Spamメールの解析と Spamメール対策

中京大学
生命システム工学部
長谷川　明生

今日の内容

- なぜ，メールを解析しようと思ったか？
- メール解析で新しいことが判明したか？
- Spamは，どこから来るか？
- Spamとメールゲートウェイの関係はどうか？
- どんなSpam対策が考えられるか？
- Spamの特徴抽出には，今後何をするか？

メール解析のきっかけ

- bsfilterの導入によるSpamの自動処理
  - 誤検知の問題の解決
  - SpamとHamを保存して, 解析してみたい。
  - 自動処理と目による処理の差をみつけたい。
  - Spamのアーカイブが20000件を超えた。
- bsfilterと他の手段の比較
  - Spamassassinを試用中
- Spamを根絶したい！

新しい発見はあったか？

- あらためてDNSを当てにできないことを確認
  - 例　emailplanet.com　MX　1. 1. 1. 1
- ベイズフィルタの効果とSpamの手口
- メールの配送遅延とSpamの関係
- 大胆なフィルタを仕掛けるとDDoS?

メールゲートウェイ環境

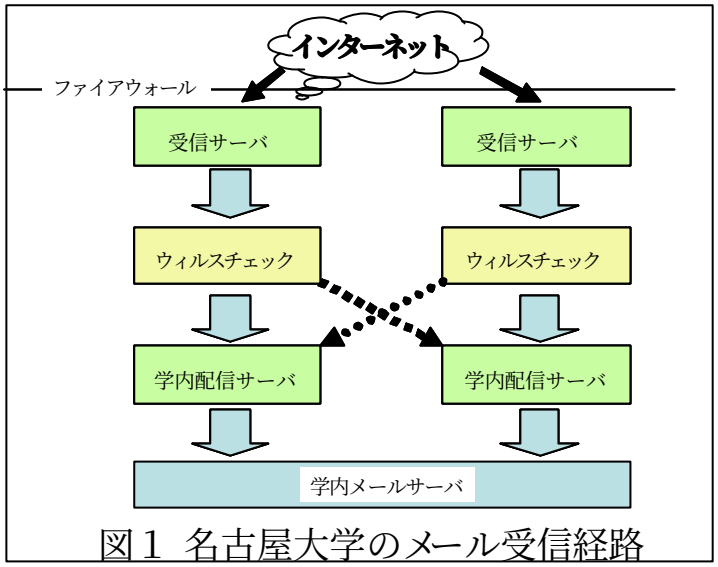


図１ 名古屋大学のメール受信経路

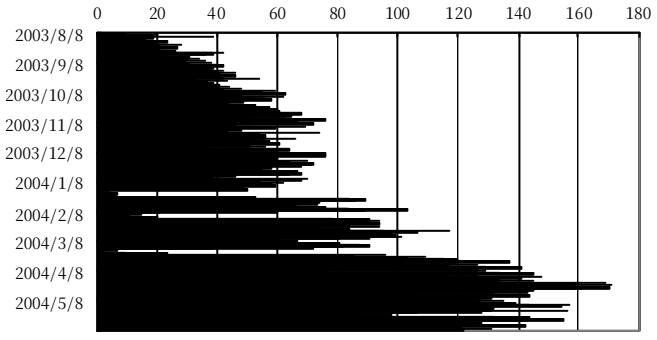
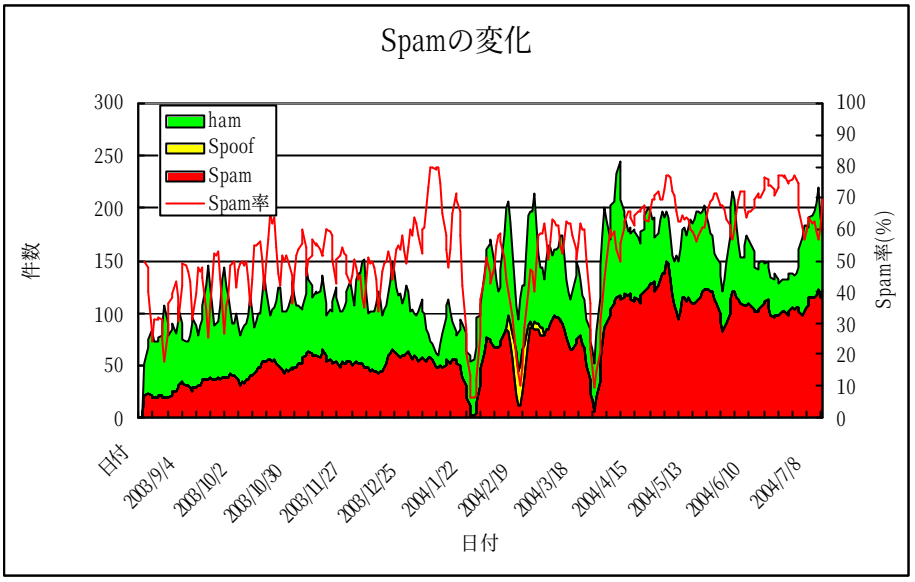


図１　実Ｓpamメール解析によるSpam件数の推移



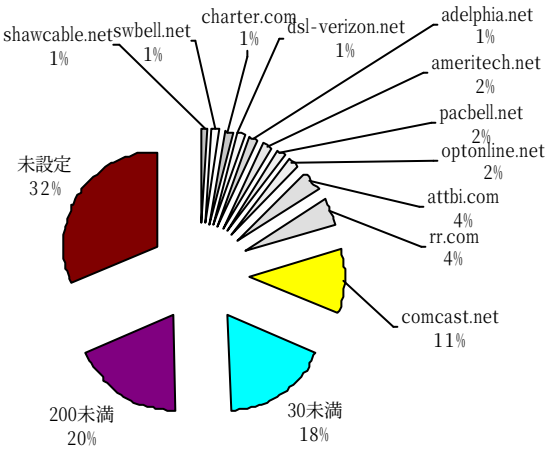


図3　Spam送信ドメイン別分類

表1　IPアドレスから見たSpam送信ドメイン

ドメイン	送付件数	比率（%）
shawcable.net	221	1.13
swbell.net	262	1.34
charter.com	282	1.44
dsl-verizon.net	284	1.45
adelphia.net	293	1.50
ameritech.net	320	1.63
pacbell.net	362	1.85
optonline.net	390	1.99
attbi.com	711	3.63
発信1回	731	3.73
rr.com	822	4.20
発信10回以下	1360	6.95
発信200回以下	1432	7.31
発信100回以下	3806	19.44
comcast.net	2181	11.14
未設定	6125	31.28
合計	19582	100.00

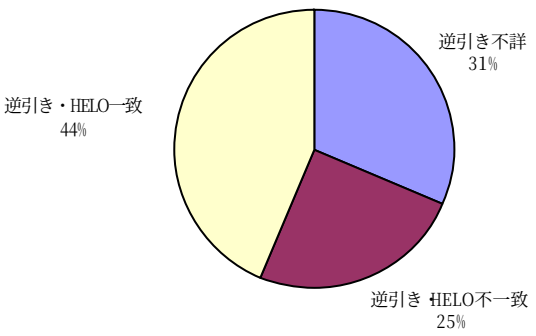


図4　逆引きとHELOパラメータの関係

表　Fromに出現するドメインと件数

ドメイン	件数
yahoo.com.hk	101
yahoo.ca	109
juno.com	122
earthlink.net	136
email.com	148
LYCOS.COM	156
excite.com	184
mail.com	190
attbi.com	239
aol.com	501
MSN.COM	751
HOTMAIL.COM	994
yahoo.com	1665
100回未満	4547
1回きり	4727
１０回未満	5819

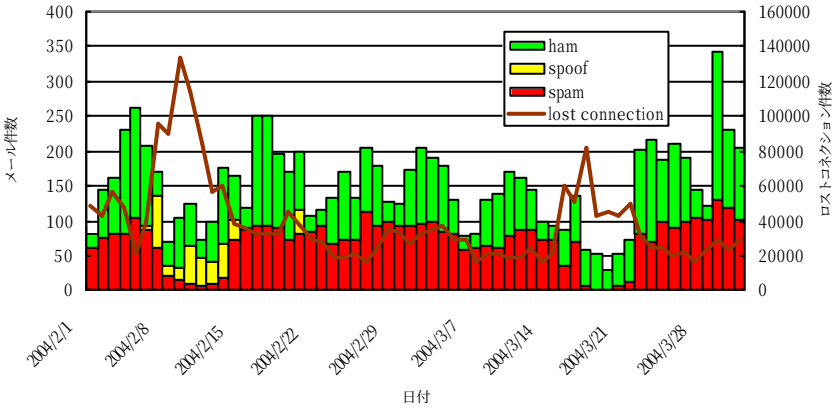


図５　Spamフィルタによるメール分類の時間変化とSMTP接続

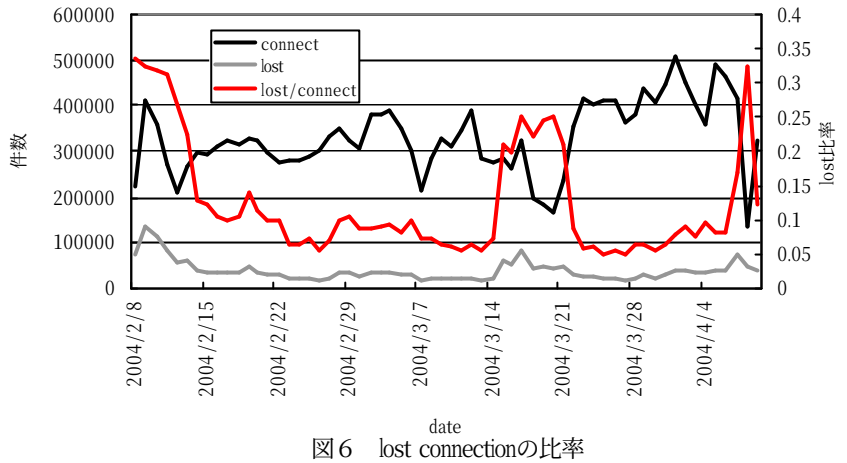
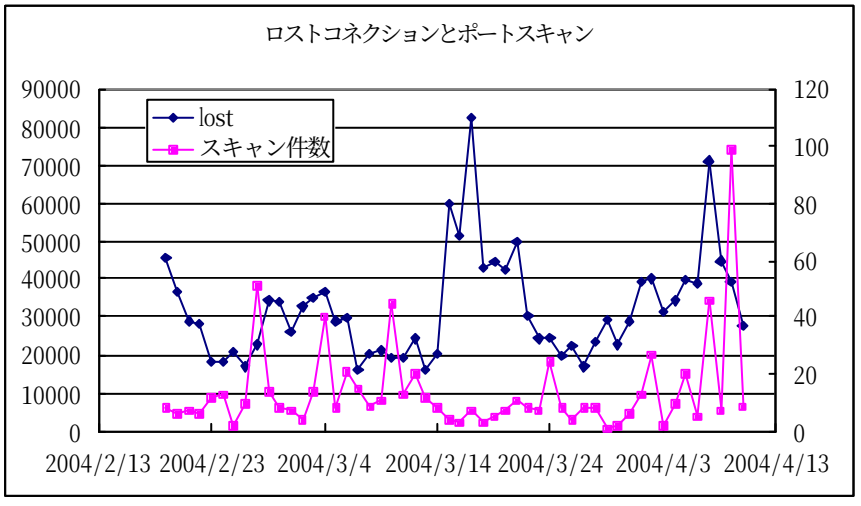


図６　lost connectionの比率



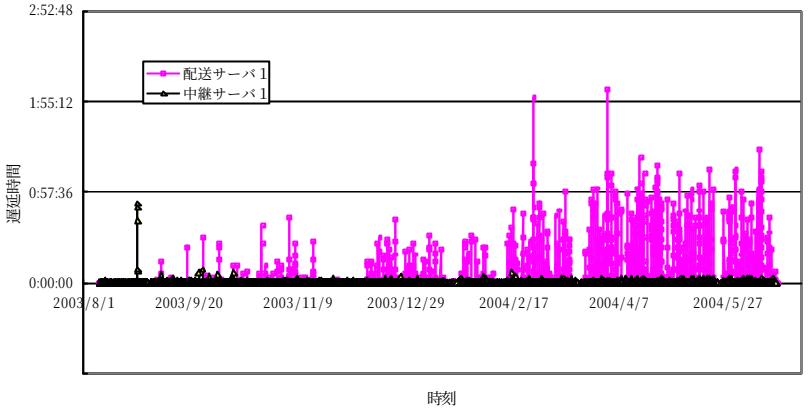
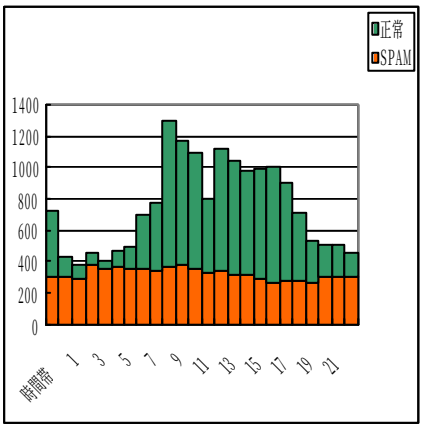
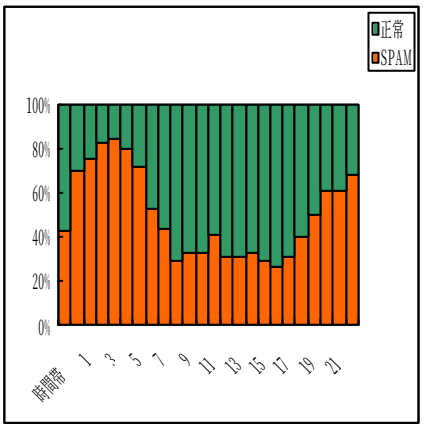


図7　遅延時間の推移 系統1)

Dateでみた時間帯別





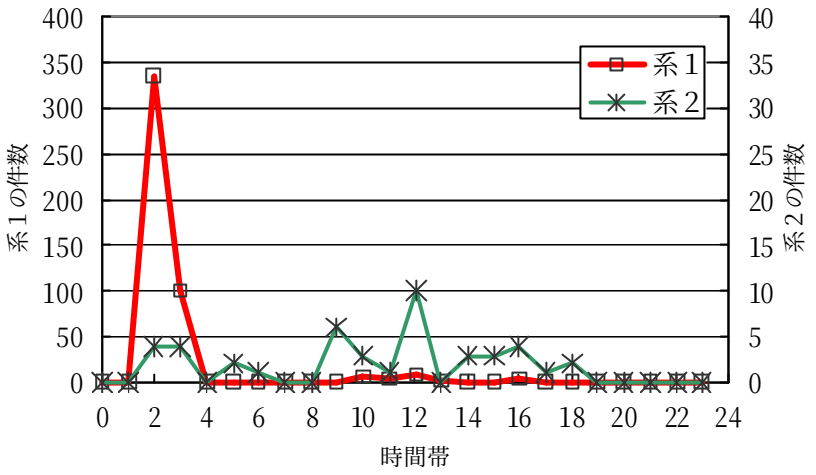


図８　メール配送遅延時間の系統別・時間帯別分布

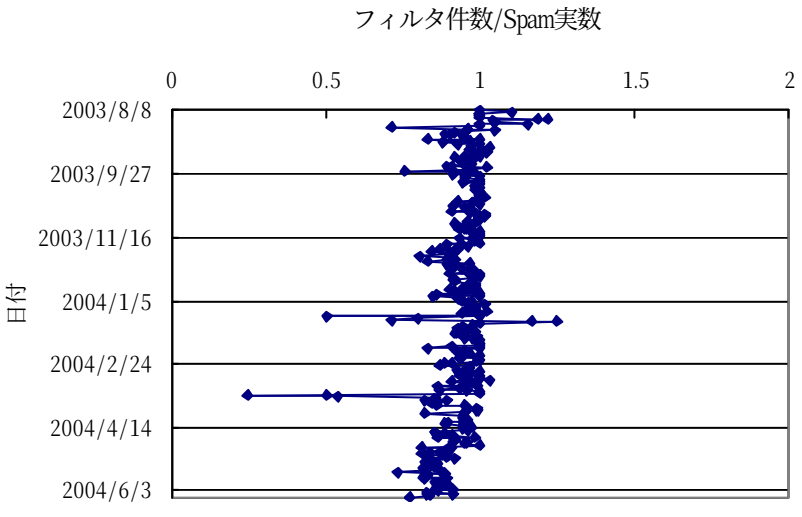


図9　Spamフィルタの検出率の推移

Spamメール対策

- 大口ISPからの接続のフィルタが簡単で効果大
  - が，反撃の可能性も？
    ゲートウェイの能力が重要
  - 正常なものを門前払い
- ベイズフィルタ等
  - 徐々に効果減少
  - 誤検知
  - Spamassassin？
  - 入り口でSpamをマークするのが最善！
- 遅延を許容すれば
  - 一見さんお断り方式？
- SpammerはDNSを見ないらしい！？
  - ときどきゲートウェイのIPアドレスを変更？

FromとMessage-ID

Spamメール	
全数	26383
一致	6218
不一致	20165

正常なメール	
全数	3533
一致	1647
不一致	1886

まとめ

- ベイズフィルタには一定の効果が期待できる。
  - でも，徐々に効率が低下
- Spam送信者は少数らしいが悪質
- 他の指標の調査もしたい
  - Message-IDとheloやFromの関係等
- 人がみたらSpamが簡単に判定できるのに！
  - メールの中身の解析（無意味単語, 偽装単語）
    例　Vi@g@とかV.i.c.o.d.i.n, s p a m