日本オペレーションズ・リサーチ学会
2004年春季研究発表会

メール型コンピュータウィルス拡散過程モデルの提案

01206600 NTT サービスインテグレーション基盤研究所 *佐藤大輔 SATOH Daisuke
NTT サービスインテグレーション基盤研究所 内田真人 UCHIDA Masato
01013530 NTT 情報流通プラットフォーム研究所 石橋圭介 ISHIBASHI Keisuke
02103840 東京工業大学大学院情報理工学研究科 小林真 KOBAYASHI Makoto

1 はじめに

コンピュータネットワークが普及するに伴い，コンピュータウィルスの脅威は深刻なものになっている．それに伴い，様々なアプローチから研究がなされている．その中のひとつに，ウィルスの拡散を微分方程式によってモデル化する研究があり，CodeRed ウィルスがロジスティック曲線モデルに適合するという報告 [1] がある．しかし，ロジスティック曲線モデルは，全てのウィルスに適合するわけではない．本論では，コンピュータウィルスの感染方法に注目し，多くのウィルスで使われている電子メールを使った感染方法についてウィルスの拡散過程のモデル化を行う．また，実データによる比較検証を行う．

2 ネットワーク型とメール型モデル

鈴木 [2] によれば，コンピュータウィルスは，大きくファイルに感染するもの，ファイルに感染しないものの 2 種類に分類され，ファイルに感染しないものは，さらにトロイの木馬，コンピュータワームの 2 つに分類される．現在，特に問題視されているのは，コンピュータワームに分類されるものである．

コンピュータワームには，電子メールによって感染していく電子メール型と電子メールとは異なる通信プロトコルによって感染していくネットワーク型がある [2]．本論では，この電子メール型とネットワーク型の感染形態のモデル化について議論する．

ネットワーク型のコンピュータウィルスのひとつである CodeRed の拡散がロジスティック方程式によってモデル化されることは既に報告されている [1]．このことは，感染した全てのノードが継続的に未感染ノードの感染に寄与していることを意味している．

一方，電子メール型では，感染したノードのアドレス帳を元に感染を拡大していくもので，感染したノードは，アドレス帳に載っている他のノードにウィルスを感染した後は，継続的に未感染ノードの感染に寄与しない．そこで，次のような差分方程式によるモデルを提案する．

$$M_{n+1} - M_n = \delta\alpha(N - M_n)(M_n - M_{n-1}) \qquad (1)$$

ここで，M_n は n ステップ時における既感染ノード数であり，N はネットワーク内の全ノード数である．このモデルは，ロジスティック方程式を前進差分した式

$$M_{n+1} - M_n = \delta\alpha(N - M_n)M_n \qquad (2)$$

と比較してわかるように，新規感染ノード数がロジスティック方程式の場合，既に感染したノード数と未感染ノード数との積に比例するのに対して，本提案モデルは，未感染ノード数と直近に感染したノード数の積に比例するモデルになっている．このことは，メール型ウィルスの感染形態をモデル化したものであると言える．

3 微分方程式とその解

式 (1) において，

$$k = N - \frac{1}{\delta\alpha} \qquad (3)$$

とおき，両辺から $M_n - M_{n-1}$ を引いて，両辺を δ^2 で割ると

$$\frac{M_{n+1} - 2M_n + M_{n-1}}{\delta^2} = \alpha(k - M_n)\left(\frac{M_n - M_{n-1}}{\delta}\right) \qquad (4)$$

となる．ここで $\delta \to 0$ とすると次の微分方程式

$$\frac{d^2M}{dt^2} = \alpha(k - M)\frac{dM}{dt} \qquad (5)$$

が得られる．ここで

$$K = \sqrt{k^2 + \frac{2C_1}{\alpha}} \tag{6}$$

とおくと，この微分方程式の解は

$$M(t) = k + K \tanh\left(\frac{\alpha K}{2}(t + C_2)\right) \tag{7}$$

である．ここで，C_1, C_2 は積分定数である．以下のような極限値となる．

$$\lim_{t\to\infty} M(t) = k + K \tag{8}$$

$$\lim_{t\to-\infty} M(t) = k - K \tag{9}$$

また，式 (5) は，

$$\frac{d^2M}{dt^2} = \frac{d}{dt}\left(-\frac{\alpha}{2}(M-k)^2\right) \tag{10}$$

と書き直せるので，両辺積分して

$$\frac{dM(t)}{dt} = \frac{\alpha}{2}\left(K^2 - (M(t)-k)^2\right) \tag{11}$$

となる．式 (11) は Riccati 方程式である．

4 実データによる評価

実データによる評価を行うためには，パラメータ推定をする必要がある．式 (11) を基にパラメータ推定を行う．Riccati 方程式の精度の高いパラメータ推定については，厳密解を持つ差分方程式 [3] によるパラメータ推定法が提案されている [4]．式 (11) に対応する厳密解を持つ差分方程式は

$$M_{n-1} - M_n = \delta\frac{\alpha}{2}(K+k-M_{n+1})(K-k+M_n) \tag{12}$$

であり，厳密解は，

$$M(t) = k - K\frac{(1-\delta\alpha)^{\frac{n-n_0}{2}} - (1-\delta\alpha)^{-\frac{n-n_0}{2}}}{(1-\delta\alpha)^{\frac{n-n_0}{2}} + (1-\delta\alpha)^{-\frac{n-n_0}{2}}} \tag{13}$$

となる．この差分方程式を使用してパラメータ推定を行う．ちなみに，式 (1)，式 (5) に対応する厳密解を持つ差分方程式はそれぞれ

$$M_{n+1} - M_n = \delta\alpha(k + \frac{1}{\delta\alpha} - M_n)\left(\frac{M_{n+1} - M_{n-1}}{2}\right) \tag{14}$$

$$M_{n+1} - 2M_n + M_{n-1} = \frac{\delta^2\alpha}{1-\frac{\delta\alpha K}{2}}(k - M_n)\left(\frac{M_{n+1} - M_{n-1}}{2\delta}\right) \tag{15}$$

となる．

実データは，メール型ウィルスである Aliz のデータ (データ提供：トレンドマイクロ社) を使用した．図 1 からわかるように，明らかにロジスティック曲線モデルとは異なる成長曲線である．提案モデルによる推定値は，実データと良く適合していることがわかる．

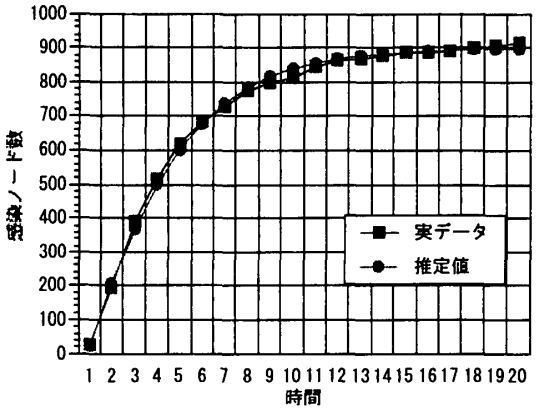

図 1: 実データとモデルによる推定値との比較

5 まとめ

本論では，メール型ウィルスの感染過程のモデルを提案し，実データで検証した．このモデル化により最終的な感染ノード数の予測や感染率の推定などが可能となる．

参考文献

[1] S. Staniford, V. Paxson, N. Weaver: How to Own the Internet in Your Spare Time, *Proceedings of the 11th USENIX Security Symposium*, (San Francisco , 2002) 149–167.

[2] 鈴木光勇: なぜコンピュータウィルスは悪さが出来るのか？, 毎日コミュニケーションズ (2003).

[3] R. Hirota: Nonlinear partial difference equations. V. Nonlinear equations reducible to linear equations. *Journal of the Physical Society of Japan*, **46** (1979) 312–319.

[4] D. Satoh: A Discrete Bass Model and its Parameter Estimation, *Journal of the Operations Research Society of Japan*, **44-1** (2001) 1–18.