

キャンパスネットワークにおける 低コスト迷惑メール対策とその効果について

相馬崇宏[†], 永井謙芝[†], 大島雅明[†], 小宮由里子[‡], 南 弘征[‡], 高井昌彰[‡], 水田正弘[‡]

[†]北海道大学 情報基盤センター 〒060-0811 札幌市北区北 11 条西 5 丁目

[‡]北海道大学 情報基盤センター/大学院情報科学研究科 〒060-0811 札幌市北区北 11 条西 5 丁目

E-mail: [†]{soma, nagai, ohshima}@iic.hokudai.ac.jp, [‡]{komiya, min, takai, mizuta}@iic.hokudai.ac.jp

あらまし 昨今, 特に日本語による性的表現を含む迷惑メールが急増し, 本学の電子メール流量の相当数を占めるに至っている. また, ウイルスや不正侵入後の不当操作による迷惑メールの発出も度々発生することがある. 本報告では, コストをできるだけかけないことを念頭に, これらの問題に対して採った対策ならびに実践例について述べる. 具体的には, ウイルス感染メールの流布防止とコスト面の問題から, 外部とのメール送受信に際して強制的に SMTP コネクションの中継を行うサーバ上で, 過去の迷惑メールの特徴をふまえ, 正規表現によるヒューリスティクスフィルタを導入することで, 誤検出のない迷惑メールフィルタリングを試みる. その効果についての集計結果を示し, 考察と今後の課題を述べる.

キーワード SPAM, SMTP, メールフィルタ

Low cost Anti-SPAM strategy and its effectiveness in our campus network

Takahiro SOMA[†], Norishige NAGAI[†], Masaaki OHSHIMA[†], Yuriko KOMIYA[‡], Hiroyuki MINAMI[‡],
Yoshiaki TAKAI[‡] and Masahiro MIZUTA[‡]

[†]Information Initiative Center, Hokkaido University N11 W5, Kita-ku, Sapporo, 060-0811 Japan

[‡]Information Initiative Center / Graduate School of Information Science and Technology, Hokkaido University
N11 W5, Kita-ku, Sapporo, 060-0811 Japan

E-mail: [†]{soma, nagai, ohshima}@iic.hokudai.ac.jp, [‡]{komiya, min, takai, mizuta}@iic.hokudai.ac.jp

Abstract SPAM mails, especially including some sexual words in Japanese, have been annoying us and the amount is approximately 20% of the total in our campus LAN. Moreover, we, administrative sections, have sometimes detected sending SPAMs from ours, caused by computer viruses or a result of unwilling operations from outside.

In the study, we offer a kind of anti-SPAM tactics whose primary concept is 'low-cost' and show some tentative results in a month. We have a few special MTAs, whose primary function is to check any computer viruses. To set E-mail filters by regular expressions based on many SPAMs we received, we expect to reduce the amount of SPAMs, without false-positive misclassification. Through the tentative numerical result in a recent month, we investigate its effectiveness with discussion.

Keyword SPAM, SMTP, E-mail filtering

1. はじめに

インターネットが社会インフラとして浸透しつつある現在, 電子メールはきわめて基本的なアプリケーションの1つである. ほぼすべてのインターネットユーザが自身のメールアドレスを有すると推定でき, その流量も増加の一途をたどっていると推察される.

しかし, 「流量」の割合は, 俗に SPAM とよばれる迷惑メールと考えられる. 従来の「手紙」と異なり, 電子メールは複数通の送出コストが1通とほぼかわらず, 複製が機械的になされることなどから, 近年, 増加の一途をたどっていると思われる. 従来は英語によるもの, つまり, 欧米圏の SPAM の流れ弾とも呼ぶべき内容が圧倒的であったが, 最近, 特に 2005 年前半

から, 日本語による猥褻語彙を含んだ, いわゆる「出会い系サイト」への誘導, あるいは直接のコンタクトを求めているかのような内容など, 直感的に不快感を覚えるものが無視できなくなっている. 一方で, これら日本語迷惑メールには, 内容のみならず, 送出元がある程度共通しているなどの特徴があるように思われる.

迷惑メール対策としてはすでに多くの報告例があり, また製品化されたものも多いが, これらでは常に誤検出, つまり, 正当なメールであるにもかかわらず, 配送を行わないという問題が生じるおそれがあるほか, 電子メール本文中の単語出現率などに依るものもあり, 管理部署において包括的に導入することについては,

誤検出に伴う技術的な欠点に加え、内容をスキャンすることに対する利用者心理の面からの抵抗も少なからずあるものと思われる。

以上のような背景をふまえ、北海道大学において学内外との電子メール配送の事実上の主管部署である情報基盤センターでは、電子メールの中継箇所において、先に述べたような特徴を正規表現で記載し、合致したメールの配信を一時的に停止させることで、迷惑メールの縮減を試みることとし、2006年3月上旬より運用を継続して行うこととした。

本報告では、4月3日までの約1ヶ月間の結果とその有効性について、種々の経緯をふまえてまとめるとともに、本対策の実施により判明した、電子メールにまつわる種々の問題点を議論する。

2. 迷惑メールについて

2.1. 全体の動向

MITでは、年に1度、SPAM Conference[1]を開催しており、ある程度の動向を知ることができる。2006年は高度なメールフィルタやユーザ認証などについての議論がなされたようであるが、2004年と2005年の講演の中で同一の講演者[2]が、

- ・ SPAMは直接読まれなくてもいい(URLだけ、等)
- ・ サイズは小さく
- ・ 偽装ヘッダーには著名サイトを使用
- ・ Word Salad (フィルタ回避のため、関係のない単語を文中に含める)
- ・ 製品による判別率は98%以上
- ・ 8割弱のエキスパートユーザが撲滅不可能と考えている。
- ・ 85%が迷惑メールフィルタを使用
- ・ 週平均で300通以上の迷惑メール

という興味深い内容を示している。

2.2. 国内の動向

国内において、従来は海外発の英語による迷惑メールが多く見られたが、2005年には、日本語による、主に猥褻な語彙を多く含む迷惑メールが多くみられるようになり、現状に至っている。これらの発信先は、国内の(いわゆる2次ISPを多数有するような)大手ISPや、海外のISPが圧倒的である。また、送出先メールアドレスも、Webなどからの収集に加え、アトラダムな文字列にドメイン名を付して送付し、エラー返送がない場合にはそのまま再送し続けるなど、機械的な対応に依っているケースも見受けられる。

2.3. 学内ネットワーク管理と迷惑メール

大学でネットワーク管理を行う主体部署において、迷惑メールは、既知のものも含め、以下のような問題を起こしている。

(a) 正当なメールとの判別労力

問い合わせアドレスは多くの場合、別名定義されてWebサイトなどに掲載されているため、迷惑メールのターゲットになりやすく、受信対象ユーザにはあまたの迷惑メールが届き、正当なメールとの仕分けの労力が(無駄に)求められる。中には、海外からの正当な警告メールが、先方の善意翻訳サイトを介して日本語で送付されたにもかかわらず、不自然な文体から、迷惑メールと判断し、数日そのままになった、という笑えない経験談もある。

(b) 関連機器への負荷

後述の結果からみて、迷惑メールを捌くためのマシンの負荷は決して無視できるものではない。これはネットワーク回線に対しても同様と思われる。

(c) 利用者からの苦情処理

一般的なアプリケーションレイヤー利用者は、管理部署は万能と思われるのか、迷惑メール阻止の対応を求めてくる。反面、エキスパートは自衛策をとっており、抜本的な対策の導入は自身の方策と相容れないことが多いため、個人分のみを対象から除外するよう、独善的な依頼を寄せることもあり、対応に苦慮することがしばしばである。

(d) 間接的な迷惑行為

北海道大学では、コンピュータウイルスの対外的流布を防止するため、学内と学外の間で強制的に中継MTAを入れ、ウイルスチェックを行っている。そのため、学外から見た場合、本学発のメールの発信元は数台の機材に特定されることになる。ところが、仮に学内から不幸にして迷惑メールが相当数発信された場合、上記の数台がSpamcop.netのRBL(Realtime Blackhole List)に掲載され、本来であれば発信した機材のみが登録されるべきところ、学内発のメールすべてが副作用として中継拒否にあうことがある。抜本的には、Relayを行うMTAすべてにウイルスチェックの機能を持たせるべきところだが、主に予算的な問題から困難と思われる。

また、宛先が実在せずにbounceさせるメールを恣意的に発信し、上記RBLに掲載されることを狙った悪質なものもある。掲載元に「当方の問題ではない」旨を申し入れたところ「SMTPのSourceである以上、自主的に対策を講じよ」と反論されてしまい、昨今の管理の難しさを実感しているところである。

3. 現実的対策の検討

3.1. 方針

これまでに紹介したような内容と、従来の経緯を踏まえ、迷惑メールへの対策として、以下のような方針を立てた。

- ・あらゆる技法を用いても抜本的な迷惑メール削減は困難である。
- ・しかし、現状の機器構成をかえず、大がかりなシステム投資を行わない範囲で、可能な対策があれば、迷惑メールの「削減」を行える可能性がある。
- ・特に、大学構成員の実情に鑑みた場合、日本語による迷惑メールを削減することには、代表照会アドレスに届いたメールを読むことの多い職員に対して、相当の効果が期待できる。
- ・日本語による迷惑メールは発信先、発信方法、必須でないヘッダーの存在など、共通部分が多く、本文をスキャンするまでもなく、メールヘッダーを調べることで、中継を阻止できる可能性が高い。
- ・大学という性格上、メール本文の内容から単語ベースでのシソーラスを作成した上で迷惑メールの判別を行う方法は、出現語彙が非常に多岐にわたることが想定されるため、全学的な制御になじまない。

幸い、前節に示したように、学内～学外間の電子メール配送では、1度、強制的にMTAを介する。このMTA上で、できるだけ簡単な設定を施すことによる対策を採ることとした。

3.2. 既着迷惑メールの分析

これまでに不本意ながら受信した迷惑メールのうち、主に日本語のものについて、メールヘッダーを子細に調べたところ、以下のような類型が伺われた。

(A) SMTP HELO の偽装

SMTP HELO では、自身の FQDN(Fully Qualified Domain Name)を伝えてくるのが原則と思われるが、たとえば到達先である本学の FQDN を名乗るもの、あるいは第三者のドメイン名を名乗るもの、同じ別組織リソースでも、本学の IP アドレスや第三者の IP アドレスを名乗るものが見受けられた。

(B) DNS 未登録、ないし、ポータブル IP ホストからの直接送信

ISP から PPPoE など割り当てられたと思われる機材から直接、メールが送付されている場合が見受けられた。一般には、各 ISP で運用しているメールサーバを介するはずであり、特に最近では、OP25B(Outbound Port 25 Block)の導入により、直接の SMTP 送付は原則

禁止となっている ISP も少なくない。また、DynamicDNS などを用いた固有ドメイン名を割り当てている場合ではその DNS 名が引けるはずであるところ、DNS 登録名からもポータブル IP アドレスであることが強く伺われる場合や、DNS 未登録である場合が伺われた。

(C) RFC 違反ヘッダー

RFC2822 は、事実上、インターネット電子メール配送におけるフォーマットを規定する文書であるが、たとえば、Message-ID: に「@」を含んでいない、ヘッダー名とヘッダー内容のセパレータである「:」の後にホワイトスペースがない、などに違反する場合は迷惑メールに多く見受けられた。

(D) その他

特定の迷惑メールのみが有する付加ヘッダーや、SMTP HELO 時、実際に DNS 登録はなされているものの、特定のパターンを有する場合など、(A)～(C)あるいはそれ以外の条件をあわせた形で表現可能なヘッダーを有する場合が見られた。

3.3. 実現方法

以上のまとめを踏まえ、学内←→学外の強制中継 MTA である postfix に標準装備された、メールヘッダーに関する、正規表現に基づく配送制御機能を用いて、迷惑メールの削減を図ることとした。これは、header_checks として標準導入手順で実装されるファイル中に正規表現を記載することで、メールヘッダーの内容に応じた配送制御(配信拒否、一時待避、別アドレスへの転送など)を可能とする機能である[3]。図 1 は、Message-ID: ヘッダーに「@」を含まないメールを一時待避させ、配送を見合わせる記載例である。

```
if /Message-ID:/  
!/Message-ID:.*@.**/ HOLD  
endif
```

図 1: postfix header_checks ファイル記載例

もっとも、3.2 の分析結果はあくまで「傾向」である以上、すべてを一律に配送拒否するには利用者各位の相応の理解が求められるところである。よって、開始からしばらくの間は、条件に抵触するメールを1度別エリアに待避させ、照会があれば履歴などに基づいて迷惑メールか否かを(本文に依らずに)検討し、正着なものであれば手動で再配送する形を採ることとした。

4. 導入結果と問題点

本節では、具体に実施を始めた 2006 年 3 月 8 日から 4 月 3 日までの結果と、実施中に遭遇した種々の問題点について記す。

なお、本報告の表題に鑑み、現在までの header_checks 中の行数は約 60 行であることを明示しておく。

4.1. 導入結果

表 1 に全体での導入結果を示す。「受信数」は学外発で正着（中継済）メール数、「受信 HOLD 数」は学外発ながら、条件抵触につき配送を差し止めたメール数である。

これによれば、約 1 ヶ月の間に、正着と判断されたメール数とほぼ同数の、およそ 270 万通について配信を回避できたことがわかる。また、中継を拒否したメールの数は、正当に配送されたものとほぼ同数であることや、中継配送停止となった（迷惑メール疑いの）メールは、週末に多く発生していることが伺える。

さらに、3.2 節に示した区分を踏まえ、細分したルール別に集計した結果を表 2 に示す。なお、表 2 では本学発信分のメールも含むため、合計数は一致しない。表 2 の各列の意味は以下の通りである。

- ・ SPPAT: SMTP HELO と DNS 逆引きの組み合わせによる特定のパターン。
From mail.abcd.com (unknown...) など。
- ・ FOD: SMTP HELO 時に本学ドメイン名を詐称したもの。
- ・ FPA: SMTP HELO において、PPPoE などにより一時取得した IP アドレスを使用していると思われるもの。
- ・ SPIPB: 過去の経緯などから、明らかに迷惑メールを送出していると思われる特定の IP アドレスブロック発で、DNS 未定義のもの。
- ・ SPDOM: 過去の経緯などから、明らかに迷惑メールを送出していると思われる特定のドメイン名。
- ・ FOIP: SMTP HELO 時に本学 IP アドレスブロックを詐称したもの。
- ・ IMID: Message-ID: のフォーマットが RFC2822 に従っていないもの。多くは「@」の欠落。
- ・ SPXMI: X-Message-Info: ヘッダーをもつもの。
- ・ FPID: DNS 逆引きによるホスト名に相当数以上の数字列(16 進数を含む)をもち、機械的な割り当てと思量されるもの。
- ・ SPXIP: X-IP ヘッダーのフォーマットが異常なもの。
- ・ SPXIF: X-Info: ヘッダーをもち、特定の内容を有するもの。
- ・ ITZ: Date: におけるタイムゾーンの異常。

たとえば、+1300 など。

表 2 から、SPPAT の多さが際だっていることが読み取れる。これは、info@mail.hoge.com など、最近顕著な日本語迷惑メールの類に合致する。次いで、FOD には、大量送付に使用しているツールの影響か、SMTP HELO で、送付先の MTA の FQDN を名乗る場合が大半である。FPA は、ポータブル IP アドレスからの直送を意味する。また、SPIPB は DNS 逆引きできない場合で、多くは SMTP HELO での名称を偽装している。たとえば、著名な国内組織名を名乗りながら、DNS では引けない、といったブロックについて、各レジストリの Whois などを用いて、属する IP アドレスブロックからのメールのうち、DNS 逆引きできない(ヘッダーには unknown と出る)ものに合致したものである。個別の集計は発表時に示すが、これに合致する IP アドレスブロックはアジア圏が圧倒的である。

4.2. 遭遇した問題

実施から現在までに、以下のような種々の問題に遭遇した。一部は現在も継続中である。

- ・ 正着とおぼしき発信元ながら、ルールに抵触するメールの取り扱い

メールマガジンの配送に際し、迷惑メール発信者と同様の自動送付ソフトウェアを用いている場合や、一部のメールソフトウェアにおいて、SMTP HELO の際に、送信先となる MTA の FQDN や IP アドレスを名乗る場合があるとわかり、その都度、場当たり的な対応を迫られることが少なくない。

また、異動などで学外者となったにもかかわらず、学内で利用していた設定のままメール受信を行おうとして、ルールに抵触した例が出たため、異動時期には当面、設定を緩やかにせざるを得なかった。そのほか、著名なドメインからの送付でありながら、発信元がポータブル IP アドレスの場合もあり、臨機応変な対応が求められる。

- ・ 本方策の運用に依らないメール配送事故に対する照会の増加

本方策の実施までには、相当な周知期間を設けるとともに、どのようなメールが対象となるかの告知も行ったが、単なる先方の誤送信にもかかわらず、本方策に依るものではないか、と照会してくる利用者が後を絶たず、苦慮しているところである。幸か不幸か、実施後に数度、spamcop.net の RBL に送付 MTA が掲載されたこともあり、区別して利用者に啓発を行うのは相応以上に労力を求められることである。

表 1: 正規表現による迷惑メール対策結果

月日	曜日	受信数 (R)	受信HOLD数 (RH)	受信HOLD率 (RH/(R+RH))
3月8日	Wed	141640	76354	35.03%
3月9日	Thu	119535	125903	51.30%
3月10日	Fri	131568	134619	50.57%
3月11日	Sat	76821	110273	58.94%
3月12日	Sun	67419	125527	65.06%
3月13日	Mon	98251	111894	53.25%
3月14日	Tue	120746	109746	47.61%
3月15日	Wed	119504	103394	46.39%
3月16日	Thu	121429	105227	46.43%
3月17日	Fri	122574	106020	46.38%
3月18日	Sat	85664	94216	52.38%
3月19日	Sun	65530	96438	59.54%
3月20日	Mon	101126	122584	54.80%
3月21日	Tue	85304	99847	53.93%
3月22日	Wed	109488	102166	48.27%
3月23日	Thu	117002	93380	44.39%
3月24日	Fri	109157	91618	45.63%
3月25日	Sat	66582	105152	61.23%
3月26日	Sun	69151	95192	57.92%
3月27日	Mon	104101	98977	48.74%
3月28日	Tue	106163	88325	45.41%
3月29日	Wed	97035	82531	45.96%
3月30日	Thu	101147	92176	47.68%
3月31日	Fri	96792	100065	50.83%
4月1日	Sat	68211	101782	59.87%
4月2日	Sun	67434	105287	60.96%
4月3日	Mon	94728	108616	53.41%
合計		2664102	2787309	51.13%

表 2: 細分ルール別 受信 HOLD 数

月日	曜日	Total	(SPPAT)	(FOD)	(FPA)	(SPIPB)	(SPDOM)	(FOIP)	(IMID)	(SPXMI)	(FPID)	(SPXIP)	(SPXIF)	(ITZ)
3月8日	Wed	82023	22794	13531	12156	2863	9523	7065	4235	1938	2798	3577	635	808
3月9日	Thu	128740	36428	25068	21628	3434	11043	10234	6442	3611	2881	5976	941	1056
3月10日	Fri	137534	44388	25239	20998	4537	13771	8285	6207	3683	2915	6150	820	541
3月11日	Sat	112474	44877	17527	15725	5427	9428	7104	3096	2166	2201	3942	891	90
3月12日	Sun	131513	46432	17321	19414	6858	12873	8232	7171	2613	5986	4042	510	61
3月13日	Mon	113142	42522	16446	16099	5311	7645	8574	5050	3015	1786	4906	548	1240
3月14日	Tue	108508	43142	17935	14258	5803	10561	6896	3164	1833	1384	2138	791	603
3月15日	Wed	104347	43844	23872	11171	10545	5738	5460	1819	324	953	31	415	175
3月16日	Thu	106334	45679	19418	13907	8699	6137	5052	3139	2549	1107	18	470	159
3月17日	Fri	105818	38160	19404	12978	14174	4073	4212	5310	5866	926	29	723	963
3月18日	Sat	95444	36560	13895	11372	14914	6948	3860	2578	3261	1228	47	640	141
3月19日	Sun	101029	35608	15400	13216	13697	3866	4228	5617	3334	4591	79	613	780
3月20日	Mon	124557	31581	40481	13394	18850	5636	3999	2978	5255	1978	76	274	55
3月21日	Tue	100719	28672	18454	12266	20357	2549	5881	5989	2650	872	1705	1270	54
3月22日	Wed	104095	26322	21965	16652	17492	5308	5030	2281	3688	1929	1087	1708	633
3月23日	Thu	98004	15529	21109	20724	17236	6021	3840	5177	1196	4667	20	1345	1140
3月24日	Fri	93918	17254	18934	20737	20349	3880	4022	1620	5671	2300	37	519	595
3月25日	Sat	106830	20928	23519	17974	16919	5584	4324	2659	11760	1678	17	753	715
3月26日	Sun	96982	26118	20777	16582	16703	4372	3782	4496	409	1790	15	527	1411
3月27日	Mon	103740	23470	24659	16512	18384	5022	3463	5730	500	4763	46	820	571
3月28日	Tue	89821	19898	23145	15852	18746	4056	3433	1777	320	1496	39	835	224
3月29日	Wed	83924	21561	20026	13046	16600	4129	3260	1887	365	1393	186	1197	274
3月30日	Thu	93410	32042	17536	14977	17785	3476	3043	1690	255	1234	77	1258	37
3月31日	Fri	102731	28404	23958	15794	20882	3423	3342	3037	371	2694	59	885	82
4月1日	Sat	102860	32179	27773	11817	18086	4089	3652	2222	280	1078	26	1428	230
4月2日	Sun	105791	33672	22882	13771	17406	1849	5059	6646	1249	1504	1123	1515	135
4月3日	Mon	110399	34507	23036	16547	16118	5781	5723	3405	1342	1783	1215	876	66
合計		2846687	872569	571290	419567	368075	166781	141055	105422	69504	59915	36663	23007	12839
		100.00%	30.65%	20.07%	14.74%	12.93%	5.86%	4.96%	3.70%	2.44%	2.10%	1.29%	0.81%	0.45%

5. おわりに

本報告では、基幹 MTA に最低限の設定を施すことで迷惑メールを削減した事例の概要と結果について述べた。設定ファイルにおけるわずか 60 行程度の記載量に比して、100 万単位での効果が得られたことは、実施主体としても率直に驚いているところである。

本方策は現在もなお継続中であるが、国内での常識のとおり、「低コスト」の中には人的コストが含まれていない。表 2 のうち、SPIPB、SPDOM など、定常的に動向を調査して随時更新させていく必要のあるルールもあり、全自動的な制御には至っていない。学内利用者の理解が進み、定常的な運用が可能となれば、かなりの部分が定常的に無人対応できると思われるが、現在は多くの面を個別の操作に依っており、担当者を疲弊させている実情は否定できない。ルールファイルのメンテナンス、待避メールの吟味、再送などの諸操作を簡便化するための各種ツールの開発や、主に学内利用者への実効性ある啓発活動が今後の課題である。

文 献

- [1] <http://www.spamconference.org/>
- [2] <http://www.jgc.org/>
- [3] Postfix Manual.
(<http://www.postfix.org/documentation.html>)