ、送信組織の多くが DNS-BL 回避策を講じていると推定された。マルウエア添付メールはおもにユーラシア大陸北部の大国、東南アジア、東アジアの太鼓等の spam bot 感染 PC からの送信されていた。国内からの送信も増加しつつあり、詐称送信者アドレスに国内の大学・医療機関のドメインや、研究者、医師が用いている個人アドレスが使われるようになってきた。論文検索サイト等で公開されている研究者・医師のアドレスを送受信者に用いることで、受信者の警戒心を解いて送信者の企図した行動をとらせやすくする攻撃の増加が予測される。

6 謝辞

本研究の一部は JSPS 科研費 263307337 の助成を受けた。本研究に関して開示すべき COI はない。

参考文献

1) 渡辺淳, 仲野俊成, 松本掲典, 新貝欣久, 高木真平, 西野典宏. 某大手フリーメール業者から送信される詐欺メールにおける流出ログイン情報の利用. 医療情報学 2013; 33S :1 006-1009.
2) 渡辺 淳,仲野俊成, 松本掲典, 新貝欣久. SMTP エンベロープの特徴解析による 419 scam メッセージの検出. 医療情報学 2011; 31s : 699-702.
3) 渡辺淳, 仲野俊成, 松本掲典, 新貝欣久, 高木真平. SMTP ヘッダにおける限定された情報を用いた詐欺メール送信者の意思決定の道筋解析と展開予測に及ぼす暗黙知の影響. 医療情報学 2012; 32S : 642-645.
4) Butler D. Investigating journals: The dark side of publishing. Nature 2013; 495:433-435.
5) Bloudoff-Indelicato M. Backlash after Frontiers journals added to list of questionable publishers. Nature 2015; 526 : 613.
6) 渡辺 淳, 新貝欣久, 松本掲典, 宮田康央, 仲野俊成. SMTP セッション情報を用いた spam フィルタの構築. 医療情報学 2007; 27s: 1118-1121.
7) 渡辺 淳, 仲野俊成, 松本掲典, 新貝欣久. 医療機関における SMTP セッション情報を用いた迷惑メール対策. 医療情報学 2007; 27s: 1114-1117.
8) 渡辺 淳, 仲野俊成, 松本掲典, 新貝欣久, 畑森浩孝. APML を用いた隠蔽・暗黙関係の描出と展開予測-SMTP snowshoe 攻撃組織の関係解析を用いた検証-. 医療情報学 2009; 29s : 859-864.
9) 樋口耕一. テキスト型データの計量的分析 -2 つのアプローチの峻別と統合-. 理論と方法, 2004;19: 101-115.
10) Hurricane Electric BGP Toolkit. [http://bgp.he.net/ (cited 2017-Sep-7)].
11) TALOS [https://talosintelligence.com/ (cited 2017-Sep-7)].
12) Beall J. Beall's List: Potential, possible, or probable predatoryscholarly open-access. [https://scholarlyoa.com/publishers/ (cited 2017-Jan-10)].
13) Strielkowski W. Predatory journals: Beall's List is missed. Nature 2017; 544 ; 416.
14) Spamhaus [https://www.spamhaus.org/ (cited 2017-Sep-7)].
15) BarracudaCentral [http://www.barracudacentral.org/ (cited 2017-Sep-7)].