

C3 迷惑メールは発信時刻を偽るか？

鈴木 孝彰 水沼 暁

1. はじめに

迷惑メールは多いときに数百件以上届くこともあり、Email の利便性を大きく損なう。[1]では、迷惑メールの送信元である懸賞サイトに登録し、最初の迷惑メールが届くまでの日数や、懸賞サイトの迷惑メール発信率が報告されている。しかし、送信元でアドレスや時間を偽装することは容易であり、これらが考慮されていない。

そこで本研究では、迷惑メールの送信元の送信時刻と、受信側の受信時刻の差をとり、偽装時刻の大きさを可視化するシステムを開発した。本システムの実装と、それを用いて抽出した特徴について報告する。本システムを応用し、送信時刻を偽装しているものを迷惑メールと判断すれば、効果的に受信メールから除外することができる。

2. 研究機材システム概要

ふらっとホーム社 OpenMicroServer(CMS-AL400/128)を用い、Debian GNU/Linux 4.0 etch、gnuplot 4.0、Java JDK 1.5.0、Perl 5.8.8 を利用して可視化システムを実現している。

3. 迷惑メールデータ

3.1 収集方法

[1]と同様に、(1)無料の出会い系サイトへの登録、(2)Web上にメールアドレスを公開するという2つの手法を行った。前者は「Yahoo!メール」を利用し、無料出会い系サイト内で、有料サイトに登録を促してくるサクラに対しメールを送信した。後者は、ウェブ上で公開された実メールアドレスへ送られたメールから抽出した。

3.2 解析方法

2008年4月1日から10月31日までに届いた迷惑メールを使用した。ランダムに100個のサンプルを抽出し、迷惑メールの受信時刻と送信時刻の差を観測した。

4. 解析結果

図1に、(1)と(2)の迷惑メールの送信元IPアドレスの第一オクテットの比率を示す。図より、二つのデータに、共通するアドレス帯と類似(60～100, 190～220)がある一方で、(2)のみメールが届くアドレス(40～55)などがあることがわかる。(2)では、中継に用いられた最初のメールサーバのアドレスを送信元とみなしているが、それがプライベートに割り当てられていることがあるためと考えられる。

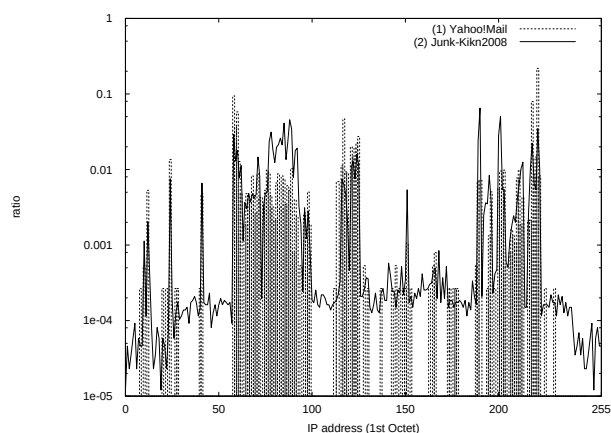


図1. 迷惑メールの第一オクテットアドレスについての分布図

図2は、X軸を迷惑メールの受信時刻、Y軸を送信時刻とした散布図である。対角線に近いほど、受信時刻と送信時刻の差は小さい。図2より、1ヵ月後と2ヵ月後、および1ヶ月前に大きなピークが見える。

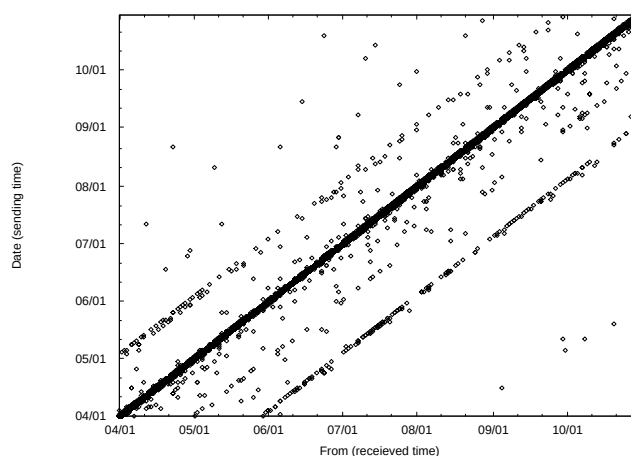


図2. 迷惑メールの送信時刻と受信時刻の関係

5. おわりに

発信時刻を偽っている迷惑メールを解析ツールで表した。中には送信時刻と受信時刻の差が数年といったメールもあり、そういったメールは明らかに信憑性がない。しかし、その差が数時間から数日の場合、意図的に偽装をしているのか、時刻の設定の誤りなのか、またはメールの配送中に遅延が発生しているのか明らかではない。今後は、時刻に差がでる理由を追求し、迷惑メールであるかどうか明確にしていきたい。

参考文献

[1] 斉藤, 望月, “懸賞サイトに登録するとスパムは来るの?”, 2006年度東海大学卒業研究, 2007.