コンピュータウィルスの生態学

豊泉　洋

コンピュータウィルスのモデリング手法に対する解説と評価を行う．疫学的モデルから，Lotka-Volterra 方程式，出生死滅過程，確率微分方程式，さらには，スケールフリーネットワーク上での percolation モデルなどの最新の確率的モデリング手法を，実際のコンピュータウィルスを紹介しながら解説し，コンピュータウィルスの数理的な生態モデルがネットワークセキュリティ上，重要であることを示す．

キーワード：コンピュータウィルス，情報セキュリティ，微分方程式，スケールフリーネットワーク

1. はじめに

近年，ネットワークセキュリティという言葉が，日常のニュースにも登場するようになった．不正アクセス，DoS 攻撃，スパムメール，スパイウェア，ボット，フィッシングメールなど，手を変え品を変え，ネットワークの快適性・安全性・信頼性を脅かす新たな技法がアンダーグラウンドな世界から編み出されている．その中でも，特に，コンピュータウィルスはネットワークを利用するユーザー，そしてネットワークの管理者に共通する悩みの種である．

しかし，我々ネットワーク管理者やユーザーにも，日々現れる新種のコンピュータウィルスに対して，対応策がある．クライアントマシンにアンチウィルスソフトを導入する，ネットワーク上でコンピュータウィルスをフィルタリングするといった物理的な対策から，「あやしげな添付ファイルは開かないようにしましょう」といった啓蒙的な対策まで，さまざまな対策が提案，実行されている．

しかし実態として，これらの対策にもかかわらず，コンピュータウィルスの被害はなくなっていない．最近でも，日本で No. 1 の市場占有率を誇る P2P ファイル共有ソフトである Winny を介して広がる情報暴露型のコンピュータウィルス Antinny による情報漏洩の被害が深刻な社会問題を引き起こしたのは記憶に新しい．コンピュータウィルスに限らず，ネットワークセキュリティのための対策には，100%はあり得ない．コンピュータウィルスの特性を知り，対策のコスト，効果，利便性への影響を評価することは重要である．コンピュータウィルスが，どのような経路で，どのようにネットワーク上を拡散していくのか？　どのような防御手法が，どのような段階で有効なのか？　このような問いに答えるためには，コンピュータウィルスの数理モデルを作らなければならない．この論文では，コンピュータウィルスの既存のモデリング手法に対する解説を行い，今後のコンピュータウィルスのモデリングの研究に方向性を示す．

2. コンピュータが感染する病気

最初で，かつ最も基本的なコンピュータウィルスのモデルは，1990 年代初頭に，Kephar, White[11]-[14]らの IBM グループによって研究された疫学的数理モデルであろう（一般の疫学数理モデルについては，例えば，巌佐[10]参照）．コンピュータウィルスを感染性の病気と捉え，その病気が集団の中でどのように広まっていくかを表す数理モデルをコンピュータウィルスに応用したものである．

コンピュータウィルスに感染可能な K 台のコンピュータのうち，ある時刻 t において，$I(t)$ 台が既に感染しているとする．我々の興味は，将来どのようにコンピュータウィルスという感染性の病気がネットワーク内に広まっていくかである．感染しているコンピュータのうち特定の一台に注目しよう．このコンピュータが，次の感染先として選んだコンピュータが未感染（susceptible）であれば，感染することでコンピュータウィルスの数は増える．しかし，選んだ先が既に感染している場合（infected）には，コンピュータウィルスの数は増えない．したがって，実効的な感染

とよいずみ　ひろし
早稲田大学 大学院会計研究科
〒169-8050 新宿区西早稲田 1-6-1

レートは，$1-I(t)/K$ だけ割り引かれる．また，ワクチンソフトなどの効果により，感染しているコンピュータが未感染の状態に復帰するとしよう．

このような数理モデルは疫学モデルと呼ばれる．このモデルでは，各コンピュータが感染可能（susceptible），感染（infected）の二つの状態をもつ．その可能な状態遷移により，SIS（susceptible-infected-susceptible）モデルといわれる．

コンピュータウィルス数 $I(t)$ に対して次のように，SIS モデルの微分方程式を次のように考えることができる．

$$\frac{dI(t)}{dt}=\lambda I(t)\left(1-\frac{I(t)}{K}\right)-\mu I(t), \tag{1}$$

$$I(0)=I_0. \tag{2}$$

ここで，λ は感染レート，μ は治癒レートを表す定数である．この微分方程式が，次のようなロジスティック曲線を解にもつことは，すぐわかる．

$$I(t)=\frac{I_0\left(K-\frac{\lambda}{\mu}\right)}{I_0+\left(K-I_0-\frac{\mu}{\lambda}\right)e^{-(\lambda-\mu)t}}. \tag{3}$$

この解を使えば，感染レート λ，治癒レート μ，それに感染可能なコンピュータの台数 K がわかれば，将来のコンピュータウィルスの繁殖は，確定的に予測できる．後付けではあるが，2001 年に発生したポートスキャン型のウィルス Code Red の繁殖の様子を理論値と比較すると，よく一致しているのがわかる（図 1）[24]．

この解(3)からすぐわかるのは，感染レート λ と治癒レート μ の大きさによって，ウィルスの広がる様子は大きく変化するということだ．$\lambda<\mu$ ならば，$I(t)$ は 0 に収束し，コンピュータウィルスの感染爆発は起こらない．ところが，感染レートが治癒レートを上回れば，感染が拡大する．感染拡大を防止する治癒レートを見積もり，コンピュータウィルスを押さえ込むためには，感染レート λ の見積もりが重要である．また，感染レート λ を治癒レート μ 以下に押さえるような対策が大事であることがわかる．

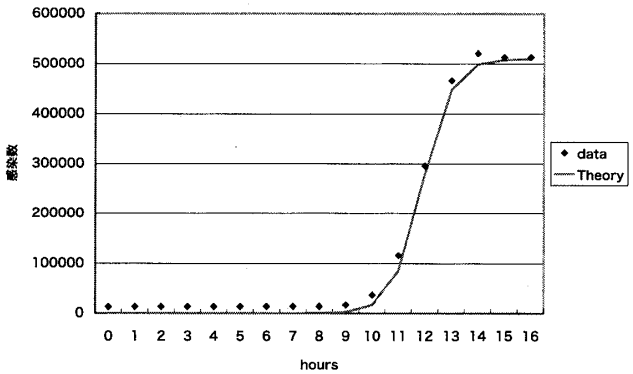

図 1　Code Red の繁殖とロジスティック曲線．Staniford [24]より引用

3. 正義のコンピュータウィルス？

2001 年，ポートスキャン型のウィルス Code Red が発生した．それまでのメール型のウィルスと異なり，その圧倒的な感染速度により，発生してから 14 時間で全世界に蔓延，36 万台が感染したといわれる[19]．このような早さで，蔓延するコンピュータウィルスには，通常のアンチウィルスソフトによる防御法では，追いつかないという問題意識から，Toyoizumi and Kara[27]は，コンピュータウィルスと同様な感染方法ももち，悪いコンピュータウィルスを排除する正義のコンピュータウィルス Predator の可能性を模索していた．

Predator は，次のような微分方程式に従って繁殖するように設計される．

$$\frac{dI(t)}{dt}=\lambda I(t)-aI(t)I_P(t), \tag{4}$$

$$\frac{dI_P(t)}{dt}=bI(t)I_P(t). \tag{5}$$

ここで，$I(t)$ は悪いコンピュータウィルス，$I_P(t)$ は predator の数を表す．Predator は，悪いコンピュータウィルスを発見すると，これを捕食する．無制限な繁殖を抑えるため，悪いコンピュータウィルスを発見したときにだけ，繁殖を行う．

$y=I_P(t)$ と $x=I(t)$ の間には，次のような関係が成立する[27, Theorem 1]．

$$x=\frac{1}{b}\{\lambda\log(y/y_0)-a(y-y_0)\}+x_0. \tag{6}$$

図 2 で，$y=I_P(t)$ と $x=I_B(t)$ の典型的な軌道がわかる．当初，悪いコンピュータウィルスはさかんに繁殖するが，途中から，Predator が増えすぎた悪いコンピュータウィルスを捕食し，繁殖を開始する．悪いコンピュータウィルスは次第に数を減らし，最後には消滅する．Predator は，悪いコンピュータウィルスが消滅したことにより，$dI_n(t)/dt=0$ となり，繁殖を自動的に停止する．

その後，Predator のようなアクティブなディフェンスのモデルは，実現可能性，有効性，社会的問題点などを中心に，勢力的に研究がなされている[2][9][15][16][23]．

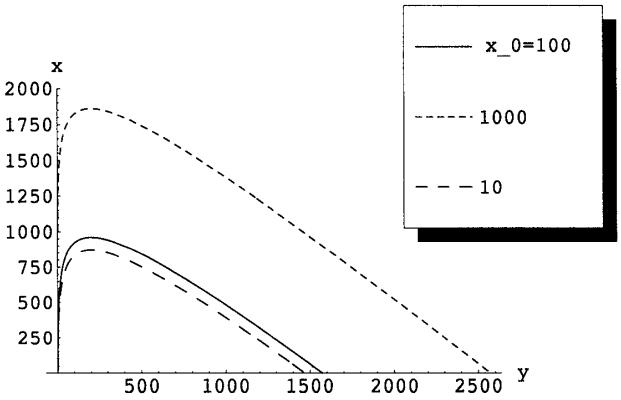

図2　初期ウィルス数を変化させた場合の軌道の変化．ただし，$y_0=1$, $r=2$, $a=1/100$, $b=1/100$.

4．正義の現実

2003年，強い感染力をもつポートスキャン型ウィルスBlasterが発生し，それを追いかけるようにNachiという同種のコンピュータウィルスが発生した．Nachiは，Blasterを除去し，Blasterが侵入したコンピュータのセキュリティホールを塞ぐという機能をもった正義のコンピュータウィルスであった．しかし，実際には，Blasterを遥かに凌ぐ感染速度をもっていたため，Nachiの侵入を許したネットワークは，そのリソースを食い尽くされ，ダウンするという運命だった[28][1]．

BlasterとNachiのような関係にあるコンピュータウィルスもLotka-Volterra捕食系としてモデル化できる．時刻tで，$I_b(t)$のBlaster，$I_n(t)$のNachiが存在していたとすると，次のようになる．

$$\frac{dI_b(t)}{dt}=\lambda I_b(t)-aI_b(t)I_n(t), \tag{7}$$

$$\frac{dI_n(t)}{dt}=\lambda_n I_n(t). \tag{8}$$

Blasterは，Nachiに見つかると補食される．その一方で，Nachiは，Blasterの存在とは独立に，λ_nの速度で自由に繁殖ができる．Predatorの繁殖を記述する(4)と比較すれば，その差は明らかだ．詳しい解析をするまでもなく，Nachiは正義のウィルスであっても，ネットワーク上に蔓延し，Blasterを凌ぐような危険な存在になりうることがわかる．

[1] 時系列的にみれば，2002年のToyoizumi and Karaの論文が2003年のNachiの出現と関わりがある可能性もあるが，その可能性は少ないであろう．実際，もしウィルスの作者が筆者らの論文を目にしていれば，もう少しましな繁殖戦略を考えたことであろう．

5．スケールフリーネットワーク上のコンピュータウィルス

コンピュータウィルスの中には，ユーザーがコンピュータ上に残した他人のメールアドレスを頼りに感染ルートを探すものが多い．このようなメール型のコンピュータウィルスは，メールアドレスのつながりによってできるメールアドレスネットワークの中を拡散していくと考えモデル化することができる[12][17][21]．今までのモデルのように，自由に次の感染先を探せるというわけではないのである．

ここでは，Newman[21]やToyoizumi and Kaiwa[26]の議論に従い，一般のネットワーク上でのコンピュータウィルスの繁殖の様子を分析しよう．個人のメールアドレスをノードとして考え，同一のマシン上に保存されているメールアドレスにはリンクが張られると考え，メールアドレスの空間をネットワーク化する．各ノードのリンク数は，それぞれ独立で，同一な確率分布に従うと仮定する．ネットワーク上でランダムに選ばれたノードのリンク数をLとし，ランダムに選ばれたリンクの先につながっているノードのリンク数をL_eとする．リンクの多いノードは，そのリンク数に比例して選ばれやすいため，L_eの分布はLの分布を用い，次式で与えられる（図3）．

$$P\{L_e=k\}=\frac{kP\{L=K\}}{E[L]}. \tag{9}$$

特にその期待値は，

$$E[L_e]=\frac{E[L^2]}{E[L]} \tag{10}$$

となる．ベキ分布のような裾の長い分布では，$E[L_e]\geq E[L]$となり，裾が長くなるに従い，$E[L_e]$は大きくなることが知られている．

あるノードがコンピュータウィルスに汚染された場合に，リンクごとに確率pでリンク先にあるノードが感染すると仮定し，ランダムに選んだノードから感染が始まった場合のoutbreakのサイズ（最終的な感染ノード数）をSとし，その大きさを評価する．この問題は，個体物理学や統計物理学では，液体が物質の中に浸透していく様子を分析する，percolation問題として知られている[25]．

ここで，ランダムにリンクを選んだ場合のoutbreakのサイズをS_eとする．ネットワークのサイズが十分大きく，感染のループができないと仮定すると，SとS_eの関係は次のように表すことができる．

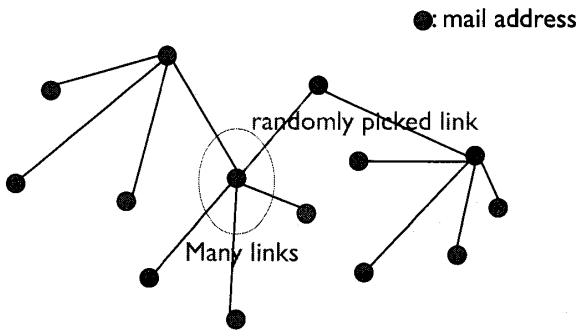

図3　メールネットワーク：リンクをランダムに選び，その先に繋がっているノードを選ぶ．すると，リンクをたくさんもつノードは，選ばれやすい．

$$S=1+\sum_{m=0}^{M} S_{e,m}. \tag{11}$$

ただし，M は最初に選んだノードの汚染されたリンク数で，$S_{e,m}$ は S_e と同一で独立な分布である．一方，最初に選んだリンクに接続されたノードの感染リンク数が M_e の場合には，次のような再帰式が得られる．

$$S_e=1+\sum_{m=0}^{M_e-1} S_{e,m}. \tag{12}$$

(11)と(12)の両辺の期待値をとって，Wald の方程式（例えば[22]）を使い，適当に整理すると，outbreak のサイズの期待値が次のように与えられる．

$$\begin{aligned} E[S]&=1+E[M]E[S_e]\\ &=1+\frac{E[M]}{2-E[M_e]}\\ &=1+\frac{pE[L]}{2-pE[L_e]}. \end{aligned} \tag{13}$$

(10)に注意すれば，(13)より $p>2E[L]/E[L^2]$ のときに，outbreak サイズの期待値 $E[S]$ は発散することがわかる．すなわち，感染爆発が起こる．

それでは，リンク数 L はどのような分布に従うと考えるのが妥当であろうか？　その答えがスケールフリーネットワークである．1999 年の Barbasi らの研究[6]をきっかけに，さまざまなネットワークがスケールフリーという特性をもつことが指摘されている[1][5]．スケールフリーネットワークとは，ノード数が増え，進化するネットワークであり，各ノードのリンク数がベキ法則分布に従うようなネットワークである．近年インターネット，感染症の伝染，遺伝子のネットワークなどが，スケールフリーネットワークに従うことが実証され，盛んに研究がなされている．

Albert and Barabasi[1, p. 72]を参考に，リンク L の分布がべき分布に従うことを（数学的には厳密ではないが，直感的に）示す．ネットワークは，新しいノードやリンクを付け加えて「進化」する．今，あたらしいノード（メールアドレス）が追加されたとしよう．このノードは，既存のノードと新しいリンクを結ぶ．リンクを結ぶ先は，ランダムに選ばれるのではない．すでに，リンクをたくさんもつ人気者ほど，選ばれる確率が高い．すなわち，結ぶ先は，既存のノードがもつリンク数に比例した形で決定される．単位時間あたりに付け加わるリンクの数を m とする．時刻 t で，ネットワークは，mt のリンクをもつ．各リンクは二つのノードをもつことに注意すれば，$L(t)$ 本のリンクをもつノードがあらたに得るリンクは，$m(L(t)/2mt)$ となる．したがって，リンク数の時間発展は，以下の式で表すことができる．

$$\frac{dL(t)}{dt}=\frac{L(t)}{2t}. \tag{14}$$

このノードがネットワークに加わった時刻を T_0 とすると，

$$L(T_0)=m, \tag{15}$$

という初期条件が得られる．この初期条件の元で，(14)を解くと，$L(t)$ の時間発展は以下のようになる．

$$L(t)=m\left(\frac{t}{T_0}\right)^{1/2}. \tag{16}$$

これを使うと，

$$P\{L(t)\geq k\}=P\left\{T_0\leq\frac{m^2}{k^2}t\right\}. \tag{17}$$

ここで，新規ノードがネットワークに加わった時間 T_0 が $(0, t)$ でランダムだとすると，$P\{T_0\leq x\}=x/t$ なので，$k\geq m$ に対して，

$$P\{L(t)\geq k\}=m^2k^{-2}, \tag{18}$$

となり，$L(t)$ がいわゆる指数 2 の Pareto 分布に従うことがわかる．大学内でのメールネットワークを調べると，実際にそのリンク数分布は，ベキ分布（指数 1.81）に従うという Ebel らの研究結果もある[8]．

Pareto 分布では，その指数が 2 以下の場合には，$E[L^2]$ は発散し，感染サイズ $E[S]$ も発散する．したがって，スケールフリーネットワークでは，感染率の低いコンピュータウィルスであっても，感染拡大しやすく，非常に危険であることがわかる．

スケールフリーネットワークおよびその上でのコンピュータウィルスの繁殖については，物理学者が参入し，盛んに研究が行われている[4][17][18]．

6.　利己的なコンピュータウィルス

ネットワークをより快適に生きるために，我々は，現在研究されているコンピュータウィルスのみの数理

モデルから，より包括的な数理モデルをつくり，その研究をする必要がある．

AndersonとTaylor[3]のサーベイ論文に詳しいが，ゲーム理論やミクロ経済学的な観点をネットワークセキュリティに当てはめることで，よりセキュアで快適なネットワークを作るために，ネットワークを構成する各プレイヤー（闇世界の住人，ネットワーク管理者，ソフトウェア開発者，ユーザーなど）にインセンティブを与えるルールや社会組織を作ることができるかという新しい問題意識でも研究が始まっている．

また，コンピュータウィルスの進化という観点でも研究をしはじめる時期かもしれない．Nachenberg[20]は，共進化という観念をコンピュータウィルスにもち込んでいる．もちろん，文字通りの意味で，コンピュータウィルスは，生きていて，生物のように進化すると考える人はいないだろう．しかし，ある意味で，その生態は，Dawkins[7]の言う利己的な遺伝子と同じように考えることができるかもしれない．利己的な遺伝子は，宿主の身体を乗り物にし，自分の複製を増やす．Dawkins流に言えば，コンピュータウィルスは，その複製の目的のためにコンピュータウィルスの製作者を乗り物として使っているのだ．コンピュータウィルスの作者は，一度じっくり考えてみた方がよい．どちらが，コントロールしているのか？　自分がコンピュータウィルスを操っているのか，それともコンピュータウィルスに自分が操られているのか….

参考文献

[1] R. Albert and A. Barabasi: Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 (1): 47-97, 2002.

[2] K. G. Anagnostakis, M. B. Greenwald, S. Ioannidis, A. D. Keromytis and D. Li: A cooperative immunization system for an untrusting internet. In *Proceedings of the 11th IEEE International Conference on Networks (ICON '03)*. IEEE, October 2003.

[3] R. Anderson and T. Moore: The economics of information security. *Science*, 314 (5799): 610-613, 2006.

[4] J. Balthrop, S. Forrest, M. E. J. Newman and M. M. Williamson: Computer science: Technological networks and the spread of computer viruses. *Science*, 304 (5670): 527-529, 2004.

[5] A.-L. Barabasi: *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Plume, 2003.

[6] A.-L. Barabasi and R. Albert: Emergence of scaling in random networks. *Science*, 286 (5439): 509-512, 1999.

[7] R. Dawkins: *The Selfish Gene: 30th Anniversary Edition-with a new Introduction by the Author*. Oxford University Press, USA, 3 edition, 5, 2006.

[8] H. Ebel, L.-I. Mielsch and S. Bornholdt: Scale-free topology of e-mail networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 66 (3): 035103, 2002.

[9] A. Gupta and D. C. DuVarney: Using predators to combat worms and viruses: a simulation-based study. In *Computer Security Applications Conference, 2004. 20th Annual*, pages 116-125, 2004.

[10] Y. Iwasa: *Mathematical Biology*. Kyouritsu, 1999.

[11] J. O. Kephart: A biologically inspired immune system for computers. In *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 130-139, Cambridge, MA, US, 1994. MIT Press.

[12] J. O. Kephart and S. R. White: Directed-graph epidemiological models of computer viruses. *1991 IEEE Symposium on Security and Privacy*, 00: 343, 1991.

[13] J. O. Kephart and S. R. White: Measuring and modeling computer virus prevalence. In *1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 2-15, 1993.

[14] J. O. Kephart, S. R. White and D. M. Chess: Computers and epidemiology. *IEEE Spectrum*, pages 20-26, MAY 1993.

[15] H. Kim and I. Kang: On the functional validity of the worm-killing worm. In *2004 IEEE International Conference on Communications*, volume 4, pages 1902-1906, 2004.

[16] M. Liljenstam and D. M. Nicol: Comparing passive and active worm defenses. pages 18-27, 2004.

[17] S. F. M. E. J. Newman and J. Balthrop: Email networks and the spread of computer viruses. *Phys. Rev.*, E 66: 035101, 2002.

[18] C. Moore: Exact solution of site and bond percolation on small-world networks. *Physical Review E*, 62: 7059. 7065, 2000.

[19] D. Moore: The spread of the code-red worm (CRv 2). http://www.caida.org/analysis/security/code-red/coderedv2_analysis.xml, July 2001.

[20] C. Nachenberg : Computer virus-antivirus coevolution. *Commun. ACM*, 40 (1) : 46-51, 1997.
[21] M. E. J. Newman : The spread of epidemic disease on networks. *Physical Review E*, 66 : 016128, 2002.
[22] S. M. Ross : *Applied Probability Models With Optimization Applications*. Dover Pubns, 1992.
[23] S. Sidiroglou and A. D. Keromytis : Countering network worms through automatic patch generation. *IEEE Security and Privacy*, 03 (6) : 41. 49, 2005.
[24] S. Staniford : Analysis of spread of July infestation of the code red worm. http://www.silicondefense.com/cr/
[25] D. Stauffer and A. Aharony : *Introduction to Percolation Theory*. Taylor and Francis, 2 edition, 7 1994.
[26] H. Toyoizumi and K. Kaiwa : Obeservation and modeling method of dynamics of computer virus spread. *IEICE 3rd QoS Workshop*, 2005.
[27] H. Toyoizumi and A. Kara : Predators : Good will mobile codes combat against computer viruses. *New Security Paradigm Workshop 2002*, pages 11-17, 2002.
[28] C. L. Webb : Worm vs. worm. Washington Post, August 19, 2003.