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

コンピュータウィルス拡散過程モデルの解析

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
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

1 まえがき

インターネットの普及に伴い，ネットワークを介して感染するコンピュータウィルスの脅威が増している．そのため，コンピュータウィルスの挙動のモデル化や，それがもたらす被害の大きさに関する検討を行う事は重要である．こうした背景の下，コンピュータウィルスの拡散過程に関する以下のような研究結果が報告されている．[1] では，Code-Red や Nimda のように，現時刻における累積感染ノードが次時刻における未感染ノードの感染に寄与する場合の拡散過程がロジスティック方程式で表現できることが報告されている．一方，[2] では，Aliz や Sobig のように，現時点における新規感染ノードのみが次時刻における未感染ノードの感染に寄与する場合の拡散過程を表現する差分方程式が提案されている．そこで，本稿では，これらの結果を確率論の視点から再考し，いくつかの解析結果を示す．なお，本稿は [3] の内容を詳細化したものである．

2 モデル化

本稿では，[2] に従い，コンピュータウィルスの感染形態のモデルとして，以下の二種類を検討する．なお，以下では，ネットワークを構成する端末（ノード）の個数を N とおく．

- 時刻 n における累積感染ノードが互いに独立に，時刻 $n+1$ において，ランダムに選択した A $(1 \le A \le N)$ 個のノードを攻撃し，確率 p で感染させる（Repetitive Attack モデル：RA モデル）．
- 時刻 n における新規感染ノードのみが互いに独立に，時刻 $n+1$ において，ランダムに選択した A $(1 \le A \le N)$ 個のノードを攻撃し，確率 p で感染させる（Sequential Attack モデル：SA モデル）．

ここで，A は確率変数である．このとき

$$\Pr\{A=a\} = \binom{N}{a} q^a (1-q)^{N-a}$$

とすることによって，時刻 n における攻撃元ノードと攻撃先ノードのなすグラフ構造は，(古典) ランダムグラフとみなすことができる $(0<q<1)$ [4]．一方，時刻 n における攻撃先ノード数を固定（a 個）とするためには

$$\Pr\{A=a\} = 1$$

とすれば良い．また，RA モデル，SA モデルはそれぞれワーム型ウィルス，メール型ウィルスの感染形態を表現するモデルであることを付記しておく [2]．なお，現実の感染形態においては，攻撃元ノードが互いに独立に攻撃を行うとは限らないと考えられるが，このことについては今後の課題とする．

3 解析

時刻 n における累積感染ノード数を表す確率変数を，RA モデル，SA モデルについてそれぞれ R_n，S_n とおく．なお，以下では，ある確率変数 X,Y について $\mathbb{E}[X]$ は X の期待値を表し，$\mathbb{E}[X|Y]$ は Y を条件とする X の条件付き期待値を表す．さらに，記法の簡単のため，ある確率変数 X,Y について $X=Y$, w.p.1 が成り立つ場合，誤解の恐れが無い限り単に $X=Y$ と書く事とする．

まず，S_n について解析する．N 個の要素からなるノード集合を V とおく．さらに，時刻 n における新規感染ノード数を表す確率変数を $\bar{S}_n = S_n - S_{n-1}$ とおき，新規感染ノード集合を $V_n = \{v_{n,1}, v_{n,2}, \cdots, v_{n,\bar{S}_n}\} \subset V$ とおく．ただし，一般性を失うことなく，V_n の各要素は $v_{n,1}, v_{n,2}, \cdots, v_{n,\bar{S}_n}$ の順に互いに独立に $A_{n,1}, A_{n,2}, \cdots, A_{n,\bar{S}_n}$ 個のノードに対してそれぞれ攻撃を行うとする．ここで，$A = A_{n,i}$ である $(i = 1, \cdots, \bar{S}_n)$．このとき，$v_{n+1,i}$ が攻撃した $A_{n+1,i}$ 個のノードの内，ウィルス未感染ノードの数を表す確率変数を $\bar{S}_{n+1,i}$ とおき，$v_{n,i}$ の攻撃が終了した段階での累積感染ノード数を表す確率変数を $S_{n+1,i}$ とおけば，$S_{n+1,j} = S_n + \sum_{i=1}^{j} \bar{S}_{n+1,i}$ となる $(j = 1, \cdots, \bar{S}_n)$．ただし，$S_{n+1} = S_{n+1,\bar{S}_n}$ である．

以上の定義より

$$\mathbb{E}[S_{n+1,j}|S_n, \bar{S}_n, A_{n+1,j}] = \mathbb{E}[S_{n+1,j-1}|S_n, \bar{S}_n] + p\mathbb{E}[\bar{S}_{n+1,j}|S_n, \bar{S}_n, A_{n+1,j}]$$

が成り立つ $(j = 1, \cdots, \bar{S}_n)$．また

$$\Pr\{\bar{S}_{n+1,j} = \bar{s}|S_n, \bar{S}_n, \bar{S}_{n+1,1}, \cdots, \bar{S}_{n+1,j-1}, A_{n+1,j}\} = \binom{N - S_{n+1,j-1}}{\bar{s}} \binom{S_{n+1,j-1}}{A_{n+1,j} - \bar{s}} \Big/ \binom{N}{A_{n+1,j}}$$

であるので

$$\mathbb{E}[\bar{S}_{n+1,j}|S_n, \bar{S}_n, \bar{S}_{n+1,1}, \cdots, \bar{S}_{n+1,j-1}] = \alpha(N - S_{n+1,j-1})$$

が成り立つ（超幾何分布の期待値）．ただし，$\alpha = \mathbb{E}[A]/N$ である．このことから

$$\mathbb{E}[\bar{S}_{n+1,j}|S_n, \bar{S}_n] = \alpha(N - \mathbb{E}[S_{n+1,j-1}|S_n, \bar{S}_n])$$

が成り立つ．以上より

$$\begin{aligned}\mathbb{E}[S_{n+1,j}|S_n, \bar{S}_n] &= N - (N - \mathbb{E}[S_{n+1,j-1}|S_n, \bar{S}_n])(1-p\alpha) \\ &= N - (N - S_n)(1-p\alpha)^j\end{aligned}$$

が成り立つ．さらに，$\bar{S}_n = S_n - S_{n-1}$ であるので

$$\mathbb{E}[S_{n+1}|S_n, S_{n-1}] = N - (N - S_n)(1-p\alpha)^{S_n - S_{n-1}}$$

が成り立つ．R_n についても，ほぼ同様の議論を行うことで，以下の定理が導かれる．

定理 1. R_n，S_n について

$$\mathbb{E}[R_{n+1}|R_n] = N - (N - R_n)(1-p\alpha)^{R_n}$$
$$\mathbb{E}[S_{n+1}|S_n, S_{n-1}] = N - (N - S_n)(1-p\alpha)^{S_n - S_{n-1}}$$

が成り立つ $(n = 0, 1, 2, \cdots)$．□

ここで

$$\mathbb{E}[R_{n+1}] = \mathbb{E}[R_{n+1}|R_n = \mathbb{E}[R_n]] \quad (1)$$
$$\mathbb{E}[S_{n+1}] = \mathbb{E}[S_{n+1}|S_n = \mathbb{E}[S_n], S_{n-1} = \mathbb{E}[S_{n-1}]] \quad (2)$$

と仮定すると，定理 1 より

$$\mathbb{E}[R_{n+1}] = N - (N - \mathbb{E}[R_n])(1-p\alpha)^{\mathbb{E}[R_n]} \quad (3)$$
$$\mathbb{E}[S_{n+1}] = N - (N - \mathbb{E}[S_n])(1-p\alpha)^{\mathbb{E}[S_n] - \mathbb{E}[S_{n-1}]} \quad (4)$$

が導かれる $(n = 0, 1, 2, \cdots)$．以下では，$\mathbb{E}[R_0] = 1$, $\mathbb{E}[S_0] = 1$, $\mathbb{E}[S_{-1}] = 0$ と設定する．このとき，式 (3), (4) より，以下の定理が導かれる（証明は付録 A 参照）．

定理 2. S_n，R_n について

$$\lim_{n\to\infty} \mathbb{E}[R_n] = N$$

$$\lim_{n\to\infty} \mathbb{E}[S_n] = N - \frac{W((N-1)(1-p\alpha)^N \ln(1-p\alpha))}{\ln(1-p\alpha)}$$

が成り立つ．ただし，W は乗積対数関数である（Lambert の W 関数とも呼ばれる）[5]． □

また，式 (3)，(4) で与えられる差分方程式について，$n=\frac{t}{\delta}$ と特殊化し，$R(t)=R_n$，$S(t)=S_n$ とおく．さらに，式 (3) で与えられる差分方程式について $1-p\alpha=(1-p\bar{\alpha})^{\delta}$ と特殊化し，式 (4) で与えられる差分方程式について $1-p\alpha=(1-p\bar{\alpha})^{\delta}$，$N=\bar{N}-\frac{1}{\delta\ln(1-p\bar{\alpha})}$ と特殊化すると，$\delta\to 0$ のとき，以下の定理が導かれる（証明は付録 B 参照）．

定理 3. $R(t)$，$S(t)$ について

$$\frac{d\mathbb{E}[R(t)]}{dt}=-\{\ln(1-p\bar{\alpha})\}\mathbb{E}[R(t)](N-\mathbb{E}[R(t)]) \quad (5)$$

$$\frac{d^2\mathbb{E}[S(t)]}{dt^2}=-\{\ln(1-p\bar{\alpha})\}\frac{d\mathbb{E}[S(t)]}{dt}(\bar{N}-\mathbb{E}[S(t)]) \quad (6)$$

が成り立つ． □

4 考察

前節で導いた結果と，[1, 2] との関係について考察する．

式 (3)，(4) は，$p\alpha\ll 1$ のとき，それぞれ

$$\begin{aligned}\mathbb{E}[R_{n+1}]&\approx N-(N-\mathbb{E}[R_n])(1-p\alpha\mathbb{E}[R_n])\\&=\mathbb{E}[R_n]+p\alpha\mathbb{E}[R_n](N-\mathbb{E}[R_n]) \quad (7)\\ \mathbb{E}[S_{n+1}]&\approx N-(N-\mathbb{E}[S_n])\{1-p\alpha(\mathbb{E}[S_n]-\mathbb{E}[S_{n-1}])\}\\&=\mathbb{E}[S_n]+p\alpha(\mathbb{E}[S_n]-\mathbb{E}[S_{n-1}])(N-\mathbb{E}[S_n]) \quad (8)\end{aligned}$$

と近似できる．ここで，式 (7) はロジスティック差分方程式である．したがって，この結果は [1] の主張と一致していることがわかる．また，式 (8) は [2] で提案されている差分方程式に一致していることがわかる．一方，定理 3 の結果も [1, 2] に一致していることがわかる．なぜならば，式 (5) はロジスティック微分方程式であり，式 (6) は [2] で導出された微分方程式に（係数を除いて）一致しているからである．ただし，式 (5)，(6) は $p\alpha\ll 1$ の条件を用いずに導出した結果であることに注意されたい．

ところで，定理 2 より，SA モデルにおける平均累積感染数は全ノード数 N に収束しない場合があることが分かる．このことは，SA モデルにおいては，時刻 n における新規感染ノードのみが，時刻 $n+1$ において攻撃を行うことに起因する．すなわち，時刻 t における新規感染ノードが存在しない場合（時刻 $n-1$ における全ての攻撃対象が時刻 $n-1$ において既感染である場合）は，ウィルスの拡散が停止してしまうということである．このことに対し，RA モデルにおいては，時刻 n における累積感染ノードが，時刻 $n+1$ において攻撃を行う．そのため，仮に，時刻 n における新規感染ノードが存在しない場合であっても，ウィルスの拡散が継続し，最終的には全てのノードが感染するのである．

5 むすび

本稿では，コンピュータウィルスの拡散過程について，確率論の視点から解析を行った．また，この解析によって得られた結果と [1, 2] との関係を明らかにした．今後の課題としては，式 (1)，(2) により与えられる仮定の妥当性を，理論及びシミュレーションにより評価する事や，攻撃元ノードが互いに独立に攻撃を行わない場合の解析を行うことなどが挙げられる．

付録

A 定理 2 の証明

まず，R_n について証明する．

背理法を用いて証明するために $\lim_{n\to\infty}\mathbb{E}[R_n]\neq N$ と仮定する．このとき，式 (3) より $\lim_{n\to\infty}(1-p\alpha)^{\mathbb{E}[R_n]}=1$ となる．ここで，$p\alpha\neq 0$ であることから $\lim_{n\to\infty}\mathbb{E}[R_n]=0$ となるが，これは $\mathbb{E}[R_n]\geq 1$ に矛盾する．

次に，S_n について証明する．

式 (4) より

$$\begin{aligned}N-\mathbb{E}[S_{n+1}]&=(N-\mathbb{E}[S_n])(1-p\alpha)^{\mathbb{E}[S_n]-\mathbb{E}[S_{n-1}]}\\&=(N-1)(1-p\alpha)^{\mathbb{E}[S_n]}\end{aligned}$$

となる．ここで $\mathbb{E}[S_\infty]=\lim_{n\to\infty}\mathbb{E}[S_n]$ とおくと

$$\begin{aligned}&(N-1)(1-p\alpha)^N\ln(1-p\alpha)\\&=\{\ln(1-p\alpha)\}(N-\mathbb{E}[S_\infty])\exp^{\{\ln(1-p\alpha)\}(N-\mathbb{E}[S_\infty])}\end{aligned}$$

となるので，乗積対数関数の定義より

$$\begin{aligned}&W((N-1)(1-p\alpha)^N\ln(1-p\alpha))\\&=\{\ln(1-p\alpha)\}(N-\mathbb{E}[S_\infty])\end{aligned}$$

が成り立つ．

B 定理 3 の証明

まず，$R(t)$ について証明する．

式 (3) より

$$\begin{aligned}&\ln(N-\mathbb{E}[R(t+\delta)])-\ln(N-\mathbb{E}[R(t)])\\&=\delta\{\ln(1-p\bar{\alpha})\}\mathbb{E}[R(t)]\end{aligned}$$

が成り立つ．このとき，上式の全体を δ で割り，$\delta\to 0$ とすれば良い．

次に，$S(t)$ について証明する．

式 (4) より

$$\begin{aligned}&-\{\ln(1-p\alpha)\}(N-\mathbb{E}[S(t)])(\mathbb{E}[S(t)]-\mathbb{E}[S(t-\delta)])\\&=-(N-\mathbb{E}[S(t)])\ln\left(\frac{N-\mathbb{E}[S(t+\delta)]}{N-\mathbb{E}[S(t)]}\right) \quad (9)\end{aligned}$$

が成り立つ．ここで

$$1-\frac{1}{x}\leq\ln(x)\leq x-1,\quad\text{（等号は}x=1\text{のときに限る）}$$

が成り立つことに注意すると，式 (9) より

$$\begin{aligned}&-(N-\mathbb{E}[S(t)])\left(\frac{N-\mathbb{E}[S(t+\delta)]}{N-\mathbb{E}[S(t)]}-1\right)\\&\leq-\{\ln(1-p\alpha)\}(N-\mathbb{E}[S(t)])(\mathbb{E}[S(t)]-\mathbb{E}[S(t-\delta)])\\&\leq-(N-\mathbb{E}[S(t)])\left(1-\frac{N-\mathbb{E}[S(t)]}{N-\mathbb{E}[S(t+\delta)]}\right)\end{aligned}$$

が成り立つ．さらに，$\bar{\alpha}$，$\bar{N}$ の定義より

$$\begin{aligned}&(\mathbb{E}[S(t+\delta)]-\mathbb{E}[S(t)])-(\mathbb{E}[S(t)]-\mathbb{E}[S(t-\delta]])\\&\leq\delta\{\ln(1-p\bar{\alpha})\}(\bar{N}-\mathbb{E}[S(t)])(\mathbb{E}[S(t)]-\mathbb{E}[S(t-\delta)])\\&\leq(\mathbb{E}[S(t+\delta)]-\mathbb{E}[S(t)])-(\mathbb{E}[S(t)]-\mathbb{E}[S(t-\delta])\\&\quad+\frac{(\mathbb{E}[S(t+\delta)]-\mathbb{E}[S(t)])^2}{\bar{N}-\frac{1}{\delta\ln(1-p\bar{\alpha})}-\mathbb{E}[S(t+\delta)]}\end{aligned}$$

が成り立つ．このとき，上式の全体を δ^2 で割り，$\delta\to 0$ とすれば良い．

参考文献

[1] S.Staniford, V.Paxson, N.Weaver, "How to 0wn the Internet in Your Spare Time", Proceedings of the 11th USENIX Security Symposium (Security '02), 2002.

[2] 佐藤, 内田, 石橋, 小林, "メール型コンピュータウィルス拡散過程モデルの提案", OR 学会春季研究発表会予稿集, 2004.

[3] 内田, 佐藤, "コンピュータウィルス拡散過程のモデル化に関する検討", 信学総大予稿集, 2004.

[4] B.Bollobás, "Random Graphs", Cambridge University Press, 2001.

[5] R. M. Corless, et al., "On the Lambert W Function", Advances in Computatianal Mathematics, vol.5, 1996.