確率モデルによるコンピュータウィルスの特徴分析

小林尚志[†] 岡村寛之[†] 土肥正[†]

インターネット社会においてコンピュータウィルスは脅威であり，大きな社会的問題となっている．コンピュータウィルスの感染能力を測るための挙動解析手法として，微分方程式を用いた方法がある．しかしながら，微分方程式による方法は確定論的な解析であり，コンピュータウィルスの感染除去にともなう不確実性の影響を無視している．そこで本論文では，確率モデルに基づいたコンピュータウィルスの挙動解析を行う．次に，提案した確率モデルに対してコンピュータウィルスの感染能力を特徴付けるための定量的評価尺度を定義する．また，ウィルスの感染除去がマルコフ過程で記述される環境において，具体的な評価尺度の計算方法を導出する．これにより，すべてのウィルスに対して評価尺度が導出可能となる．さらに数値例では，現存する 2 種類のウィルス被害件数データからパラメータの推定および定量的評価尺度の算出を行い，コンピュータウィルスの特徴分析を行う．

Characteristic Analysis of Computer Viruses by Stochastic Models

Hisashi Kobayashi,[†] Hiroyuki Okamura[†] and Tadashi Dohi[†]

The computer virus prevalence is a severe problem in the Internet. In order to evaluate the computer virus prevalence, the virus models based on ordinary differencial equations, which are deterministic models, have been used in the literature. However, the deterministic models ignore the effect of uncertainty arising in the prevalence and removal process of computer virus, so that they are not appropriate to characterize the computer virus with quantitative measures. In this paper, we develop a stochastic model of computer viruses and propose the quantitative measures to characterize the computer virus. Moreover, the algorithm to calculate the quantitative measures is derived under the assumption that the virus prevalence is described by Markov processes. This leads to derive the measures for all the computer viruses. In numerical examples, we carry out the characteristic analysis of computer virus by using the real infection data from existing two kinds of computer viruses.

1. はじめに

コンピュータネットワークを基本とした現在の高度情報化社会において，インターネットは不可欠な情報伝達媒体として認識されている．しかし，インターネットの接続は，同時にコンピュータウィルス感染やサイバーテロなどの脅威にさらされているといっても過言ではない．コンピュータウィルスの感染予防に対して一般的に行われている対策は，定期的にワクチンソフトのウィルスの定義ファイルをダウンロードし，ウィルスの浸入を食い止めることである．しかしながら，近年コンピュータウィルスの感染方法や感染経路は複雑多岐にわたっており，ウィルス感染の対策が後手に回ることも少なくない．たとえば，2002 年の 4 月に流行したワーム KLEZ は感染者のローカルドライブにある任意のメールアドレスを差出人にしてウィルスメールを詐称するため，多くの被害が発生した．また，KLEZ は様々な亜種が存在することも特徴の 1 つであり，すべての亜種に対応するウィルス定義ファイルを準備することが遅れたことも感染爆発を招いた 1 つの要因である．また，2001 年に流行したワーム NIMDA は「メール感染」「Web 感染」「ファイル感染」という 3 つの感染経路を持つため非常に感染力が強く，感染した端末台数は一時 15 万台以上となり，ネットワークの機能不全を引き起こした．このように，きわめて感染力の強いウィルスに対しては，ユーザが適切な定義ファイルをダウンロードしたりセキュリティホールを修繕するパッチをあてたりする間に感染の拡大を起こす可能性があり，必然的に対策が後手にまわってしまう．

ウィルスの蔓延を未然に防ぐための対策として，これまでに様々なものが考案され，現実に適用されている．たとえば，ルータなどのネットワーク機器にウィ

† 広島大学大学院工学研究科
Graduate School of Engineering, Hiroshima University

ルスの検知機能を持たせる技術や，簡単なところではワクチンソフトが自動的にウィルス定義ファイルを更新する技術などがある．Okamoto ら[1] は免疫システムに注目し，自律的にウィルスを撃退するアンチウィルスシステムの開発を行っている．また，Badhusha ら[2] はアンチウィルスソフトのウィルス定義ファイルの自動更新に着目し，ネットワーク経由で自動更新を行う際のパケットを解析し，自動更新の有効性について議論している．

上述のようなウィルス蔓延予防に対する様々な方策が提案されている一方で，ウィルス感染力などコンピュータウィルス自体の拡散現象を解析し影響力の評価を行う試みがなされている．Thimbleby ら[3] はチューリングマシンモデルを構築し，コンピュータウィルスの性質を定性的に議論している．また，Kephart ら[4),5)]，Kephart[6]，豊泉ら[7]，Toyoizumi ら[8] はコンピュータウィルス感染の振舞いに対して，数理生態学や社会学において生態の繁殖過程や人口増加の時間的挙動を記述するために用いられる微分方程式モデルを利用している．この微分方程式に基づいたウィルス挙動解析は確定論的な方法として分類される．つまり，ウィルスの感染や除去をある種の力学系における軌道としてとらえることが可能となり，コンピュータウィルスに関する影響力の評価を行うことができる．しかしながら，これらの確定論的な方法はウィルスの感染除去にともなう不確実性の影響を無視している．そのため，ウィルスの感染爆発や完全な除去（死滅）のような希な事象を表現することができない．また，微分方程式は与えられた初期パラメータによって以降の振舞いが一意に決定されるため，たとえばウィルスが死滅する事象に関連した評価尺度を導出する際に，パラメータによっては死滅がまったく起こりえない場合が存在する．したがって，確定論的な解析によって，あらゆるタイプのウィルスに対して適用可能な評価尺度を求めることは困難である．他方，Wang ら[9] は不確実性を考慮したウィルスモデルを提案し，いくつかのネットワーク環境を想定してシミュレーションに基づいた挙動解析を行っている．しかし，シミュレーションを基礎とした評価は解析的に導出される評価尺度と比較すると精度の観点において劣っていることが自明である．

そこで本論文では，確率モデルに基づいたコンピュータウィルスの挙動解析を行う．次にコンピュータウィルスに共通して現れる現象（死滅など）に着目して，あらゆるタイプのコンピュータウィルスを特徴付けるために有用な定量的評価尺度を提案する．特に，ウィルスの感染と除去がマルコフ過程で記述される環境において，具体的に定量的評価尺度の導出を行う．また，実際のウィルスの被害件数データと提案したマルコフモデルに基づいて，現存する 2 種類のコンピュータウィルスに関する特徴分析を行う．

本論文の構成は以下のとおりである．2 章では Kephart ら[4),5)]，Kephart[6]，豊泉ら[7]，Toyoizumi ら[8] で議論されている微分方程式モデルに基づいたコンピュータウィルスの挙動解析方法の紹介を行う．特に Kephart ら[5] によって取り扱われている Kill Signal（以下 KS と記述）を考慮した微分方程式モデルを紹介する．ここで，KS とは，ウィルスに関する警告を意味し，「KS を受け取った端末がウィルス感染していた場合は感染を修復し，同時に感染の可能性のある端末に KS を送る」という特徴を持つ．すなわち，KS はウィルスに対する免疫力として解釈される．3 章はコンピュータウィルスの挙動を解析するための確率モデルの提案を行う．ここでは，ウィルスを特徴付けるための定量的な評価尺度の提案と，ウィルスの感染・除去がマルコフ過程で記述される際の具体的な解析方法を示す．特に KS を考慮しない確率モデルと KS を考慮した確率モデルに焦点をあて，各種評価尺度の具体的な計算法について言及する．4 章では，実際のウィルス被害件数データと提案した確率モデルを用いて，現存する 2 種類のコンピュータウィルスに対する特徴分析を行う．

2. 関 連 研 究

Kephart ら[4),5)]，Kephart[6]，豊泉ら[7]，Toyoizumi ら[8] はコンピュータウィルス感染数の時間的挙動を記述する微分方程式モデルを提案している．ここでは，最も単純なコンピュータウィルスに関する微分方程式モデルを紹介し，その後に Kephart ら[5] が提案した KS を考慮した微分方程式モデルについて述べる．

最も単純なコンピュータウィルスに対するモデルとは，ウィルスの増殖過程のみに着目した微分方程式モデルである．いま $n(t)$ を時刻 t においてウィルスに感染したことのある端末台数（ウィルスを除去したか否かは問わない）とすると，$n(t)$ は微分方程式により以下のようにモデル化される．

$$\frac{dn(t)}{dt} = \beta n(t)\{K - n(t)\}, \quad n(0) = n_0. \quad (1)$$

ここで β (≥ 0) はウィルスの感染率，K (≥ 0) はウィルスに感染する可能性のある総端末台数である．特に β は単位時間あたりに 1 台の端末が他の 1 台に感染する割合を表す．式 (1) はロジスティック方程式であ

り，その解はロジスティック曲線と呼ばれる S 字曲線となる．上記の微分方程式はウィルスの増殖過程のみに着目したモデルであるが，現実にはウィルス感染が発覚するとウィルスの除去さらに再発防止の対策などが実施されるため，上記の微分方程式よりもさらに複雑な挙動を示すことが予想される．

Kephart ら[5] は KS を考慮した微分方程式モデルを提案している．KS とはウィルスに関する一種の免疫力であり，

(1) KS を受け取った端末がウィルスに感染していた場合は除去を行う，
(2) ウィルス除去後は KS を保持する，
(3) ウィルス除去後に感染の可能性のある端末へ KS を送る，
(4) 時間の経過によって KS が失われる，

という特徴を持つ．例をあげると，A さんが所有する端末がウィルスに感染したことが分かったとき，A さんは頻繁にデータのやりとりをする B さんと C さんに自分がウィルス感染したこと知らせる．この A さんが B さんと C さんに伝える警告が KS である．また，KS の減少は警告を忘れることに対応する．Kephart ら[5] はウィルスに感染した端末台数 $n(t)$ と KS を保持している端末台数 $r(t)$ の時間変化に関する微分方程式を導出し，ウィルス感染に関する挙動について議論している．次の記号を定義する．

β (≥ 0)：ウィルス感染率（単位時間あたりに 1 台の端末が他の 1 台に感染する割合）
δ (≥ 0)：ウィルス除去率（単位時間あたりにウィルス感染している端末がウィルスを除去する割合）
β_r (≥ 0)：KS 増加率（単位時間あたりに 1 台の端末が他の 1 台に KS を送る割合）
δ_r (≥ 0)：KS 減少率（単位時間あたりに KS を保持している端末が KS を失う割合）

このとき，以下の微分方程式に基づいて挙動解析を行うことができる．

$$\frac{dn(t)}{dt} = \beta n(t)\{K - n(t) - r(t)\} - \delta n(t) - \beta_r n(t) r(t), \quad n(0) = n_0, \tag{2}$$

$$\frac{dr(t)}{dt} = \beta_r r(t)\{K - r(t)\} + \delta n(t) - \delta_r r(t), \quad r(0) = r_0. \tag{3}$$

3. 確率モデルに基づいたウィルスの挙動解析

前節で紹介した微分方程式によるコンピュータウィルスの挙動解析は，ウィルス被害件数（感染端末台数）の周期的な振舞いなどのウィルスに関する挙動を表現することができる．しかしながら，2 種類のウィルスの特徴を比較する場合は定性的な挙動に関する議論だけでは不十分であり，比較を行うための定量的評価尺度が必要となる．

本章ではすべてのウィルスに適用可能な評価尺度の導出を行うことを目的として，確率モデルに基づいたウィルスの挙動解析を行う．まず最初に，ウィルスの感染端末台数を確率過程として表現し，ウィルスを特徴付ける定量的な評価尺度の定義を行う．次に，ウィルス感染と除去がマルコフ連鎖によって記述できる環境を考え，定量的評価尺度の導出を行う．特に KS を考慮しないモデルと KS を考慮したモデルのそれぞれに対してウィルスがあるレベル以上感染するまでの時間とウィルスが死滅するまでの時間について考察を行う．

3.1 評価尺度の定義

ここではウィルス感染端末台数を確率過程で表現した場合に，ウィルスの感染能力を特徴付ける定量的な評価尺度の導出を行う．いま $\{N_f(t);\ t \geq 0\}$ を時刻 t においてウィルスに感染した端末台数を表す確率過程として定義する．また，時刻 $t = 0$ における感染端末台数を n_0 とする．このとき，感染端末台数があるレベル x に初めて到達する時刻を表す確率変数を T_x，感染端末台数が初めて 0 に到達する時刻を T_0 とすると

$$T_x = \inf\{t \geq 0; N_f(t) \geq x \mid N_f(0) = n_0\} \tag{4}$$

および

$$T_0 = \inf\{t \geq 0; N_f(t) = 0 \mid N_f(0) = n_0\} \tag{5}$$

となる．

本論文では，ウィルス感染力とウィルス継続力に着目する．ウィルス感染力とはウィルスが拡散する能力そのものを表し，他の端末への影響度としてとらえることができる．一方，ウィルス継続力とはウィルスの除去に関する評価尺度であり，継続力が高いものほどウィルスが長期にわたって存在する．もちろん，これらのパラメータは，定性的な議論においては，ウィルスの感染経路や対象としている OS の普及率により決定されるべきものである．しかしながら，本論文ではウィルス感染端末台数の時間的な振舞いのみに着目することで，あらゆるウィルスに対しても適用可能な評価尺度を導出することを目指す．これによってウィルスの感染経路や OS に依存しない，ウィルスの評価を行うことが可能となり，ひいてはウィルス予防対策の評価を行うことが可能となる．

ウィルス感染力を定量的に表現するために，ここではハザードという概念を取り入れる．ハザードとはウィルス感染端末台数に関する臨界レベル c（たとえ

ば，すべての端末台数や全端末の 80%以上など）を設定し，ウィルスが死滅することなく設定した臨界レベル c に達することと定義する．このとき，ウィルス感染力はいかに早くハザードが発生するか，いい換えるとハザードが発生する時間によって計測することができる．

時刻 t でウィルスが死滅せずにハザードが発生しない確率は，前述した初到達時間 T_x および T_0 を用いると

$$R_c(t) = \Pr\{T_c > t \mid T_c < T_0\} \tag{6}$$

となる．また，ハザードが発生するまでの平均時間（Mean Time to Hazard: MTTH）は

$$\mathrm{MTTH}(c) = \mathrm{E}[T_c | T_c < T_0] = \int_0^\infty R_c(t)dt \tag{7}$$

によって与えられる．上記のハザード発生確率 $R_c(t)$ および $\mathrm{MTTH}(c)$ によって，ウィルス感染力を定量的に表現することが可能となる．

一方，ウィルス継続力に関して，ウィルス感染力の表現と同様に，ウィルスがいかに早く死滅するか，つまりウィルス感染端末台数が 0 に到達するまでの時間によって計測することができる．

時刻 t でウィルスが死滅しない確率は

$$R_0(t) = \Pr\{T_0 > t\} \tag{8}$$

となり，死滅するまでの平均時間（Mean Time to Extinction: MTTE）は

$$\mathrm{MTTE} = \mathrm{E}[T_0] = \int_0^\infty R_0(t)dt \tag{9}$$

となる．この $R_0(t)$ および MTTE によりウィルス継続力に関する特徴分析を行う．

3.2 マルコフ過程によるモデル化

ここでは KS を考慮しないコンピュータウィルスの感染・除去過程として出生死滅過程を仮定したモデルを提案し，前述した定量的評価尺度を導出する．

いま，ウィルス感染端末が n 台存在するという条件のもとで，ウィルスの感染率 λ_n と除去率 μ_n を以下のように仮定する．

$$\lambda_n = \beta n(K-n), \qquad n = 0, 1, \ldots, K, \tag{10}$$

$$\mu_n = \delta n, \qquad n = 1, \ldots, K. \tag{11}$$

ここで，K はウィルスに感染する可能性のある総端末台数であり，β と δ はウィルスの感染率と除去率に関するパラメータである．これは，ウィルス感染端末台数が現在感染している台数 n と感染していない端末数 $K-n$ に比例することを示した単純なマルコフモデルである．

時刻 t において n 台の端末がウィルス感染している確率を $P_n(t)$ とおくと，次の微分差分方程式が成立する．

$$\frac{d}{dt}P_0(t) = \mu_1 P_1(t), \tag{12}$$

$$\frac{d}{dt}P_1(t) = \mu_2 P_2(t) - \nu_1 P_1(t), \tag{13}$$

$$\frac{d}{dt}P_n(t) = \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) - \nu_n P_n(t), \quad \text{for } n = 2, \ldots, K-1, \tag{14}$$

$$\frac{d}{dt}P_K(t) = \lambda_{K-1}P_{K-1}(t) - \mu_K P_K(t). \tag{15}$$

ここで $\nu_n = \lambda_n + \mu_n$ である．

いま，臨界レベル c を設定した場合を考える．すなわち，ウィルス感染した端末台数が c になる状態をハザードとして定義する．このとき，問題は状態 $N_f(t) = 0$ と $N_f(t) = c$ を吸収状態とした連続時間マルコフ連鎖の解析に帰着され，状態 $N_f(t) = c$ に吸収されるまでの時間がハザードが発生するまでの時間に対応する．つまり，ウィルス感染端末台数が c のときのウィルスの感染率とウィルス除去率を

$$\lambda_c = 0, \qquad \mu_c = 0 \tag{16}$$

としたもとで式 (12)–(15) の微分方程式の解を $\tilde{P}_n(t)$ とすると，状態確率 $\tilde{P}_n(t)$ を用いてウィルス感染率に関する諸量を得ることができる．

時刻 t までにウィルスが死滅することなく臨界レベルに達する同時確率は

$$\Pr\{T_c \le t, T_c \le T_0\} = \tilde{P}_c(t) \tag{17}$$

で与えられる．したがって，時刻 t でハザードが発生しない確率は

$$R_c(t) = 1 - \frac{\tilde{P}_c(t)}{\tilde{P}_c(\infty)} \tag{18}$$

となる．ここで $\tilde{P}_c(\infty)$ は次の手順で計算できる．F_i をウィルス感染端末台数が i のもとでウィルスが死滅することなく感染端末台数が $i+1$ に到達する確率とする．このとき $\tilde{P}_c(\infty)$ はウィルス感染端末台数が n_0 のもとでウィルスが死滅することなく c 台まで増加する確率であるため，$\tilde{P}_c(\infty) = \prod_{i=n_0}^{c-1} F_i$ となる．さらに F_i は以下の漸化式によって得られる（付録 A.1. 参照）．

$$F_0 = 0, \tag{19}$$

$$F_i = \frac{\lambda_i}{\lambda_i + \mu_i(1 - F_{i-1})}, \quad \text{for } i = 1, \ldots, K-1. \tag{20}$$

$\mathrm{MTTH}(c)$ は，ウィルス感染端末台数が i のもとでウィルスが死滅することなく $i+1$ 台感染するまでの期待時間 M_i を用いて，以下の手続きにより算出することができる．

$$\mathrm{MTTH}(c) = \sum_{i=n_0}^{c-1} M_i/F_i. \tag{21}$$

ここで，

$$M_0 = 0, \tag{22}$$

$$M_i = \frac{\lambda_i + \mu_i(F_{i-1} + \nu_i M_{i-1})F_i}{\nu_i\{\lambda_i + \mu_i(1 - F_{i-1})\}}, \quad \text{for } i = 1, \ldots, K-1. \tag{23}$$

同様な考察から，式 (12)–(15) の解である状態確率 $P_n(t)$ を用いて，ウィルス継続力に関する評価尺度は

$$R_0(t) = 1 - P_0(t) \tag{24}$$

となる．また MTTE は，ウィルス感染端末台数が i から 1 台減少して $i-1$ になるまでの期待時間 $\overline{M}_i$ を用いて，

$$\mathrm{MTTE} = \sum_{i=n_0}^{1} \overline{M}_i \tag{25}$$

で与えられる．ここで，

$$\overline{M}_K = 1/\mu_K, \tag{26}$$

$$\overline{M}_i = 1/\nu_i + \lambda_i/\mu_i \overline{M}_{i+1}, \quad \text{for } i = K-1, \ldots, 1 \tag{27}$$

である．

3.3 Kill Signal を考慮したウィルスモデル

Kephart ら[5] が提案した KS を考慮したウィルスモデルをマルコフモデルへと拡張する．

いま $\{N_f(t);\ t \geq 0\}$ と $\{N_r(t);\ t \geq 0\}$ を，時刻 t においてウィルス感染した総端末台数と KS を保持している端末台数を表す確率過程として定義する．このとき，ウィルス感染した端末台数 $N_f(t)$ と KS を保持している端末台数 $N_r(t)$ を状態空間に持つ連続時間マルコフ連鎖を考える．状態 $(n,m) = \{N_f(t) = n, N_r(t) = m\}$ から他の状態への遷移を考えると，2 章の式 (2) と式 (3) と同様な記号を用いて，推移率は以下のようになる．

(1) ウィルスは KS を保持していない他の端末へ感染するため，状態 $(n+1, m)$ への推移率は

$$\lambda_{n,m} = \beta n(K - n - m). \tag{28}$$

(2) ウィルスが除去された端末は必ず KS を保持するため，状態 $(n-1, m)$ への推移は存在しない．

(3) ウィルスの除去は感染した端末が自分自身で除去を行う場合と，KS を受け取ることによって除去される場合の 2 つの場合が考えられるため，状態 $(n-1, m+1)$ への推移率は

$$\mu_{n,m} = \delta n + \beta_r mn. \tag{29}$$

(4) KS は自身で増加するため，状態 $(n, m+1)$ への推移率は

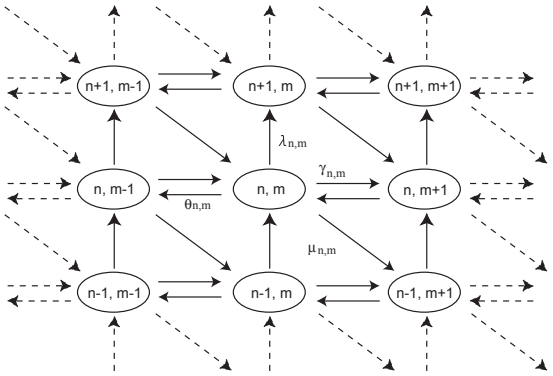


図 1 KS をともなうウィルスモデルの状態遷移図
Fig. 1 Transition diagram of the virus model with KS.

$$\gamma_{n,m} = \beta_r m(K - n - m). \tag{30}$$

(5) KS は時間の経過とともに減少するため，状態 $(n, m-1)$ への推移率は

$$\theta_{n,m} = \delta_r m. \tag{31}$$

このとき，マルコフ連鎖の状態遷移図は**図 1** のようになる．

状態確率 $P_{n,m}(t)$ を時刻 t で感染した端末台数が n，KS を保持している端末台数が m である確率とすると，次の微分差分方程式が得られる．

$$\begin{aligned}\frac{dP_{n,m}(t)}{dt} = {} & \lambda_{n-1,m}P_{n-1,m}(t) \\ & + \mu_{n+1,m-1}P_{n+1,m-1}(t) \\ & + \gamma_{n,m-1}P_{n,m-1}(t) \\ & + \theta_{n,m+1}P_{n,m+1}(t) \\ & - \nu_{n,m}P_{n,m}(t), \\ & \text{for } n = 0, \ldots, K, \\ & \quad m = 0, \ldots, K-n.\end{aligned} \tag{32}$$

ここで，総端末台数が有限であるため $K \geq n+m$ であることと，ウィルスが死滅し KS が完全に消滅した状態 $(0,0)$ が吸収状態であることに注意する．また，$\nu_{n,m} = \lambda_{n,m} + \mu_{n,m} + \gamma_{n,m} + \theta_{n,m}$ である．上記の微分方程式を数値的に解き，時刻 t における状態確率 $P_{n,m}(t)$ を算出した後に，ウィルスの感染力および継続力に関する諸量を導出することが可能である．たとえば時刻 t における期待感染端末台数は

$$\mathrm{E}[N_f(t)] = \sum_{n=0}^{K}\sum_{m=0}^{K-n} nP_{n,m}(t) \tag{33}$$

で与えられ，これは従来の微分方程式から得られる結果とほぼ同じ振舞いを示す．

ハザードに関する評価尺度は，上記の微分方程式において，ウィルス感染した端末台数が c 以上になった状態を吸収状態とすることで得られる．つまり，ウィ

ルス感染した端末台数が c のときのウィルス感染率と除去率，および KS 増加率と減少率をすべて 0 とした微分方程式を解いたときの状態確率 $\tilde{P}_{n,m}(t)$ を用いて，

$$R_c(t) = 1 - \frac{\sum_{m=0}^{K-c} \tilde{P}_{c,m}(t)}{\sum_{m=0}^{K-c} \tilde{P}_{c,m}(\infty)} \tag{34}$$

を得る．また，ウィルス継続確率は

$$R_0(t) = 1 - \sum_{m=0}^{K} P_{0,m}(t) \tag{35}$$

として与えられる．

次に，MTTH(c) と MTTE に対する計算手続きについて議論する．いま，$\boldsymbol{A}_n$，$\boldsymbol{B}_n$，$\boldsymbol{C}_n$ を，それぞれウィルス感染端末が n 台という条件のもとで感染端末が 1 台減少する，感染端末が変化しない（KS 保持端末台数が変化），感染端末が 1 台増加するという事象に対する確率行列とする．ここで，端末台数は有限であるので，$\boldsymbol{A}_n$，$\boldsymbol{B}_n$，$\boldsymbol{C}_n$ はそれぞれ $(K-n+1)\times(K-n+2)$，$(K-n+1)\times(K-n+1)$，$(K-n+1)\times(K-n)$ の行列であり，各要素はその時点での KS 保持端末台数に対応する．すなわち，$\boldsymbol{A}_n$ の i，j 成分は状態 (n,i) から状態 $(n-1,j)$ への推移確率を表す．具体的に，$\boldsymbol{A}_n$，$\boldsymbol{B}_n$，$\boldsymbol{C}_n$ の各要素（$[\boldsymbol{A}]_{ij}$ は行列 $\boldsymbol{A}$ の i，j 成分を示す）は以下のようになる．

$$[\boldsymbol{A}_n]_{ij} = \begin{cases} \mu_{n,i}/\nu_{n,i} & \text{for } i = j+1 \\ 0 & \text{otherwise,} \end{cases} \tag{36}$$

$$[\boldsymbol{B}_n]_{ij} = \begin{cases} \theta_{n,i}/\nu_{n,i} & \text{for } i = j-1 \\ \gamma_{n,i}/\nu_{n,i} & \text{for } i = j+1 \\ 0 & \text{otherwise,} \end{cases} \tag{37}$$

$$[\boldsymbol{C}_n]_{ij} = \begin{cases} \lambda_{n,i}/\nu_{n,i} & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

このとき，$\boldsymbol{F}_n$ をウィルス感染端末が n 台ある条件のもとでウィルスが死滅することなくウィルス感染端末が $n+1$ 台に増加する確率を表す行列とすると

$$\boldsymbol{F}_0 = \boldsymbol{O}, \tag{39}$$

$$\boldsymbol{F}_n = (\boldsymbol{I} - \boldsymbol{A}_n\boldsymbol{F}_{n-1} - \boldsymbol{B}_n)^{-1}\boldsymbol{C}_n, \quad \text{for } n = 1,\ldots,K-1 \tag{40}$$

となる．また，行列 $\tilde{\boldsymbol{A}}_n$，$\tilde{\boldsymbol{B}}_n$，$\tilde{\boldsymbol{C}}_n$ を，ウィルス感染端末が n 台の条件のもとで感染端末が 1 台減少する，KS 保持端末台数のみが変化する，感染端末が 1 台増加するまでの期待時間を表す行列とすると，

$$[\tilde{\boldsymbol{A}}_n]_{ij} = \begin{cases} \mu_{n,i}/\nu_{n,i}^2 & \text{for } i = j+1 \\ 0 & \text{otherwise,} \end{cases} \tag{41}$$

$$[\tilde{\boldsymbol{B}}_n]_{ij} = \begin{cases} \theta_{n,i}/\nu_{n,i}^2 & \text{for } i = j-1 \\ \gamma_{n,i}/\nu_{n,i}^2 & \text{for } i = j+1 \\ 0 & \text{otherwise,} \end{cases} \tag{42}$$

$$[\tilde{\boldsymbol{C}}_n]_{ij} = \begin{cases} \lambda_{n,i}/\nu_{n,i}^2 & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{43}$$

として与えられる．

上述の行列を用いて，臨界レベル c に対する MTTH(c) は以下の手順により算出することができる（付録 A.1. 参照）．

$$\text{MTTH}(c) = \frac{\sum_{i=n_0}^{c-1} \boldsymbol{\alpha}\boldsymbol{F}_{n_0}\cdots\boldsymbol{F}_{i-1}\boldsymbol{M}_i\boldsymbol{F}_{i+1}\cdots\boldsymbol{F}_{c-1}\boldsymbol{e}}{\boldsymbol{\alpha}\boldsymbol{F}_{n_0}\cdots\boldsymbol{F}_{c-1}\boldsymbol{e}}. \tag{44}$$

ここで，$\boldsymbol{\alpha}$ は初期時刻において KS を保持している端末の台数を表す確率ベクトル，$\boldsymbol{e}$ はすべての要素が 1 の列ベクトルである．また，$\boldsymbol{M}_n$ はウィルス感染端末が n 台存在するという条件のもとでウィルスが死滅することなく感染端末が 1 台増加するまでの期待時間であり，以下の漸化式によって算出できる．

$$\boldsymbol{M}_0 = \boldsymbol{O}, \tag{45}$$

$$\boldsymbol{M}_n = (\boldsymbol{I} - \boldsymbol{A}_n\boldsymbol{F}_{n-1} - \boldsymbol{B}_n)^{-1} \times (\tilde{\boldsymbol{A}}_n\boldsymbol{F}_{n-1}\boldsymbol{F}_n + \tilde{\boldsymbol{B}}_n\boldsymbol{F}_n + \tilde{\boldsymbol{C}}_n + \boldsymbol{A}_n\boldsymbol{M}_{n-1}\boldsymbol{F}_n), \quad \text{for } n = 1,\ldots,K-1. \tag{46}$$

次に，MTTE を算出するために，$\overline{\boldsymbol{F}}_n$ をウィルス感染端末が n 台ある条件のもとでウィルス感染端末が 1 台減少する確率を表す行列とする．このとき，

$$\overline{\boldsymbol{F}}_K = \boldsymbol{A}_K, \tag{47}$$

$$\overline{\boldsymbol{F}}_n = (\boldsymbol{I} - \boldsymbol{B}_n - \boldsymbol{C}_n\overline{\boldsymbol{F}}_{n+1})^{-1}\boldsymbol{A}_n, \quad \text{for } n = K-1,\ldots,1 \tag{48}$$

となる．次に，$\overline{\boldsymbol{M}}_n$ をウィルス感染端末が n 台存在するという条件のもとで感染端末が $n-1$ 台に減少するまでの期待時間とすると，MTTE は

$$\text{MTTE} = \frac{\sum_{i=n_0}^{1} \boldsymbol{\alpha}\overline{\boldsymbol{F}}_{n_0}\cdots\overline{\boldsymbol{F}}_{i+1}\overline{\boldsymbol{M}}_i\overline{\boldsymbol{F}}_{i-1}\cdots\overline{\boldsymbol{F}}_1\boldsymbol{e}}{\boldsymbol{\alpha}\overline{\boldsymbol{F}}_{n_0}\cdots\overline{\boldsymbol{F}}_1\boldsymbol{e}} \tag{49}$$

となる．ここで，

$$\overline{\boldsymbol{M}}_K = \tilde{\boldsymbol{A}}_K, \tag{50}$$

$$\overline{\boldsymbol{M}}_n = (\boldsymbol{I} - \boldsymbol{B}_n - \boldsymbol{C}_n\overline{\boldsymbol{F}}_{n+1})^{-1} \times (\tilde{\boldsymbol{A}}_n + \tilde{\boldsymbol{B}}_n\overline{\boldsymbol{F}}_n + \tilde{\boldsymbol{C}}_n\overline{\boldsymbol{F}}_{n+1}\overline{\boldsymbol{F}}_n + \boldsymbol{C}_n\overline{\boldsymbol{M}}_{n+1}\overline{\boldsymbol{F}}_n), \tag{51}$$

for $n = K-1, \ldots, 1$.

4. 実データに基づいたコンピュータウィルスの特徴分析

ここでは現存するコンピュータウィルスの被害件数に関するデータに基づいた解析例を示し，確率モデルに基づいたコンピュータウィルスの定量的な特徴分析を行う．

次のような 2 種類のコンピュータウィルスに着目する☆．

WORM BLAID.A（以下 BLAID）： 2002 年 11 月に発見された WindowsOS 上で繁殖するワームであり，他のウィルスを感染マシン内に作成するという特徴を持つ．また，自身もメール添付によって増殖するワームとして活動する．ワームメール送信の際には，コンピュータ内のファイルからメールアドレスを取得して宛先（To:）と送信者名（From:）に利用する．

WORM DATOM.A（以下 DATOM）： 2002 年 7 月に発見された WindowsOS 上で繁殖するワームであり，読み取りや書き込みが可能なネットワークドライブ上に自身をコピーする．単体で動作する独立なプログラムであり，他のファイルへは感染しない．また，感染した端末のシステム情報を特定のメールアドレスに送信する．破壊活動としてはファイアウォールソフトの停止を行う．

本論文で提案する解析はすべてのコンピュータウィルスに対して適用可能であるが，(i) 総感染端末台数がほぼ同じ，(ii) 感染端末台数の増加に特徴があるという理由で上記のウィルスに着目する．

解析に用いるデータはそれぞれのウィルスに関する 60 日間（2003 年 1 月 1 日〜）の総被害件数であり，トレンドマイクロ（株）が運営するウィルストラッキングセンター☆☆に掲載されている日単位の被害件数に基づいている．これらの値は報告された被害件数であり，実際の被害件数はここで示されるものよりも多いことが予想される．しかしながら，全体のウィルス被害件数に対する標本値と考えることで，全体のウィルス被害件数の時間的振舞いを予測することができる．**図 2** および**図 3** は，データ採取期間全体における日ごとの被害件数およびその累積値（総被害件数）を示したグラフである．これらの図において BLAID と DATOM の特徴がよく現れている．つまり，BLAID は初期段階で多数の被害件数が報告されているが，およそ 30 日を経過した時点から報告される被害件数が少なくなっているのに対して，DATOM に対する日単位の被害件数はデータ採取期間を通じてほぼ一定の値を示している．これらは，BLAID がメールを介した感染によって増殖するのに対して，DATOM はネットワークドライブを主な感染経路にしていることに起因している．つまり，メールを介した感染ではインターネットを経由して不特定多数のユーザに感染する可能性があるため，ウィルスが発生した直後の初期段階において多数の被害件数が報告される傾向がある．

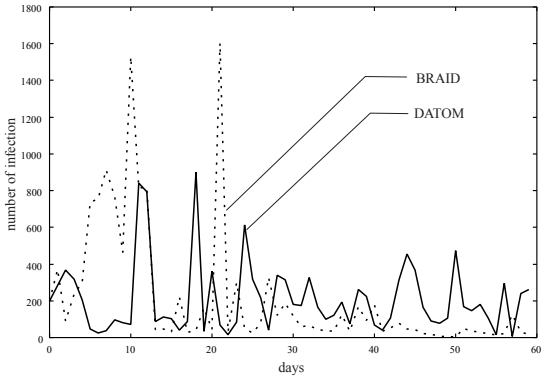


図 2 日単位の被害件数
Fig. 2 Behavior of the number of infection.

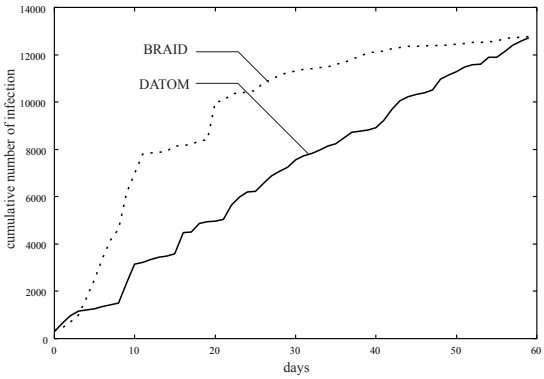


図 3 総被害件数
Fig. 3 Behavior of the cumulative number of infection.

本論文では，図 3 で示された総被害件数データに基づいて解析を行う．解析の手順は以下のとおりである．

(1) 総被害件数からウィルス感染率などのパラメータを推定．

(2) 推定値をもとに MTTH および MTTE を算出．

次に，パラメータ推定のため設定した幾つかの仮定について説明した後に，最小二乗法により推定したパ

☆ トレンドマイクロ（株）：ウィルスデータベース
http://www.trendmicro.co.jp/vinfo/virusencyclo/

☆☆ トレンドマイクロ（株）：ウィルストラッキングセンター
http://wtc.trendmicro.com/

ラメータ値と定量的評価尺度に基づいてウィルスの特徴分析を行う．特にここでは簡単のため，KS を考慮した確率モデルについてのみ取り扱うこととする．

マルコフ連鎖に基づいた解析を行う場合，状態空間の大きさが計算時間に強く影響を及ぼすことは周知の事実である．これは，状態空間が大きくなると非常に大きな行列に対する演算を必要とするからにほかならない．本論文においても，3 章で示したマルコフ解析を適用するために，状態空間の大きさ，すなわち総端末台数の値の設定に注意をはらう必要がある．図 3 から分かるように，BLAID あるいは DATOM の被害にあった総端末数は 60 日間で 20,000 件以下である．ここでは，状態空間の減少のために，1 単位の端末台数を 1,000 台とし，$K = 20$ とした．このように 1 単位の端末台数を調整することで，得られる評価尺度に大きな影響を与えることなく感染端末台数の上限（パラメータ K）を決定することができる．次に，KS の増加率と減少率を決定する．本論文で対象としたワームは，KLEZ など過去に大きな被害をもたらしたワームと比較して，一般的にあまり知られていないため，他端末からの警告による KS 増加が極端に少ない．そこで，KS の増加率に関して $\beta_r = 0$ とし，KS の増加はウィルスの感染・除去によるもののみであると仮定する．さらに，KS の減少率に関しても $\delta_r = 0$ とした．この理由として，現実的に KS を受け取った端末（ウィルス感染した後に修復された端末）が KS を失うために要する期間はウィルスが感染するために必要な期間と比較して非常に長い期間を必要とするが，ここでは 60 日間という短い期間でのデータを取り扱っているため，KS の減少は 60 日間中で起こらないと仮定した．また，$n_0 = 2$ とした．これらのパラメータ設定は得られる評価尺度に影響するが，ここでは 2 種類のウィルスの特徴比較に着目しているのでパラメータに対する仮定は 2 種類のウィルスに対して同じ値であれば十分であることに注意する．

上述の仮定のもとで，ウィルスの感染率 β とウィルス除去率 δ を推定し，ウィルスの挙動に関する定量的な評価尺度の導出を行う．ウィルス感染率 β とウィルス除去率 δ の推定に関しては計算手続きの簡略化のため，総被害件数の期待値と観測期間中の総被害件数との二乗誤差を最小にする最小二乗法を用いた．いま，N_k, $k = 1, \ldots, 60$ をマルコフモデル上における経過日数 k での総被害件数を表す確率過程とすると，実測データとの二乗誤差

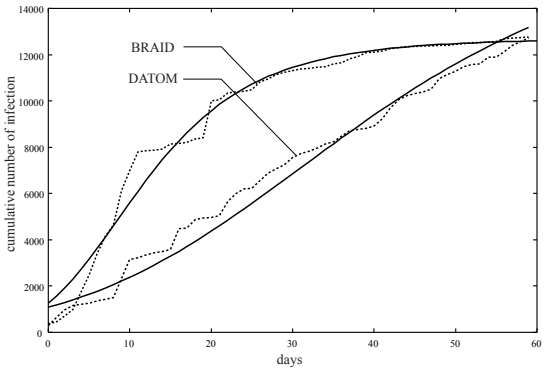


図 4 実測データと推定値の総被害件数による比較
Fig. 4 Comparison between real data and estimates in terms of the cumulative number of infection.

表 1 BLAID と DATOM に対する評価尺度の比較結果
Table 1 Comparison of measures for BLAID and DATOM.

	MTTE	MTTH(10)	MTTH(14)
BLAID	41.23	14.63	14.51
DATOM	273.94	51.64	44.10

$$\Theta(\beta, \delta) = \sum_{k=1}^{60} (\mathrm{E}[N_k] - n_k)^2 \tag{52}$$

を最小にする $\hat{\beta}$, $\hat{\delta}$ を推定量として定義する．ここで，n_k は第 k 日目における総被害件数である．いい換えると，パラメータ推定は

$$\begin{aligned} &\text{minimize} && \Theta(\beta, \delta) \\ &\text{subject to} && \beta > 0, \ \ \delta > 0 \end{aligned} \tag{53}$$

の非線形計画問題に帰着される．この問題を解くために本論文では準ニュートン法[10] を用いた．

下記は，実際に 2 種類のコンピュータウィルス BLAID と DATOM に対して観測された 60 日間の総被害件数データから，最小二乗法によって推定されたウィルス感染率とウィルス除去率である．

$$(\beta, \delta) = (1.3276 \times 10^{-2}, 8.2045 \times 10^{-2}),$$

BLAID：

$$(\beta, \delta) = (4.6716 \times 10^{-3}, 1.2927 \times 10^{-2}).$$

DATOM：

また，**図 4** はマルコフモデルで推定したパラメータを用いたときの総被害件数の期待値（実線）と実測した総被害件数（破線）を比較したものである．この図において，推定された期待総被害件数は実際の総被害件数の振舞いをよく表現していることが分かる．

次に，2 種類のウィルス BLAID と DATOM に対する推定値からマルコフモデルに基づいた定量的評価尺度の導出を行い，ウィルスの挙動に関する比較を行

表 2 BLAID と DATOM において臨界レベルに到達する確率
Table 2 Hazard probabilities of BLAID and DATOM.

	$c = 10$	$c = 14$
BLAID	0.357	0.049
DATOM	0.819	0.477

う．**表 1** は，BLAID と DATOM に対する MTTE，MTTH(10)，MTTH(14) の値を示している．ここで MTTH(10) と MTTH(14) は，臨界レベルを 10,000 台あるいは 14,000 台としたときの MTTH の値を示している．また，**表 2** は BLAID と DATOM それぞれが臨界レベルに到達する確率を表したものである．

表 1 から，BLAID は DATOM よりもウィルスが死滅するまでの平均時間が短く，ハザードが発生するまでの平均時間 MTTH(10) と MTTH(14) に関しても BLAID の方が短いことが見てとれる．これは，BLAID の方が感染力は高いが継続力が低いことを意味している．逆に，DATOM は感染力が低く継続力が高いことが分かる．また，表 2 から，BLAID は DATOM よりも感染端末台数が臨界レベルに到達する確率が低いことが分かる．すなわち，ウィルスの性質が MTTE，MTTH および臨界レベルへの到達確率の 3 つの評価尺度のバランスによって特徴付けられることが分かる．特に，これら 2 種類のウィルスは 60 日間の総被害件数がほぼ同等であることから，ウィルスの感染率のみの情報によってウィルスの被害規模の影響を測ることはできないことが分かる．また，これらの結果から,（期待値の意味で）40 日程度で被害が終息する BLAID よりも，感染力は低いものの除去が困難な DATOM に対して予防策を注力することで，全体的な被害件数の低減を図ることが可能である．

ここで示した量は 2 種類のウィルスを同環境（初期感染端末台数，感染端末台数の上限）において算出した結果であり，相対的に 2 つのウィルスの特徴をよく表している．しかしながら，より多くの種類のウィルスを比較する場合や，MTTH，MTTE の精度（どのくらい現実に即しているか）を追求する場合は，初期感染端末台数や感染端末台数の上限などのパラメータを慎重に決定する必要がある．

5. むすび

本論文では，コンピュータウィルスの感染・除去の挙動に関する定量的な評価を行った．特に，コンピュータウィルスの感染・除去に付随する不確実性を考慮して，従来の微分方程式に基づいたコンピュータウィルスの解析では困難であった定量的評価尺度の導出を行った．また，具体的にウィルス感染端末台数をマルコフ過程によって表現する確率モデルを構築し，KS を考慮しない場合と KS を考慮した場合の 2 つの環境に対して解析的に評価尺度の導出を行った．数値例では，現存する 2 種類のウィルスに関する総被害件数データから，確率モデルのパラメータ推定を行い，本論文で提案した評価尺度に基づいた特徴分析を行った．

今後は，MTTH や MTTE の精度を高めるために感染端末台数の上限に対する決定方法や，最尤推定によるパラメータ推定を行う．また，それらの結果に基づいて他のコンピュータウィルスに対する被害件数データから MTTH および MTTE の導出を行い，より広範にウィルスの特徴分析を行う予定である．

謝辞 本研究の一部は文部省科学研究費若手研究（B）Grant No.15700060（2003–2004），萌芽研究 15651076（2003-2005）および基盤研究（B）Grant No.13480109（2001–2004）による助成のもとで行われたものである．

参考文献

1) Okamoto, T. and Ishida, Y.: A distributed approach to computer virus detection and neutralization by autonomous and heterogeneous agents, *Proc. 4th International Symposium on Autonomous Decentralized Systems*, pp.328–331 (1999).
2) Badhusha, A., Buhari, S., Junaidu, S. and Saleem, M.: Automatic signature files update in antivirus software using active packets, *Proc. ACS/IEEE International Conference on Computer Systems and Applications*, pp.457–460 (2001).
3) Thimbleby, H., Anderson, S. and Cairns, P.: A framework for modelling Trojans and computer virus infection, *The Computer Journal*, Vol.41, No.7, pp.445–458 (1998).
4) Kephart, J.O. and White, S.R.: Directed-graph epidemiological models of computer viruses, *Proc. 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pp.343–359 (1991).
5) Kephart, J.O. and White, S.R.: Measuring and modeling computer virus prevalence, *Proc. 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pp.2–15 (1993).
6) Kephart, J.O.: A biologically inspired immune system for computers, *Proc. International Joint Conference on Artificial Intelligence*, pp.20–25 (1995).
7) 豊泉 洋，加羅 淳：プレデター型のウィルス撃退手法,「待ち行列理論とその応用：未来への展

望」シンポジウム報文集，pp.21–30 (2002).
8) Toyoizumi, H. and Kara, A.: Predators: good will codes combat against computer viruses, *ACM SIGSAC New Security Paradigms Workshop 2002* (2002).
9) Wang, C., Knight, J.C. and Elder, M.C.: On computer viral infection and the effect of immunization, *Proc. 16th Annual Computer Security Applications Conference*, pp.246–256 (2000).
10) 坂和正敏：非線形システムの最適化，森北出版 (1986).

付　　録

A.1 MTTH（MTTE）の導出

ここでは式 (39)，式 (40)，式 (44)–(46) の導出を行う．他の同様な式 (19)–(23) および式 (25)–(27) においても基本的な考え方は同じである．

いま，ウィルス感染端末台数が n の状態でウィルスが死滅することなく感染端末台数が $n+1$ へ移行する確率 $\boldsymbol{F}_n$ は，$\boldsymbol{A}_n$，$\boldsymbol{B}_n$，$\boldsymbol{C}_n$ がそれぞれ感染端末台数が 1 台減る，感染端末台数がそのまま，感染端末台数が 1 台増える確率を表す行列であることを利用して，

$$\boldsymbol{F}_n = \boldsymbol{A}_n\boldsymbol{F}_{n-1}\boldsymbol{F}_n + \boldsymbol{B}_n\boldsymbol{F}_n + \boldsymbol{C}_n \qquad (54)$$

となる．これを $\boldsymbol{F}_n$ に関して解くと式 (40) が得られる．同様な手順で，ウィルスが死滅することなくウィルス感染台数が n から $n+1$ へ増加するまでの期待時間 $\boldsymbol{M}_n$ を考えると

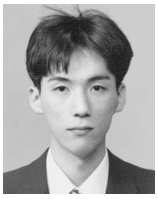
$$\begin{aligned}\boldsymbol{M}_n = {} & \tilde{\boldsymbol{A}}_n\boldsymbol{F}_{n-1}\boldsymbol{F}_n + \boldsymbol{A}_n\boldsymbol{M}_{n-1}\boldsymbol{F}_n \\ & + \boldsymbol{A}_n\boldsymbol{F}_{n-1}\boldsymbol{M}_n + \tilde{\boldsymbol{B}}_n\boldsymbol{F}_n \\ & + \boldsymbol{B}\boldsymbol{M}_n + \boldsymbol{C}_n \qquad (55)\end{aligned}$$

を得る．これを解いて $\boldsymbol{M}_n$ に対する漸化式が得られる．また，$\boldsymbol{M}_n$ は条件付き期待値ではないため，式 (44) でウィルスが死滅しない確率で除算し，条件付き期待値へ変換している．MTTE の場合も同様である．

(平成 15 年 7 月 10 日受付)
(平成 16 年 3 月 5 日採録)

小林　尚志

昭和 55 年生．平成 15 年広島大学工学部第二類卒業．平成 15 年同大学大学院工学研究科情報工学専攻博士課程前期入学．現在，システムセキュリティに関する研究に従事．



岡村　寛之（正会員）

昭和 48 年生．平成 7 年広島大学工学部第二類卒業．平成 9 年同大学大学院工学研究科システム工学専攻博士課程前期修了．同年 CSK（株）入社．平成 10 年広島大学工学部助手．平成 15 年広島大学大学院工学研究科助教授．博士（工学）．主として，待ち行列システム，信頼性・保全性システムの研究に従事．日本 OR 学会，電子情報通信学会，IEEE 各会員．

土肥　　正

昭和 40 年生．平成元年広島大学工学部第二類卒業．平成 4 年同大学大学院工学研究科システム工学専攻博士課程後期中途退学．同年広島大学工学部助手．平成 8 年同助教授．平成 14 年広島大学大学院工学研究科教授．博士（工学）．平成 4 年ブリティッシュ・コロンビア大学（カナダ）客員研究員．平成 12 年デューク大学（アメリカ）客員研究員．主として，信頼性理論および数理システムの研究に従事．日本 OR 学会，日本応用数理学会，電子情報通信学会，システム制御情報学会，計測自動制御学会，IEEE 各会員．