情報処理学会論文誌 Vol. 51 No. 3 682–690 (Mar. 2010)

# メーリングリストを考慮したマスメール型ワームの感染数理モデル

金岡 晃[†1] 勝野恭治[†2] 岡本栄司[†1]

電子メールを通じて感染するマスメール型ワームは数多い種類のワームの中で多くの割合を占め，さらに繁殖性や生存力の強さから社会的な問題になっている．現在ワームに焦点を当てた研究が多く行われているが，マスメール型ワームの感染の特徴であるメーリングリストを経由した感染を扱った研究はほとんど行われていないため，マスメール型ワームの感染プロセスの特徴がいまだ詳細に解析されていない．本論文では，メーリングリストの効果を包含したマスメール型ワームの感染数理モデルを提案する．提案モデルは，メーリングリストの効果に加えて，現実社会に沿った電子メールによる社会ネットワーク構造を反映することが可能である．さらに本論文では提案モデルに基づいてシミュレーションを行い，従来研究では分からなかった，メーリングリストの影響効果が感染初期で強く働いていることを解明することができた．

# Propagation Model for a Mass-mailing Worm with Mailing List

Akira Kanaoka,[†1] Yasuharu Katsuno[†2] and Eiji Okamoto[†1]

Mass-mail type worms have threatened to become a large problem for the Internet because of its amount and surviving property. Although many researchers have analyzed such worms, there are few studies that consider worm propagation via networks and mailing list. In this paper, we present a mass-mailing type worm propagation model including how the use of a mailing list could affect the propagation. We study the propagation through simulation with a model of an actual e-mail social network. We show that the impact of the mailing list on the mass-mail worm's propagation is significant, even if the mailing list is not large.

## 1. はじめに

ワームによる被害は年々増加し深刻な社会問題となっている．特に，電子メール（以下メール）を通じて感染をするワーム（本論文では以後，マスメール型ワームと呼ぶこととする）は数多くある種類のワームの中で多くの割合を占めているため[1)]，その対策が特に重要視されている．マスメール型ワームでは，メール利用者がファイル添付されたメールを受信し，それを実行することで感染する．その後ワームはコンピュータ内のメールアドレスを検索し，自分自身を添付してすべてのメールアドレスにメールを送信する．マスメール型ワームはその感染メカニズムから，メーリングリスト経由で爆発的に感染すると考えられているが，感染プロセスの特徴はいまだ詳細に解析されていない．

古くから研究が行われている疫学による感染数理モデルをワームに適用して，ワームの感染プロセスの特徴の解析を試みる研究[2)–4)]が注目されている．しかし，これらの研究ではワームやその感染経路となるネットワークの特性，特にメーリングリストの特性を反映できるものとなっていないため，マスメール型ワームに適応することが困難であった．

本論文では，疫学による感染数理モデルを拡張することでマスメール型ワームの感染メカニズムの数理表現を行ったマスメール型ワームの感染数理モデルを提案する．本モデルは，従来の感染数理モデルを用いたワーム遷移では実現されていなかった以下の 3 点，1）任意のネットワークトポロジの適用と，2）感受性状態から隔離状態への遷移，3）メーリングリスト効果の適用，を同時に実現するモデルである．さらに本論文では，提案モデルに基づいてシミュレーションを行い，その結果，従来研究では分からなかった，感染初期におけるメーリングリストの強い感染拡大効果を示した．

2 章では，本論文の背景と関連研究を紹介する．3 章では，本論文で対象とするマスメール型ワームの数理モデル化に必要な条件と，従来数理モデルの行列表現を示す．そして 4 章でマスメール型ワームのモデル化を提案し，5 章では提案モデルを利用したシミュレーションの結果を示す．最後に 6 章でまとめる．

†1 筑波大学
University of Tsukuba
†2 日本アイ・ビー・エム株式会社東京基礎研究所
IBM Research - Tokyo

## 2. 背景と関連研究

### 2.1 マスメール型ワーム

本章では本論文の背景として，マスメール型ワームの感染と拡大のメカニズムを解説し，いくつかの関連研究を紹介する．

マスメール型ワームに感染する場合，まず最初に利用者はファイルが添付されたメールを受け取る．この添付ファイルはマスメール型ワームを含んでいる．利用者が添付ファイルを実行すると，コンピュータは感染する．感染後は，ワームはメール閲覧ソフトウェアのアドレス帳やコンピュータ上のハードディスクドライブからメールアドレスを収集する．そして収集アドレスに対して自分自身を添付したメールを送信し拡大を図る．ワームの種類によっては，収集アドレスのドメイン情報と，存在する可能性の高いユーザ名を組み合わせて，宛先数を増幅させるものもある．また政府やセキュリティ関連会社のドメインを持つメールアドレスを宛先から排除することで，感染事実の発覚を回避するなど，様々な送信方法がとられる．

感染したコンピュータは，サービス妨害（Denial-of-Service: DoS）攻撃や，迷惑メール（スパムメール）の送信など様々な用途で利用者の知らない間に悪用されてしまう[5),6)]．

### 2.2 電子メールネットワーク

マスメール型ワームは，メールを通じて感染するため，メールユーザがいかにつながっているかというメールの社会ネットワークを知る必要がある．本論文ではその社会ネットワークをメールネットワークと呼ぶこととする．

近年の複雑ネットワーク分野での研究により，メールネットワークはある特徴を持つことが判明している．Ebel らはメールの送信元と宛先のデータを SMTP サーバログから抽出し，そのネットワークを解析した[7)]．メールアドレスをグラフにおけるノード，送信元と宛先の組をエッジとすると，そのグラフ（メールネットワーク）の次数分布がべき乗則に従うスケールフリーネットワークであることを示した．Newman らも同じくメールネットワークを解析している．彼らは大学のコンピュータシステム上のユーザが持つアドレス帳からメールネットワークを構築し，そのネットワークの次数分布が指数則に従ったものであることを解明した[8)]．

### 2.3 感染数理モデル

#### 2.3.1 SI/SIR モデル

感染数理モデルの基本形は Kermack らによって提案された[14)]．モデルは，感受性状態（S: Suceptible），感染状態（I: Infectious），隔離状態（R: Removed）の 3 つの状態が存在するため SIR モデルと呼ばれる．それぞれの状態にある人口数は以下の差分方程式で表される．

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

ここで感染率は $\beta$ で表され，隔離率は $\gamma$ で表されている．R の状態がない SI モデルは $\frac{dS}{dt} = -\beta SI$，$\frac{dI}{dt} = \beta SI$ で表される．ただし，これらのモデルは，構成される人がすべて直接つながっている完全ネットワークに基づいたものであり，ネットワークのトポロジを考慮したものではない．

#### 2.3.2 SIS モデル

感染数理モデルには，感染状態（I）から再び感受性状態（S）に戻る SIS モデルも存在する[15)]．

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \gamma I \\ \frac{dI}{dt} &= \beta SI - \gamma I \end{aligned} \tag{2}$$

### 2.4 スケールフリーネットワークでのワーム感染の解析

近年，スケールフリーネットワーク上でのワーム感染に関する研究がいくつか行われている．それら研究はマスメール型ワームを対象にしたものと，TCP/IP のサービス提供ソフトウェアの脆弱性などを通して直接感染するネットワーク型ワームを対象にしたものの 2 種類に大別される．

ネットワーク型ワームを対象にした研究では，インターネット上のルータが作り出すトポロジや AS（Autonomous System）どうしのつながりで表現されるトポロジといったスケールフリー性に着目した解析が行われている．Breisemeister らは感染数理モデルのトポロジによる影響を考慮し，ワームの感染がトポロジによりいかに変わるかを示した[9)]．Nikoloski らはペア近似（Pair Approximation）の手法を用いて従来の感染モデルを拡張し，スケールフリーネットワーク上でのワーム感染について研究している[11)]．Pastor-Satorras らは SIS

モデルを拡張し，ノードの次数に関連した遷移状態を表現した[10]．モデルの表現は，感染状態にある総数 $I$ の代わりに，次数 $k$ を持つノードの感染状態確率 $\rho_k(t)$ を用いた差分方程式が用いられている．

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \beta k \left(1 - \rho_k(t)\right) \Theta(\beta) \tag{3}$$

ここで $\Theta(\beta)$ は，与えられたリンクが感染ノードに接続している確率であり，次数 $k$ を持つノード数の割合 $P(k)$ を用いて以下のように表される．

$$\Theta(\beta) = \sum_k \frac{kP(k)\rho_k}{\sum_s sP(s)} \tag{4}$$

一方で，Zou らはマスメール側ワームの感染についての研究を行っており[12),13)]，スケールフリーネットワークとランダムグラフの差異に注目している．しかし，ここでは感染におけるメーリングリストの存在は考慮されていない．感染のモデルについては Pastor-Satorras らによるモデルを用いている[10]．感染状態から再び感受性状態に戻るというモデルは，すべての種類のワームに対して感染状態がどう変わるかを表現したものであり，特定のワーム感染状態がどう行われるかを示すのに適した方法ではない．さらに，このモデルでは感染拡大においてメーリングリストがどのように影響しているかの考慮はされていない．

これらの関連研究は SIS や SIR モデルなどを採用してモデル提案や解析がなされていることから，本論文でも Kermack らのモデルを拡張したモデルを提案する．

## 3. 行列による感染モデルの表現

### 3.1 モデル拡張要件

従来の感染数理モデルは感染者数の状態を差分方程式で表すものであるが，メーリングリストの効果やトポロジの影響を反映できないため，現状の感染モデルを直接適用することはできない．モデル化にあたり必要な条件を以下に示す．

#### 3.1.1 任意のネットワークトポロジ

メールネットワークはスケールフリーなどの性質を持つことが分かってきていることから，任意のネットワークトポロジの性質を反映可能とする感染モデルでなければならない．

#### 3.1.2 ウイルス対策ソフトウェアによる特殊状態遷移

初期の感染モデルでは，状態 S（Susceptible，感受可能状態），I（Infectious，感染状態），R（Removed，隔離状態）のそれぞれの遷移は，S から I，I から R とされている．しかし，ウイルス対策ソフトウェアによる隔離状態への移行は，感染状態と感受可能状態（非感染状態）を問わず行われるため，S から R の状態遷移も考える必要がある．

#### 3.1.3 感染におけるメーリングリストの効果

メールネットワークではメールアドレスはネットワーク上の 1 つのノードとして表現され，そこにはメーリングリストのアドレスも含まれる．しかし，メーリングリストノードは通常のメールアドレスとは違う動作を行う．メーリングリストは一種のメールメッセージの増幅器であり，その機能は配信されるメールがワームであるかどうかを問わないものである．そういったメーリングリストの特殊動作をモデルに反映する必要がある．

### 3.2 SI/SIR モデルの行列表現

2.3 節で示されたモデルは，3.1 節で示された条件を満足するものではない．特にメーリングリストを考慮するためには，モデル内におけるメーリングリストの扱いを明確にしなければならないが，そこには通常のメールアドレスとは異なる動作をすることによる状態遷移や感染拡大の差異の存在が予想される．

一方，これまでに紹介された差分方程式により表現された感染数理モデルは，各個体の種別差による遷移特徴などの差異を包含したものではなく，すべての個体が同種であることを前提にモデル化されている．そのため本研究で目指すメーリングリストと通常メールアドレスの 2 種類が存在する環境に対して直接適用することは困難である．そこで本論文では，まず感染数理モデルを差分方程式ではなく行列で表現することとした．

いくつかの表記と状態を表す行列の定義を以下に示す．

#### 3.2.1 表記の定義

$\mathbf{A} = \{a_{xy}\}$ : $M \times N$ 行列 $\mathbf{A}$ とその要素
$\mathbf{E}$ : 単位行列
$\mathbf{AB}$ : 行列積
$\mathbf{A} \cdot \mathbf{B}$ : 行列要素ごとの積

ここで，$F(\mathbf{A})$ は $\mathbf{A}$ の各要素の値が 2 以上になる場合，その要素を 1 にする関数である．

#### 3.2.2 状 態 行 列

各ノードの状態を示す行列を 2 つ用意する．$\mathbf{I}(t)$ は時刻 $t$ における各ノードの感染状態を示す $1 \times N$ 行列であり，要素 $i_x(t)$ が 1 であるときにノード $x$ が時刻 $t$ で感染していることを示し，0 であるときには感受性状態または隔離状態であることを示す．$\mathbf{R}(t)$ は時刻 $t$ における各ノードの隔離状態を示す $1 \times N$ 行列であり，要素 $r_x(t)$ が 1 であるときにノード $x$ が時刻 $t$ で隔離されていることを示し，0 であるときには感受性状態または感染状態であ

ることを示す．またここでは，$N$ はノード数を示している．

#### 3.2.3　ネットワークの表現

ネットワークは各ノード間のつながりを示す隣接行列で表現される．隣接行列 $\mathbf{G}$ は $N \times N$ 行列であり，各要素 $g_{xy}$ は，ノード $x$ と $y$ 間にエッジが存在する場合に 1 をとり，存在しない場合には 0 をとる．

#### 3.2.4　状態遷移の表現

##### 3.2.4.1　$\mathbf{R}(t)$ の状態遷移

$$\mathbf{R}(t+1) = F\left(\mathbf{R}(t) + \mathbf{I}(t) \cdot \mathbf{\Lambda}\right) \tag{5}$$

$$r_x(t+1) = f\left(r_x(t) + i_x(t)\lambda_x(t)\right) \tag{6}$$

ここで $\mathbf{\Lambda}$ は，確率 $\gamma$ で 1 となり確率 $(1-\gamma)$ で 0 となる $\lambda_x(t)$ からなる行列である．$\lambda_x(t)$ はノード $x$ の隔離状態への移行をランダムに決定するために用いられる．

##### 3.2.4.2　$\mathbf{I}(t)$ の状態遷移

$$\mathbf{I}(t+1) = F\left(\mathbf{I}(t)\left(\mathbf{E} + \mathbf{D}(t) \cdot \mathbf{G}\right)\right) \cdot \left(\mathbf{1} - \mathbf{R}(t)\right) \tag{7}$$

$$i_x(t+1) = f\left(\sum_y i_y(t)(e_{yx} + d_{yx}(t)g_{yx})\right)(1 - r_x(t)) \tag{8}$$

ここで $\mathbf{D}$ は，確率 $\beta$ で 1 となり確率 $(1-\beta)$ で 0 となる $d_{xy}(t)$ からなる行列である．$d_{xy}(t)$ はリンク $(x, y)$ を通じてワームが届けられた際に，ノード $y$ が感染するかどうかをランダムに決定するために用いられる．

#### 3.2.5　時刻 t の単位

一般にネットワーク型ワームは感染速度が速く，数秒や数分単位で拡大をするものもある．一方で，マスメール型ワームは，そのメカニズムから感染拡大には人間の行動が必要であり，メール受信からメール開封，そしてメール添付ファイルの実行に至るまでの時間幅は利用者により異なり数分から数時間の幅で推移がされるものと考えられる．これらのことからワームの感染拡大における推移は，時刻の粒度がワーム種別によって大きく異なることが分かる．

しかし，本論文は 3.1 節に示すように従来のモデルでは同時には表現できなかった 3 つの点を満たすマスメール型ワームの感染拡大モデルを提案するものであり，時刻 t の粒度は本論文の対象外とする．現実の感染速度はモデル内の時間粒度の差により異なることに留意されたい．

### 3.3　行列表現モデルの差分方程式への帰着

本節では，前節までで定義をした行列表現が SIR モデルに帰着可能であることを示す．

まず $S(t)$ を時刻 $t$ において状態 S にあるノードの数とし，同様に $I(t)$，$R(t)$ を状態 I, R にあるノードの数とする．時刻 $t$ と $t+1$ での各ノード数の関係は以下のように表現することができる．

$$\begin{aligned} S(t+1) &= S(t) - S(t)p(S \to I) \\ I(t+1) &= I(t) + S(t)p(S \to I) - I(t)p(I \to R) \\ R(t+1) &= R(t) + I(t)p(I \to R) \end{aligned} \tag{9}$$

ここで $p(S \to I)$ は状態 S から状態 I への遷移が起こる確率であり，各状態行列の要素は $i_x(t) = 0$，$r_x(t) = 0$，$i_x(t+1) = 1$，$r_x(t+1) = 0$ であることが求められる．しかし，SIR モデルでは S から R への遷移は起こらないため，$r_x(t+1) = 0$ となる確率は考えなくてよい．よって状態 S から状態 I への移行確率は以下で表される．

$$\begin{aligned} p(S \to I) &= p(\exists y(i_y(t)d_{yx}(t)g_{yx} = 1)) \\ &= 1 - (1-\beta)^{\sum_y i_y(t)g_{yx}} \\ &= 1 - (1-\beta)^{m_x(t)} \end{aligned} \tag{10}$$

ここで，$m_x(t) = \sum_y i_y(t)g_{yx}$ である．しかし SIR モデルでは完全グラフを前提としていることから $g_{yx} = 1$ であるため，

$$m_x(t) = \sum_y i_y(t) = I(t)$$

となる．また，$\beta << 1$ である場合，式 (10) は以下のようになる．

$$p(S \to I) = \beta I(t) \tag{11}$$

一方，状態 I から R への移行確率は，$r_x(t) = 0$ と $r_x(t+1) = 1$ であることが必要であるため，以下のようになる．

$$p(I \to R) = \gamma \tag{12}$$

式 (11)，(12) から，式 (9) は式 (13) へと変形できる．

$$\begin{aligned} S(t+1) &= S(t) - \beta S(t)I(t) \\ I(t+1) &= I(t) + \beta S(t)I(t) - \gamma I(t) \end{aligned} \tag{13}$$

$$R(t+1) = R(t) + \gamma I(t)$$

上式を差分方程式へと表現を変えると，式 (1) を得ることができる．よって，提案した行列表現は SIR モデルへと帰着可能である．

## 4. マスメール型ワームの感染拡大モデル

前章で提案した行列表現モデルは，$\mathbf{G}$ により任意のネットワークトポロジが表現可能であるため，3.1 節で述べた 3 つの拡張要件のうち，3.1.1 項の要件を満たすものである．さらに 2 つの要件を満たすため，行列表現を拡張する．

なお，本論文では隔離状態 R を「ウイルス対策ソフトがそのウイルスに対応しているため，感染前にメール添付のファイルが隔離または削除されるために感染しない状態」として議論を行い，OS の再インストール等による感染可能状態 S への復帰は考えないものとする．また本論文では，新たに発生した単一ワームに関する感染拡大について考慮し，$t = 0$ 時ではウイルス対策ソフトが新ワームに未対応であることから $R(0) = 0$ とする．

### 4.1 SIR モデルの改良（改良 SIR モデル）

本節では 3.1.2 項の要件を満たすために，式 (5) を以下のように変更する．

$$\mathbf{R}(t+1) = F\left(\mathbf{R}(t) + \mathbf{\Lambda}\right) \tag{14}$$

これにより S か I かの状態を問わずに R への移行ができるようになる．

### 4.2 メーリングリストを考慮したモデル（メーリングリストモデル）

ノードがマスメール型ワームを受け取った場合，そのワームの伝搬はユーザが感染した後に行われる．行列表現の場合，時刻 $t$ で感染したノードは，時刻 $t+1$ 以降定期的にワーム伝搬を行う．一方，メーリングリストはメールを受け取ると即座にすべてのメーリングリスト参加者にメールを転送する．その際，メールがワームを含んだマスメール型ワームか，通常のメールかは判断しない．行列表現上では，時刻 $t$ で受け取ったメールを即座にメーリングリスト参加者に転送している，と考えることができる．

本論文では，メーリングリストノードを通常のメールアドレスノードとは異なる特殊ノードとして扱う．メーリングリストノードは以下の性質を持つ．また状態遷移を**図 1** に示す．

(1)　1 時刻で感染→拡大の 2 作業を行う．
(2)　拡大活動を行った後は状態 S に戻る．
(3)　感染率 $\beta$ は 1.
(4)　定期的な拡大活動は行わない．

メーリングリストノードの特殊性をモデルに反映させるために，離散時間 $t$ の間隔を半分にした $\tau$ を用いる（**図 2**）．そして，$\tau = 2t$ のときに通常の感染拡大活動を行い，$\tau = 2t+1$ のときにメーリングリストノードによる拡大活動のみが行われることとした．

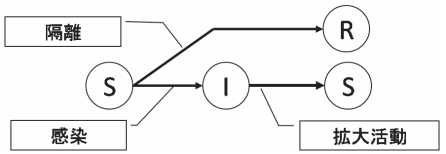


**図 1**　メーリングリストノードの状態遷移
Fig. 1　2 step epidemic dynamics of an mailing list node.

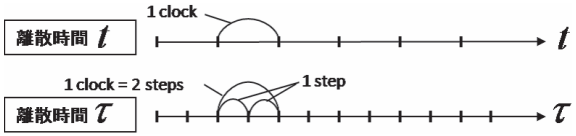


**図 2**　提案モデルの時刻表現
Fig. 2　Time scale for the proposed model.

次に，メーリングリストノードの特殊性を反映するために行列表現を拡張する．状態行列 $\mathbf{I}(\tau)$，$\mathbf{R}(\tau)$ とネットワーク行列 $\mathbf{G}$ を，メーリングリストノードの状態行列 $\mathbf{I}_{(ML)}(\tau)$，$\mathbf{R}_{(ML)}(\tau)$ とネットワーク行列 $\mathbf{G}_{\{U\to ML\}}$，$\mathbf{G}_{\{ML\to U\}}$，$\mathbf{G}_{\{ML\to ML\}}$ を使い，それぞれ $\mathbf{I}'(\tau)$，$\mathbf{R}'(\tau)$，$\mathbf{G}'$ へと拡張する．ここで，$\mathbf{G}_{\{U\to ML\}}$ は通常ノードからメーリングリストにつながるネットワーク情報を示す行列を表す．

$$\mathbf{I}'(\tau) = \left\{\mathbf{I}(\tau), \mathbf{I}_{\{ML\}}(\tau)\right\},$$

$$\mathbf{R}'(\tau) = \left\{\mathbf{R}(\tau), \mathbf{R}_{\{ML\}}(\tau)\right\}$$

$$\mathbf{G}' = \begin{pmatrix} \mathbf{G} & \mathbf{G}_{\{U\to ML\}} \\ \mathbf{G}_{\{ML\to U\}} & \mathbf{G}_{\{ML\to ML\}} \end{pmatrix}$$

$$\hat{\mathbf{G}}_{\{U\to\}} = \left\{\mathbf{G}, \mathbf{G}_{\{U\to ML\}}\right\},$$

$$\hat{\mathbf{G}}_{\{ML\to\}} = \left\{\mathbf{G}_{\{ML\to U\}}, \mathbf{G}_{\{ML\to ML\}}\right\}$$

上記の表現を用いることで，状態 R の遷移式（式 (5)）は式 (15) に拡張され，状態 I の遷移式（式 (7)）は式 (16) に拡張される．

$$\mathbf{R}'(2\tau+1) = F\left(\mathbf{R}'(2\tau) + \left\{\mathbf{\Lambda}, \mathbf{\Lambda}_{\{ML\}}\right\}\right)$$
$$\mathbf{R}'(2\tau+2) = \mathbf{R}'(2\tau+1) \quad (15)$$

$$\mathbf{I}'(2\tau+1) = F\left(\mathbf{I}'(2\tau)\left(\{\mathbf{E},\mathbf{0}\} + \mathbf{D}'(2\tau)\cdot\hat{\mathbf{G}}_{\{U\to\}}\right)\right)\cdot\left(\mathbf{1}-\mathbf{R}'(2\tau+1)\right)$$
$$\mathbf{I}'(2\tau+2) = F\left(\left\{\mathbf{I}'(2\tau+1),\mathbf{0}\right\} + \mathbf{I}_{\{ML\}}(2\tau+1)\hat{\mathbf{G}}_{\{ML\to\}}\right)\left(\mathbf{1}-\mathbf{R}'(2\tau+2)\right) \quad (16)$$

### 4.3　提案モデルの差分方程式への帰着

本節では，前節で得られたモデルを差分方程式へと帰着させる．これらの帰着は 3.3 節と同様の手法で得られる．

#### 4.3.1　改良 SIR モデル

$$\frac{dS}{dt} = -(1-\gamma)\beta S m_x(t) - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S m_x(t) - \gamma I \quad (17)$$
$$\frac{dR}{dt} = \gamma(S+I)$$

なお，$m_x(t) = \sum_y i_y(t) g_{yx}$ である．

#### 4.3.2　メーリングリストモデル

$$\frac{dS}{dt} = -(1-\gamma)\beta S\Omega - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S\Omega - \gamma I \quad (18)$$
$$\frac{dR}{dt} = \gamma(S+I)$$

ここで，$\Omega = m_x(t) + \sum_{y=N+1}^{N+M} m_y(t) g_{yx}$ である．

## 5.　シミュレーション

前節で提案した行列を用いた改良 SIR モデルとメーリングリストモデルの双方についてシミュレーションを行い，これまでのモデルとの比較を行った．比較は，1）ネットワークトポロジの差による感染の違い，2）SIR モデルと改良 SIR モデル，3）感染におけるメーリングリストの影響，の 3 点に注目した．

すべてのシミュレーションにおいて，ノード数 $N$ を 3,000，初期感染ノード数 $I(0)$ を 50 とした．シミュレーションは各条件で 10 回行い，図中に示される $E[I(t)]$ は感染ノード数の平均値である．またスケールフリー性を持つネットワークトポロジの生成には Barabasi と Albert のモデル（BA モデル）[16] を採用した．またトポロジの条件としてメーリングリストの参加者に他のメーリングリストを含まない，$\mathbf{G}_{\{ML\to ML\}} = \mathbf{0}$ とした．

### 5.1　感染におけるスケールフリートポロジの影響

本節ではネットワークトポロジの違いによる感染の差を見る．比較するネットワークトポロジは平均次数 $E[k]$ を同じにしたスケールフリー性を持つネットワークとランダムネットワークの 2 つである．

まず隔離状態を持たない SI モデルの感染者数 $I(t)$ の推移を見ることで，感染拡大速度の差を見る．平均次数 $E[k]$ は Zou らの論文[12),13)] において行われたシミュレーションと同じものとして 8 を選択した．また通常のノードの平均次数とメーリングリストのノードの平均次数は簡略化のために同じ値とした．**図 3** は感染率 $\beta = 0.005$ のときの $E[I(t)]$ を，**図 4** は $\beta = 0.01$ のときの $E[I(t)]$ をそれぞれ示したものである．双方の結果は，スケールフリーネットワークの感染拡大はランダムネットワークよりも早く起こっていることを示している．

次に SIR モデルでの $I(t)$ の推移を見ることで，感染の後の収束の差を見る．**図 5** は感染率 $\beta = 0.005$，隔離率 $\gamma = 0.01$ のときの $E[I(t)]$ を示したものであり，**図 6** は $\beta = 0.01$，$\gamma = 0.01$ のときの $E[I(t)]$ を示したものである．双方の結果より，スケールフリーネットワーク上での感染者数ピークはランダムネットワークのものよりも早く訪れて収束に転じることが分かる．

### 5.2　改良 SIR モデルの評価

本節では SIR モデルと改良 SIR モデルの違いを見る．**図 7** は感染率 $\beta = 0.005$，隔離率 $\gamma = 0.01$ のときの $E[I(t)]$ を示したものであり，**図 8** は $\beta = 0.01$，$\gamma = 0.01$ のときの

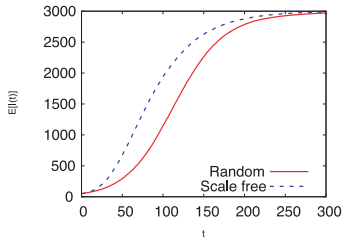


**図 3**　SI モデルでの感染者数推移：$\beta = 0.005$
Fig. 3　SI model simulation: $\beta = 0.005$.

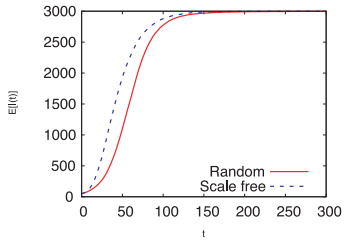


**図 4**　SI モデルでの感染者数推移：$\beta = 0.01$
Fig. 4　SI model simulation: $\beta = 0.01$.

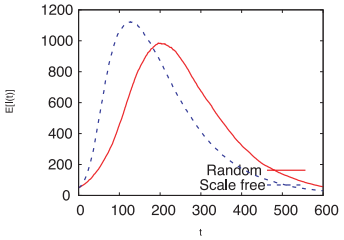


**図 5**　SIR モデルでの感染者数推移：$\beta = 0.005$，$\gamma = 0.01$
Fig. 5　SIR model simulation: $\beta = 0.005$, $\gamma = 0.01$.

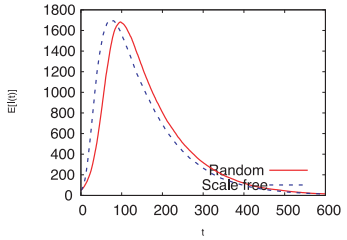


**図 6**　SIR モデルでの感染者数推移：$\beta = 0.01$，$\gamma = 0.01$
Fig. 6　SIR model simulation: $\beta = 0.01$, $\gamma = 0.01$.

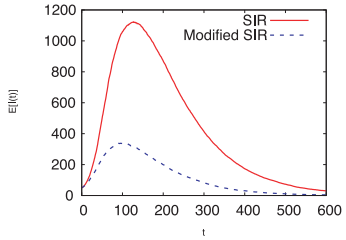


**図 7**　改良 SIR モデルの評価：$\beta = 0.005$，$\gamma = 0.01$
Fig. 7　SIR model and variant SIR model: $\beta = 0.005$, $\gamma = 0.01$.

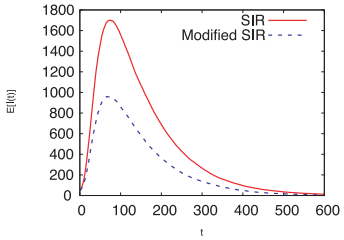


**図 8**　改良 SIR モデルの評価：$\beta = 0.01$，$\gamma = 0.01$
Fig. 8　SIR model and variant SIR model: $\beta = 0.01$, $\gamma = 0.01$.

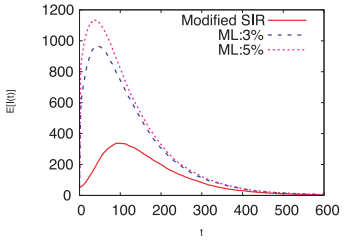


**図 9**　Mailing list impact: $\beta = 0.005$, $\gamma = 0.01$
Fig. 9　Mailing list impact: $\beta = 0.005$, $\gamma = 0.01$.

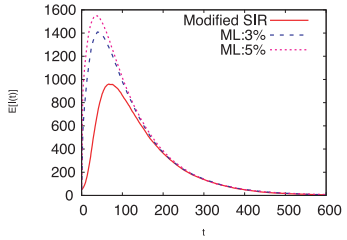


**図 10**　Mailing list impact: $\beta = 0.01$, $\gamma = 0.01$
Fig. 10　Mailing list impact: $\beta = 0.01$, $\gamma = 0.01$.

**表 1**　図 9，10 における最初の 3 時刻
Table 1　First three steps of $I(t)$ in Figs. 9, 10.

図 9 における最初の 3 時刻

| $t$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 改良 SIR | 50 | 51.3 | 53.4 | 54.2 |
| メーリングリスト：3% | 50 | 203.6 | 437.7 | 545.7 |
| メーリングリスト：5% | 50 | 230.1 | 630.2 | 808.0 |

図 10 における最初の 3 時刻

| $t$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 改良 SIR | 50 | 53.1 | 56.8 | 61.6 |
| メーリングリスト：3% | 50 | 167.3 | 409.2 | 553.2 |
| メーリングリスト：5% | 50 | 261.8 | 702.8 | 883.3 |

$E[I(t)]$ を示したものである．双方とも改良 SIR モデルでは感染者数が総じて少なくなっており，感受性状態（S）から直接隔離状態（R）に遷移する改良モデルの特徴が現れていることが分かる．また双方のシミュレーションは隔離率 $\gamma$ を揃えて行っており，感染率 $\beta$ の差が感染者数ピーク値の大きな差として現れていることが分かる．

### 5.3　マスメール型ワームの感染におけるメーリングリストの影響

本節では，改良 SIR モデルと本論文で提案したメーリングリストモデルの違いを見る．

筆者らが運用管理する SMTP サーバの 3 カ月間のログを解析したところ，31,338 のメールが転送されており，またメールアドレスの数は 2,250，送信元–宛先の組は 3,397 個存在した．そしてそれらメールアドレス群よりメーリングリスト 69 個を抽出した．調査環境におけるメーリングリストの割合は 3.1%であったため，メーリングリストを考慮したモデルでは，全ノードに対するメーリングリストアドレスノードの割合が 3%と 5%のものを 2 つ用意した．

**図 9** は感染率 $\beta = 0.005$，隔離率 $\gamma = 0.01$ のときの $E[I(t)]$ をを示したものであり，**図 10** は $\beta = 0.01$，$\gamma = 0.01$ のときの $E[I(t)]$ を示したものである．両ケースでメーリングリストモデルの感染拡大が急速に行われていることが分かる．改良 SIR モデルのパラメータは前節と同じであり，感染率 $\beta$ の差により感染者数のピークが大きくことなることが示されているが，メーリングリストの効果が感染率の差よりも非常に強いことが分かる．

さらに特徴的なのは，メーリングリストモデルでの初期段階での急激な感染拡大である．**表 1** は図 9，10 のシミュレーションでの感染初期 3 時刻での感染者数 $E[I(t)]$ を示したものであり，**図 11**，**図 12** は感染初期 10 時刻での感染者数 $E[I(t)]$ のグラフである．

これら急激な拡大は，式 (18) における $I(t)$ の増加要因 $\Omega = m_x(t) + \sum_{y=N+1}^{N+M} m_y(t) g_{yx}$

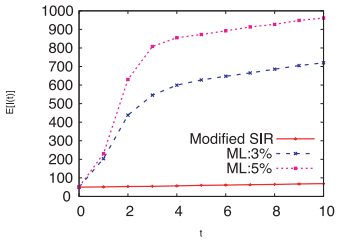


**図 11**　図 9 における最初の 10 時刻
Fig. 11　First ten steps in Fig. 9.

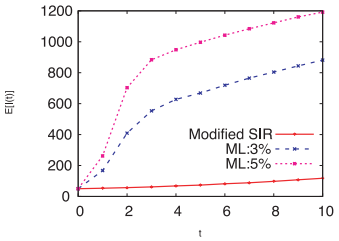


**図 12**　図 10 における最初の 10 時刻
Fig. 12　First ten steps in Fig. 10.

が初期時刻で大きな値をとることで起きている．Ω の値はメーリングリストによる感染は通常ユーザの 2 時刻分を 1 時刻で行い，さらにその感染率は 1 であることによる感染拡大力を反映する．それに加えて，スケールフリー性トポロジの特徴であるハブノードはその接続性の高さから拡大の初期段階で感染しやすいことが急激な増加をさらに加速させていることが考えられる．

これらの結果により，感染拡大の閾値である基本再生産数 $R_0 = \frac{\beta N}{\gamma}$ が通常閾値とされる 1 よりも小さな値で感染拡大が起こることが考えられるが，メーリングリストによる効果の有無によらず，スケールフリー性を持つネットワーク上での基本再生産数はほとんど 0 になることが示されていることから[17]，本論文では $R_0$ の議論は行わないこととした．

## 6. ま と め

本論文では，メールを介して感染を行うマスメール型ワームの感染数理モデルを提案した．提案モデルは，従来の感染数理モデルでは扱うことが困難であった，1）任意のネットワークトポロジの適用と，2）感受性状態から隔離状態への遷移，3）メーリングリスト効果の適用，を可能とした．さらに提案モデルを用いてシミュレーションを行い，メーリングリストの効果が感染初期で強く現れ，感染率の差と比較して感染の効果が非常に高くなることを判明することができた．

マスメール型ワームの中で代表的な NetSky は，2004 年に発生したものであるが，いまだ多くの届出と検出が報告されていることから，生存力の強さが分かる．これらの生存を説明するために，隔離率の動的変化や，再感染という要因を提案モデルに新たに加えることで，より現実に即したマスメール型ワームの感染状況のモデル化も可能であり，本論文の成果はそれら現実に近いモデルによる効果的な対策を検討することを可能にするものである．

## 参 考 文 献

1) コンピュータウイルス・不正アクセスの届出状況［3 月分および第 1 四半期］について，独立行政法人情報処理推進機構セキュリティセンター（*IPA/ISEC*）(2009). http://www.ipa.go.jp/security/txt/2009/04outline.html
2) Kephart, J., Chess, D.M. and White, S.: Computers and Epidemiology, *IEEE Spectrum*, Vol.30, No.5 (1993).
3) Kephart, J. and White, S.: Directed-Graph Epidemiological Models of Computer Viruses, *Proc. IEEE Symposium on Security and Privacy*, pp.343–359 (1991).
4) Kephart, J. and White, S.: Measuring and Modeling Computer Virus Prevalence, *Proc. IEEE Symposium on Security and Privacy* (1993).
5) Wong, C., Bielski, S., McCune, J., et al.: A Study of Mass-mailing Worms, *Proc. ACM Workshop on Rapid Malcode* (*WORM'03*) (2003).
6) Ishibashi, K., Toyono, T. and Toyama, K.: Detecting Mass-Mailing Worm Infected Hosts by Mining DNS Traffic Data, *Proc. ACM SIGCOMM 2005 Workshop on Mining Network Data* (*MineNet2005*) (2005).
7) Ebel, H., Mielsch, L.-I. and Bornholdt, S.: Scale-free topology of e-mail networks, *Physical Review E*, Vol.66, 035103(R) (2002).
8) Newman, M.E.J., Forrest, S. and Balthrop, J.: Email networks and the spread of computer viruses, *Physical Review E*, Vol.66, 035101(R) (2002).
9) Briesemeister, L., Lincoln, P. and Porras, P.: Epidemic Profiles and Defense of Scale-Free Networks, *Proc. ACM Workshop on Rapid Malcode* (*WORM'03*) (2003).
10) Pastor-Satorras, R. and Vespignani, A.: Epidemic Spreading in Scale-Free Networks, *Physical Review Letters*, Vol.86, pp.3200–3203 (2001).
11) Nikoloskia, Z., Deob, N. and Kucera, L.: Correlation Model of Worm Propagation on Scale-Free Networks, *COMPLEXUS*, Vol.3, No.1-3, pp.169–182 (2006).
12) Zou, C.C., Towsley, D.F. and Gong, W.: Modeling and Simulation Study of the Propagation and Defense of Internet E-mail Worms, *IEEE Trans. Dependable and Secure Computing*, Vol.4, No.2, pp.105–118 (2007).
13) Zou, C.C., Towsley, D.F. and Gong, W.: Email Worm Modeling and Defense, *Proc. International Conference on Computer Communications and Networks* (*ICCCN 2004*), pp.409–414 (2004).
14) Kermack, W.O. and McKendrick, A.G.: A contribution to the mathematical theory of epidemics, *Proc. Royal Society A*, Vol.115, pp.700–721 (1927).
15) Weiss, G.H. and Dishon, M.: On the asymptotic behavior of the stochastic and deterministic models of an epidemic, *Mathematical Biosciences*, Vol.11, Issue 3-4, pp.261–265 (1971).
16) Barabási, A-L. and Albert, R.: Emergence of scaling in random networks, *Science*,

Vol.286, pp.509–512 (1999).
17) Paster-Satorras, R. and Vespignani, A.: Immunization of complex networks, *Physical Review E*, Vol.65, p.036104 (2002).

(平成 21 年 4 月 20 日受付)
(平成 21 年 12 月 17 日採録)

**金岡　晃**（正会員）

2001 年東邦大学大学院理学研究科修士課程修了．2004 年筑波大学大学院博士課程システム情報工学研究科修了．博士（工学）．同年セコム株式会社入社．その後，筑波大学大学院システム情報工学研究科研究員をへて，2008 年より筑波大学大学院システム情報工学研究科助教，現在に至る．ネットワークシステムの安全設計方式，電子認証に関する研究に従事．IEEE，ACM，電子情報通信学会各会員．

**勝野　恭治**（正会員）

1998 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本アイ・ビー・エム株式会社入社．東京基礎研究所主任研究員．2009 年筑波大学大学院システム情報工学研究科リスク工学専攻後期博士課程修了．博士（工学）．2003 年ソフトウェア科学会高橋奨励賞受賞．情報セキュリティ，クラウド・コンピューティング，コンピュータ・ネットワーク，エージェント技術に関する研究開発に従事．日本ソフトウェア科学会会員．

**岡本　栄司**（フェロー）

1973 年東京工業大学工学部電子工学科卒業．1978 年東京工業大学大学院電子工学専攻博士課程修了．工学博士．同年日本電気（株）中央研究所入社．その後，北陸先端科学技術大学院大学，東邦大学をへて 2002 年より筑波大学システム情報工学研究科教授，現在に至る．グラフ理論，通信理論，アルゴリズム，情報セキュリティをはじめとする情報数理工学の教育・研究に従事．1990 年電子通信学会論文賞，1993 年本学会ベストオーサ賞．2008 年本学会論文賞，CHES2007 と CHES2009 の Best Paper Award 受賞．著書『暗号理論入門』（共立出版），『電子マネー』（岩波書店）等．2003 年電子情報通信学会フェロー，2004 年本学会フェロー．IEEE，ACM，電子情報通信学会，情報理論とその応用学会，日本セキュリティ・マネージメント学会，IACR（International Association for Cryptologic Research）各会員，IJIS（International Journal of Information Security）編集長．