LL-005

# 迷惑メールの解析

## Analysis of UCE

**長谷川明生†　山口榮作‡　鈴木常彦†**

Akiumi Hasegawa, Eisaku Yamaguchi and Tsunehiko Suzuki

## １．まえがき

迷惑メールは増加の一途をたどり，インターネットでのトラフィックのかなりの量が迷惑メールで占められているという報告がある。筆者ら[1]は，optout（迷惑メール送信者に対して送信停止を求める行為）が迷惑メールの停止にはつながらず逆に迷惑メール増加の原因になることを示した．この迷惑メール呼び込みの意図は，迷惑メール対策の効果の検証データ取得を目的としている．

2004 年 10 月の optout による迷惑メール呼び込み開始から 24 ヶ月経過時点の 2006 年 9 月末で 139 万通を超える迷惑メールが集まっている．これらのメールについて，ヘッダ情報を Microsoft Access 2003 を用いてデータベース化し解析をおこなった．本論文では，解析により明らかになった点について報告し，あわせて迷惑メール対策についても考察する．

## ２．迷惑メールの収集と時間動向

迷惑メールを呼び込むために optout を実施するとともに，optout 実施の時間，optout したメールアドレスおよび optout 先の URL もしくはメールアドレスをデータベースに記録する．受信専用ドメインのメールサーバは，そのドメイン宛のメールすべてを受信し，ファイルに残す設定になっている．手動で optout したメールアドレス数は，1269 件である．

このようにして受信した迷惑メールの総数は 1392740 通である．受信メール数の月別変動を図１に示す．

図１で，「DB にあり」は，optout に使われたメールアドレス，「DB になし」は，optout の記録がないアドレス，すなわちドメイン上に実態が存在しない捏造されたアドレス宛に送られたメール数を示している．2005 年 10 月のデータから捏造アドレスの使用量の急増が見て取れる．

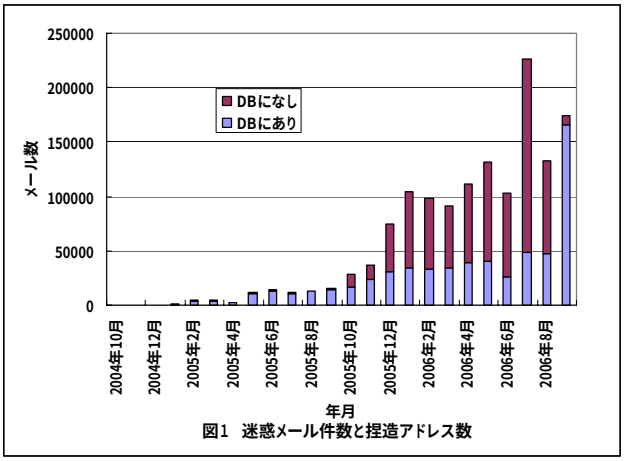


図1　迷惑メール件数と捏造アドレス数

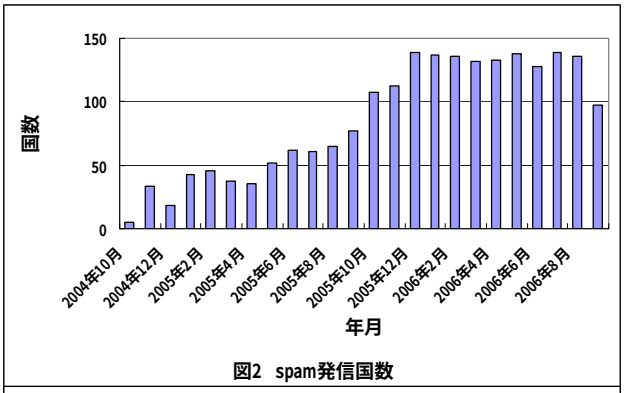


図2　spam発信国数

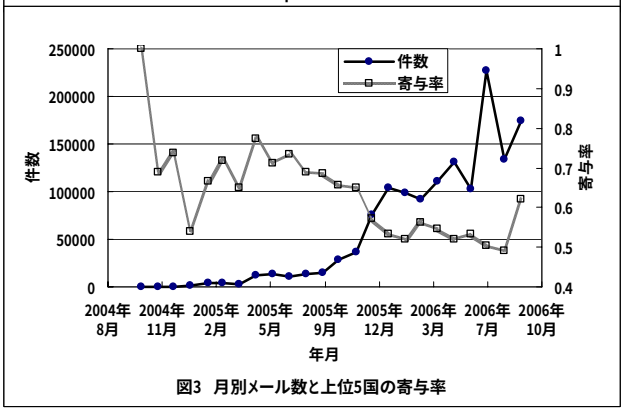


図3　月別メール数と上位5国の寄与率

各メールについて，発信元のIPアドレスの国別分布を公開されているデータ[2]により検索した．図２に迷惑メール発信国数の月別変動を示す．また，メールが発信された国の数および発信数の上位５位の全メールに対する月別寄与率を図３に示す．この図でも，上位 5 国の寄与率および発信国数の変動が 2005 年 10 月前後に起きていることが読み取れる．

## 3．helo パラメータと捏造アドレス

適当な送信者を騙り，捏造されたアドレスに対する迷惑メール送信はバックスキャッタ問題として認識されている．そこで，これらの点について調査した．また，SMTP セッションにおいて helo パラメータのチェックは迷惑メール対策として重要であるといわれている．

### 3.1　バックスキャッタの問題

不正なソフトウェアによって適当に生成されたアドレスに対するメール送信は，いわゆるバックスキャッタ問題に直結する．

†中京大学

‡愛知県立大学

図４に，メールアドレスの種別による受信メール数を示した．図の中で DB にありとは，optout の記録があり正当な受信アドレスであることを示す．不明とは，optout データベースに登録を忘れた可能性のあるアドレスで，この２つを合算しても全メール数の 46%のすぎない．残りの 54%は適当なアドレスを生成して，送られたものであり大量のバックスキャッタの原因となる．

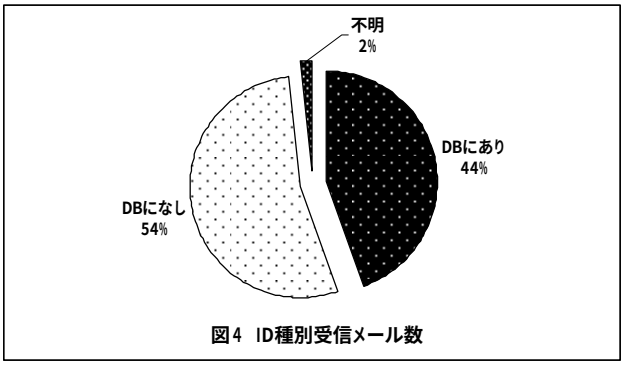


図4 ID種別受信メール数

### 3.2 helo パラメータと DNS の関係

SMTP セッションにおいて，helo パラメータのチェック有効かどうかを検証する．

helo パラメータと IP アドレスおよび DNS の逆引き設定の関係について個々の分類ごとのメールの件数に着目して調査した．メールのヘッダを解析し，メールを IP アドレスからホスト名が検索できないもの，検索可能であるが helo パラメータと一致しないもしくは helo パラメータがないものと helo パラメータと DNS 検索結果に矛盾がないものに分類し，メール数を調べた．その結果を図５に示す．この図から，helo パラメータのチェックを厳密に実施することで 8 割以上の迷惑メールが排除できることがよみとれる．

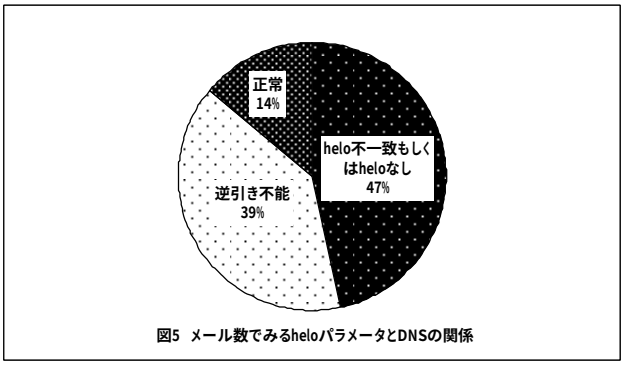


図5 メール数でみるheloパラメータとDNSの関係

つぎに，正当なアドレスかどうかの判定と helo パラメータによる判定を組み合わせると，どのような効果が得られるかを調べてみた．

その結果を表１に示す．

| 種別 | メール数 | パーセント |
|---|---|---|
| DB 登録 helo 正常 | 156651 | 11.2 |
| DB 未登録 helo 正常 | 35377 | 2.5 |
| DB 登録 DNS 未登録 | 310490 | 22.3 |
| DB 登録 helo 偽装 | 151409 | 10.9 |
| DB 未登録 DNS 未登録 | 237735 | 17.1 |
| DB 未登録 helo 偽装 | 500949 | 36.0 |
| DB 登録 DNS なし | 53 | 0.0 |
| DB 未登録 DNS なし | 76 | 0.0 |

表１ helo パラメータとアドレスチェック

表１で，網掛け表示した部分は，helo パラメータが正常で，正当な受信者がいるメールということで，あて先アドレスのチェックと helo パラメータの組み合わせでは排除できない迷惑メールである．

## 4 結論

筆者らの収集した迷惑メールの集合に偏りがないと仮定すると，本研究から，迷惑メールを受信せず，かつバックスキャッタを発生させないためには，メール受信サーバにおいて，あて先アドレスの正当性の検査と，helo パラメータの厳密なチェックの併用が有効であることが結論づけられる．この対策により 9 割近い迷惑メールが排除可能である．

## 5 まとめ

本研究では，収集した迷惑メールから得られる情報の一部を解析しただけで，メール本文を含めた大半の情報の解析は未着手である．予備的にtcpd[3]のログと迷惑メールの発信元アドレスの突合せを実施し，共通のアドレスが存在することを見出したが明確な結論を得られる段階にはいたっていない．本データと他のIDS当のログとの突合せや本文の解析を考えている．

リアルタイムのspam解析については，Ramachandran[4]の研究があるが，迷惑メール対策とあわせて，MailAvenger[5]のようなリアルタイムツールの導入によるより詳細なデータの収集を考えている．

## 参考文献

[1] 長谷川明生，山口榮作，鈴木常彦：”たかが spam，されど spam”，情報処理学会研究報告 (2005-DSM-37),VOl.2005,No.39,pp.75-78(2005.5.12-5.13)

[2] http://software77.net/cgi-bin/ip-country/geo-ip.pl

[3] Z.V.Wietse: ftp://ftp.porcupine.org/pub/security/index.html

[4] A. Ramachandran and N. Feamster:”Understanding the Network-Level Behavior of Spammers”, SIGCOMM’06,Sept.11-15,2006, Pisa, Italy, pp.291-302

[5] D. Mazi`eres: http://www.mailavenger.org/