

ネットワーク成長によるメール型ウイルスの再流行と 重点的なハブの免疫化の効果

林 幸雄*, 箕浦 正人*, 松久保 潤*

概要: 電子メールの送受信数の実測分布に基づく接触関係を考えた, コンピュータウイルスの拡散モデルを提案する. 特に, 開放系の本モデルは, 感染の鎮静化後の再流行現象がネットワーク成長を伴う新規ユーザの加入によって起ることを示唆する. 経験上典型的なこの現象は, 従来モデルでは説明できない. さらに, 重点的なハブ頂点のアンチウイルスによる免疫化の効果を明らかにする.

Recovery Prevalences of Viruses via E-mails in a Growing Computer Network and Effectiveness of the Prior Immunization for Hubs

Yukio Hayashi*, Masato Minoura*, Jun Matsukubo*

Abstract: We propose a model of computer viruses spreading via electronic mails, based on the contacts in real distributions of the sent and receive mails. In particular, this model in an open system suggests that a growing computer network with new mail users causes the recovery phenomenon of prevalences from a low level of disease, which is typical in experiences but not explained by the conventional models of computer viruses. Moreover, we verify the effectiveness of the prior immunization by anti-viruses for hubs in the scale-free network.

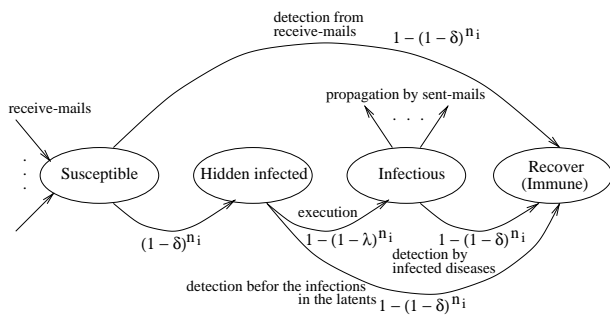


図 1: S-H-I-R 状態遷移図.

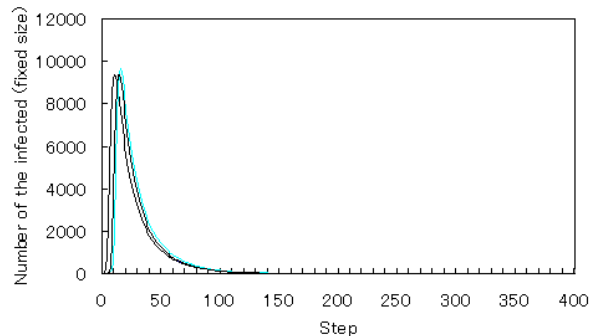


図 2: 固定規模における感染数の変化.

* 北陸先端科学技術大学院大学・Japan Advanced Institute of Science and Technology.

1 はじめに

コンピュータウィルスの被害の中で、電子メールによるものは70-80%以上で[3]、その損害も大きい[12]。そこで、個々の駆除の為のアンチウィルスの開発のみならず、その効果的なネットワーク上の配置から拡散伝搬による被害を抑えることが望ましい。

従来のコンピュータウィルスの拡散モデルでは、解析に適した一様ランダムや完全グラフ[11]、一次元格子などのネットワーク構造が扱われ、電子メールによるウィルス感染の接触機会の偏りが指摘されている[7]ものの、実際の送受信数に基づいたネットワーク構造や、アンチウィルスの配置方法は明らかでなかった。一方、社会的な知人関係などを表すSF(Scale-Free)ネットワーク構造[1]上のSIS(Susceptible-Infected-Susceptible)モデルでは、絶滅のしきい値が存在しないこと[9]や、重点的なハブの免疫化によって拡散が抑えられることが、平均場解析から最近示された[2][10]。しかしながら、これらのモデルではネットワーク規模が固定され(あるいは $N \rightarrow \infty$)、ウィルスの蔓延か絶滅の定常状態しか議論されておらず、観測データに見られる再流行現象[4]を説明できなかった。

本研究では、電子メールによる拡散モデル[7]を基に、実際の送受信関係[6]がべき乗分布に従うSFネットワーク構造となることを示し、その現実的な接触関係上の拡散伝搬の基本特性を明らかにする。また、新規ユーザによるネットワーク成長によって、再流行現象が引き起こされることを数理モデルから示唆する。さらに、重点的なハブの免疫化によって、再流行する可能性があるウィルスの拡散伝搬も効果的に抑えられることを示す。

2 電子メールによるウィルス感染

まず、電子メールの送受信関係を表すネットワークを考える。その頂点 $i = 1, \dots, N$ は各PCを、有向辺はアドレス帳などに登録されたPC間のメールの送受信関係を表す。

個々のPCの状態は図1のように遷移する[7]。ここで、 λ は潜伏状態Hから感染状態Iへの実行率、 δ は感染の発見率とする。また、頂点 i におけるPCが保持する n_i 個のウィルス(感染メール)から、どれか1つでもウィルスが発見される確率は $1 - (1 - \delta)^{n_i}$ 、 n_i 個の潜伏状態からどれか1つでも増殖状態に移行

する確率は $1 - (1 - \lambda)^{n_i}$ となる。これらの状態遷移は確率的に行われるが、時間の経過に伴って各頂点はいずれは免疫状態Rとなる点に注意しよう。

次に、実測データに基づいて、電子メールの送受信関係を表すネットワーク構造を考える。

通信総研と東大社会情報研による'00年10~11月の調査[6]では、一日の送受信数の最頻度は1~10通、平均送信数は16.3通、平均受信数は27.7通で、送受信数がある偏った分布となることが報告されている。これらの送受信数は、べき乗則(次数 k を持つ頂点数の分布が $k^{-\gamma}$ に比例)に従う。これより、送受信数の分布のべき係数はそれぞれ $\gamma_{out} = 2.5$ 、 $\gamma_{in} = 1.9$ 、平均送受信数 \bar{k} は5~20通程度と推定される。

べき乗則に従うSFネットワーク構造は、インターネットのルータ、WWW、電力網、知人関係など現実の全てのネットワークに共通する普遍特徴として極めて重要である[1]。本研究では、SFネットワーク構造上のウィルスの拡散伝搬のシミュレーションを行うため、平均場近似による解析向きモデル[1]よりも、構成的な (α, β) モデル[5][8]に着目し、上記の実測値に適合したネットワークを具体的に生成した。

3 再流行をも抑えるハブの免疫化

前章で述べた実測値に従って生成したSFネットワーク上で、図1の状態遷移に基づくウィルスの拡散伝搬の特性をシミュレーション解析した。図2は、ネットワーク規模(総頂点数)を $N = 18934$ に固定した場合の、典型的な感染数(状態H+Iの数)の挙動と、 (α, β) モデルによって確率的に生成されたネットワークおよび状態遷移による違いを示す。ここで、実行率 $\lambda = 0.1$ 、発見率 $\delta = 0.04$ 、平均辺数 $\bar{k} = 6.6$ とし、感染源Iをランダムに選んだ5個の頂点とした。

一方、図3と4はそれぞれ、初期規模 $N = 400$ で構成したSFネットワークから、ウィルスの拡散伝搬を伴う確率的な状態遷移の各ステップごとに、

線形成長: +50個ずつの頂点と対応する辺を追加、

指数成長: $\times 1\%$ ずつの頂点と対応する辺を追加、

しながら成長していく開放系のモデルの典型的な挙動(再流行現象)を示す。

次に、重点的なハブの免疫化の効果を図5に示す。グラフの上から順に記号が菱形:ウィルス発見のみによる駆除、四角:ネットワーク成長に追従した頂点の

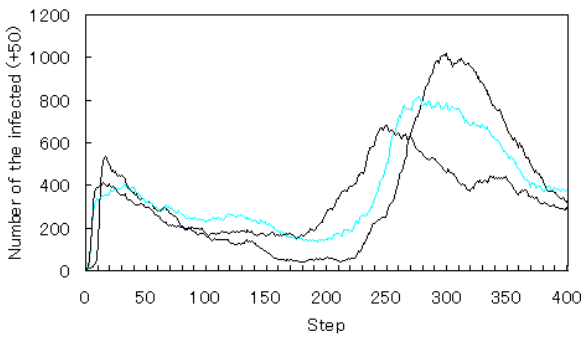


図 3: 線形成長における感染数の変化.

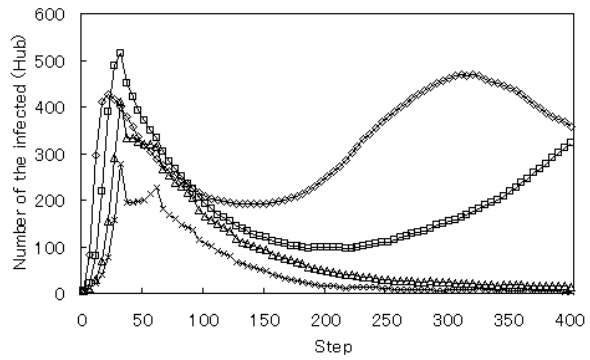


図 5: 線形成長におけるハブ免疫化.

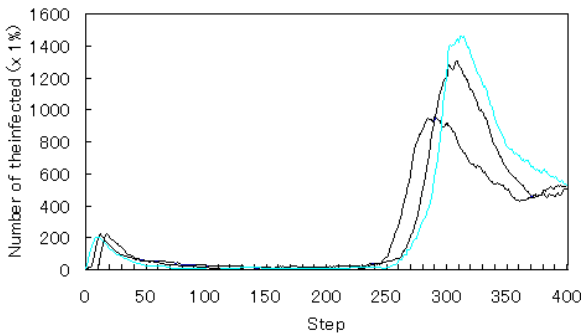


図 4: 指数成長における感染数の変化.

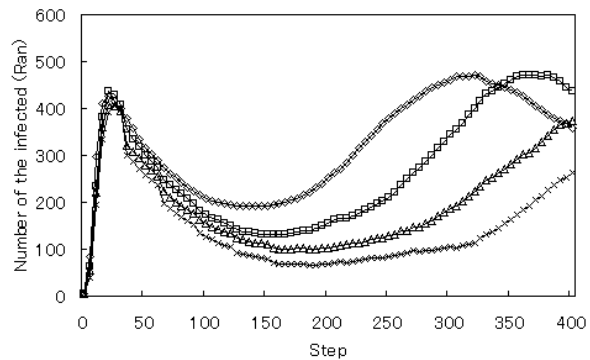


図 6: 線形成長におけるランダム免疫化.

免疫率 10%, 三角: 20%, x: 30%に対する, 100 回の試行の平均感染数 (状態 H+I の数) の変化をそれぞれ表す. 但し, ウィルス発見によって免疫状態になった頂点とは別に, 30 ステップ毎に成長した規模に対する比率で, 出次数が大きい (ハブ) 順にあるいはランダムに頂点を免疫化した. 図 5 では, 免疫率を上げるに従って感染数が減少し, 免疫率 30% (x 印) では絶滅させるまでに到った.

一方, 図 6 に示すランダムな頂点の免疫化では, 免疫率を上げるに従って感染数が若干減少するものの, 30%でも絶滅させることが出来てない (より急激に感染数が増加する点を除いて指数成長でも同様).

線形成長に対してハブを免疫化する場合で, 同様な 100 回の試行の平均値として, 未感染頂点数 (状態 S の数) と免疫頂点数 (状態 R の数) を図 7 と 8 にそれぞれ示す. 図 3 に示した感染数 (状態 H+I の数) と比べて, ネットワーク成長による未感染頂点数の増加が著しい (指数成長でも同様). これらの定量的な性質は, ネットワークの成長速度と, ウィルスの実行率や発見率 (感染数の増減の急激さ), 平均辺数 (拡

がり度) などに依存する.

さらに, 免疫率 30% の線形成長の場合で, 30 ステップ毎に免疫化された頂点数とウィルス発見数の 100 回試行の平均値を図 9 に示す. 免疫化されたハブの頂点数 (黒棒) がウィルス発見による状態 R の頂点数 (斜線付棒) よりも圧倒的に多く支配的で, 特にその傾向はステップ数が大きくなるほど顕著になっている (指数成長でも同様). これらの結果から, 感染数の激減や絶滅にはウィルス発見よりもハブの免疫化が効果的であることがより鮮明になる.

4 おわりに

本研究では, 実際の電子メールの送受信関係 [6] が SF ネットワーク構造 [1] となることを示し, その現実的な接触機会上の拡散伝搬の特性を明らかにした. 特に, 定常的な蔓延や絶滅ではなく, 従来モデルでは説明できない, 一旦は感染が鎮静化した後に再流行する現象 [4] を, (開放系の上記のモデルにおける)

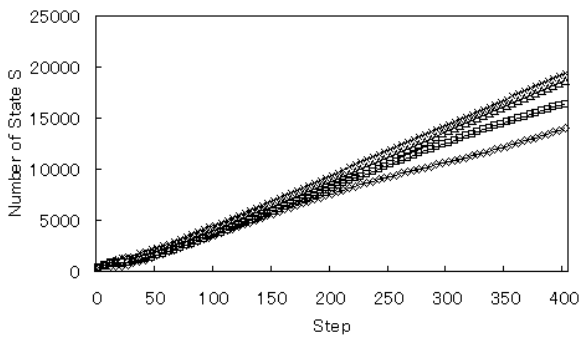


図 7: 未感染頂点数の変化 (線形成長, ハブ免疫化) .

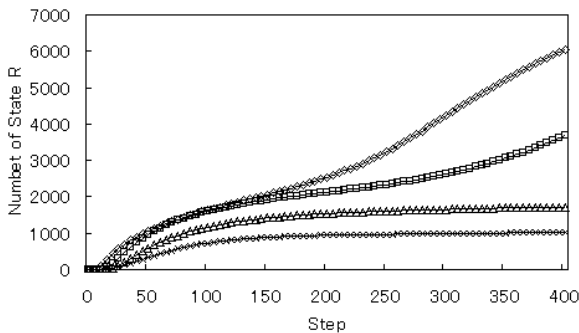


図 8: 免疫頂点数の変化 (線形成長, ハブ免疫化) .

新規ユーザによるネットワーク成長が引き起こすことを示唆した。さらに、この開放系のモデルにおいても、重点的なハブの免疫化によってウィルスの拡散が効果的に抑えられることを示した。

参考文献

- [1] R. Albert, and A.-L. Barabási. “Statistical Mechanics of Complex Networks,” arXiv:cond-mat/0106096v1, 2001.
- [2] Z. Dezső, and A.L. Barabási. “Halting viruses in scale-free networks,” *Physical Review E*, Vol. 65, 055103, 2002.
- [3] 情報処理振興事業協会. “国内におけるコンピュータウイルス被害状況調査,” 平成 13 年 2 月, 2001, 平成 14 年 3 月, 2002.
- [4] J.O. Kephart, and S.R. White. “Measuring and Modeling Computer Virus Prevalence,”

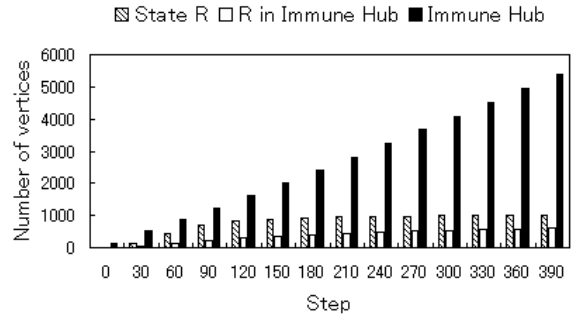


図 9: 免疫状態の頂点数 (ハブ免疫化 30%).

Proc. of the 1993 IEEE Comp. Soc. Symp. on Res. in Security and Privacy, pp. 2-15, 1993.

- [5] R. Kumar et al. “Extracting large-scale knowledge bases from the web,” *Proc. of the 25th VLDB Conf.*, pp.7-10, 1999.
- [6] 三上 俊治. “日本人とインターネット生活-ワールドインターネットプロジェクト-,” 2001. <http://sophy.asaka.toyo.ac.jp/users/mikami/info&media/>
- [7] 岡本 剛, 石田 好輝. “電子メールにより拡散するコンピュータウィルスの拡散モデルの解析,” *信学論 D-I*, Vol. J84-D-I, No. 5, pp. 474-482, 2001.
- [8] 大久保 和彦, 林 幸雄, 蛭川 繁. “Web 的ネットワークにおける情報伝搬率と速度,” *信学論 D-I*, Vol.J85-D-I, No.2, pp. 241-244, 2002.
- [9] R. Pastor-Satorras, and A. Vespignani. “Epidemic dynamics and epidemic states in complex networks,” *Physical Review E*, Vol. 63, 066117, 2001.
- [10] R. Pastor-Satorras, and A. Vespignani. “Immunization of complex networks,” *Physical Review E*, Vol. 65, 036104, 2002.
- [11] 千石 靖, 岡本 栄司, 満保 雅浩, 植松 友彦. “コンピュータウィルスの拡散と消滅の大域的振る舞いについて,” *情報論*, Vol. 37, No. 4, pp. 579-587, 1996.
- [12] US NEWS FLASH [Tech Web 特約] 2001 年のウィルス損害額は 8 月末までに推定 107 億ドル <http://itpro.nikkeibp.co.jp/free/ITPro/USNEWS/20010907/10/>