様式 C-19

科学研究費補助金研究成果報告書

平成２１年　５月１８日現在

研究種目：若手研究（B）
研究期間：２００７～２００８
課題番号：１９７００００５
研究課題名（和文）　インターネットワームに対する総合的セキュリティモデルの実装

研究課題名（英文）　Impremention of a comprehensive security model against Internet worms

研究代表者
岡村　寛之（HIROYUKI OKAMURA）
広島大学・大学院工学研究科・准教授
研究者番号：１０３１１８１２

研究成果の概要：PC に対するインターネットワームの問題は深刻な社会現象として広く認知されている．本研究ではインターネットワームの爆発的な繁殖予防を目的として，新種のワームが認知された直後（初期認知期間）における繁殖予測モデルの構築を行う．具体的には，確率統計を基礎としたモデリングを行う．一般に生態学における繁殖モデルは比較的長時間にわたる観測に対するデータに基づく場合が多い．しかしながら，初期認知期間においてインターネットワームの危険度を評価する場合，利用可能な情報が限定されている．そこで，本研究ではインターネットワームを利用される脆弱性の観点から同時に評価することを考える．つまり脆弱性の分類とその影響範囲をなんらかの確率モデルで表現した後に，その脆弱性をエサとするワームの生態を表現することによって，少数のデータに対しても高い精度で危険性を評価可能なモデルの構築およびデータからの数値予測を行う．

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	1,100,000	0	1,100,000
２００８年度	700,000	210,000	910,000
年度			
総　計	1,800,000	210,000	2,010,000

研究分野：総合領域
科研費の分科・細目：情報学・計算機システム・ネットワーク
キーワード：安全性・信頼性，インターネットワーム，繁殖予測モデル

１．研究開始当初の背景

情報通信技術（IT）の有効活用は現在我が国の重要な国策の一つである．平成 18 年にはこれまでの「e-Japan 戦略」を引き継ぐ形で「重点計画-2006」が策定され，IT 技術の有効活用という点により重点をおいた対策を行っている．これは IT 構造改革力と IT 基盤の整備と言う大きく二つの項目に分類されており，IT 基盤の整備において「セキュア・ジャパン 2006」と「高セキュリティな次世代 OS 環境の実現」が具体的施策としてあげられている．つまり，IT 活用を支える基盤技術において情報セキュリティの確保は，ネットワークコンテンツの充実やユビキタスネットワークを実現する上で避けて通れない問題となっている．ネットワーク上のセキュリティを脅かす要因の一つにコンピュータウィルスがある．コンピュータウィルスは 1984 年に F. Cohen が「自己増殖プ

ログラム」の研究成果を発表した際に使用した用語がその語源となっており，これ以後，不正を行うプログラムの総称となっている．より詳細には，インターネットワーム型，トロイの木馬型などの分類が行われており，近年のネットワーク環境において，もっとも脅威となっているのがインターネットワーム型（以下，ワーム）と呼ばれるコンピュータウィルスである．

ワームは単体で機能する自己複製プログラムであり，Web ブラウザやメーラのセキュリティホール（悪用される可能性のある欠陥）を利用して繁殖する．悪名高いワームは，2001 年に発見された Code-Red や 2003 年の SQL Slammer, MSBlaster, 2004 年の Netsky などである．例えば，Netsky は 2004 年の 3 月から 4 月の数週間で 10,000 件以上の感染報告があがっている．この 1 年程は大きく目立ったウィルスが存在しないが，これは世界全体で利用されている Windows の現バージョンが比較的枯れたソフトウェアとなっていることが起因しており，2007 年の新バージョン発表とともに，再び強力な繁殖力を持つウィルスの登場が予想される．また，ワームの幾つかは，個人情報を流出や PC 内部のデータ（ハードディスク）損失などの破壊活動を行うため，ワームの繁殖を防止することによって実現されるセキュリティの確保は現在においても非常に重要な役割を担っている．

ワームの繁殖は OS あるいはアプリケーションのセキュリティホール（近年ではより広い定義の「脆弱性」としている）を利用している．つまり，PC の管理者が定期的に OS あるいはアプリケーションのベンダーが提供する修正プログラム（パッチと呼ばれる）をチェックすることで，簡単にワームの繁殖を防ぐことができる．しかしながら，現実的に日々多くの修正パッチが発表されるため，管理者の限られた労力ですべての PC の修正パッチをチェックすることは不可能である．また，修正パッチの幾つかは適用することでシステムが不安定になることがあるため，修正パッチの公開と適用にタイムラグが生じる．現実的にワームの繁殖はこの期間で起こっていると言っても過言ではない．

この様な現状を鑑みて，近年では，米国において総合的なセキュリティ対策の研究が行われている（例えば米 Trend Micro Inc. の White Paper などに見られる）．これは，ワームの被害が認知されてからからの対処療法的なアプローチではなく，様々な情報を取り入れることで感染の爆発（アウトブレイク）あるいはどのような種類のワームが今後繁殖するかを事前に予測し，予防を行うアプローチである．主として，脆弱性のタイプからの経験的な予測，あるいは警告を効率良く送信できるシステム作りと言った，ソフトウェア工学的観点から総合的セキュリティ対策を実現する試みである．

このような傾向とほぼ時を同じくして，申請者の所属する研究チームは平成 15 年度ー平成 17 年度萌芽研究「コンピュータウィルス撃退のためのセキュリティモデルの開発と性能評価」において，コンピュータウィルスに関する統計的な観点からの総合セキュリティモデルの構築を行った．これはソフトウェア工学的なアプローチとは異なり，統計的観点から総合的なセキュリティの実現を目指した研究活動である．この特徴は生態学的立場からワームの繁殖に関する数理モデルを構築し，その性質や繁殖力について議論するものであった．また実際の繁殖データを用いた実験においてウィルスの繁殖がある種の生態学におけるモデルによって記述できることが確認されている．特に，多くのウィルスに関してこのような統計的アプローチが実際に有効であることが示されている．

一方，現在では新たな問題が浮上している．上述したようにワーム繁殖に関する定性的な議論は生態学における議論である程度完結する．しかしながら，モデルを用いた将来予測となると，繁殖が収束する時点での予測に関する精度は高いが，新種のワームが認知された直後（初期認知期間）で得られている少数サンプルでは，モデルによる予測精度が極端に落ちてしまう．つまり，現在の繁殖モデルによる予測はワームが既にある程度繁殖した場合でなければ，その効果を十分に発揮できない状況にあり，安全性という立場において矛盾した状況を生み出している．現在では新種のワームが発見される初期認知期間において高精度な将来予測できるモデルの構築が要求されている．ところが，単にワームの感染傾向のみに主眼をおいた統計解析では利用できる情報が非常に限定されているため，要求される精度の予測を行うことは不可能である．つまり，現在の研究とは異なった観点からの情報利用が必要になることは明白である．

２．研究の目的

本研究では「初期認知期間のデータを用いた繁殖予測」と言う目標を達成するために，これまでの研究とは異なったアプローチを行う．具体的には，ワームの繁殖原因となる脆弱性とワーム繁殖の因果関係を統計モデルによって表現することによって，セキュアプログラミングなどのソフトウェア工学的な知見と統計解析に基づいたワームの生態的なふるまいを融合する．これは，ソフトウェア工学と統計学に対する横断的な知識が必

要とされる．この意味で申請者が他に先がけてこの研究に着手する意義があるものと考えられる．研究の全体像は，(i) 脆弱性とワームに関するデータの収集，(ii) 統計的な特徴分析とワーム繁殖要因の解析，(iii) ワーム繁殖モデルの構築，(iv) 経験ベイズによる少数データを用いた繁殖の推定，となる．

(i)では，脆弱性の種別（バッファオーバーフローなど）に対するデータと過去に認知されたワームの発生日時と発生数に関するデータの収集を行う．(ii)は収集したデータの解析であり，脆弱性の因子とワーム発生状況の因果関係をベイジアンネットワークなどの統計ツールによってモデル化する．(iii) は (i), (ii) の成果を統合する繁殖モデルの構築を行う．最終的に (iv)において，(iii) のモデルと実データ（因子や現在のワーム認知台数）を用いて，新種ワームの繁殖確率やアウトブレイクタイミングの推定を行う．これらの上記のアプローチの特徴は，ワームの繁殖データだけでなく繁殖因子を同時に考慮したことであり，これによって，非常に情報の少ない初期認知期間において高い精度の予測が実現できる．

３．研究の方法

（１）脆弱性とワームに関するデータの収集

過去に存在した脆弱性に関する情報を収集する．収集するデータソースとしては OS ベンダー，IPA，アンチウィルスソフトの作成会社におけるデータを利用する．その際，周辺情報として対象 OS やアプリケーション，あるいはどのような種類の脆弱性（バッファオーバーフローなど）があったか詳細に調べる．同時にそれらを悪用したワーム（他のウイルスも含む）の発生時期や被害状況のデータも収集する．データの調査に関しては，数千以上の脆弱性とワームについて調べる必要がある．また，調査したデータを保持するためのデータベースシステムを構築する．

（２）統計的な特徴分析とワーム繁殖要因の解析

収集したデータをもとに，脆弱性とワームの性質を特徴付ける統計モデルの構築を行う．ここでは，特に脆弱性とワームの因果関係に注目する．基礎となる統計モデルの候補としては，単純な重回帰モデルから，非線形回帰モデル，COX 回帰モデル，ニューラルネットワーク，ベイジアンネットワークを考えている．特に，COX 回帰モデルは本来生存時間解析を行う統計ツールであることから，ワームの繁殖時間と因子間の解析が期待される．また，脆弱性の質的な要因とワームの特徴に対する関係については，因果推論を行う統計ツールであるベイジアンネットを適用することで，理論的に矛盾のない解析および現象の説明が行えるものと考えられる．これは，対象とする OS やアプリケーションの種別と脆弱性の情報から新種ワームの発生率や繁殖率が直接算出できる統計モデルとなる．

上記の研究成果として，脆弱性の特徴とワーム繁殖の因果関係に関する統計的モデリングによる解析アプローチと実データを用いた解析結果を，国内の研究集会と国際会議において発表を行う．具体的には，IEEE International Symposium on Software Reliability Engineering や IEEE International Conference on Dependable Systems and Networks であり，これらの会議で，それまでの成果発表を行う．

（３）ワーム繁殖モデルの構築

先の作業において，OS，アプリケーション，脆弱性の種別から，新種ワームの発生率と繁殖率の算出ができたことを受けて，ここでは確率的なワーム繁殖モデルの構築を行う．特に，初期認知期間に特化した繁殖モデルと長期的な予測を行う二種類のモデル構築を目指す．初期認知期間の特徴は，アンチウィルスソフトウェアによる耐ワーム効果がないことである．そのため，生態学のモデルでは最も単純な指数的増加モデル，確率モデルでは分枝過程などが適用できると考えられる．初期認知期間における短期予測モデルを用いて，ワームの繁殖を評価する様々な確率的な尺度，具体的にはアウトブレイク確率や死滅確率を算出する公式を示す．これは，ワームの危険度を総合的に判断するための指標であり，これは脆弱性の危険度評価へと反映されることとなる．次に，長期的な予測モデルの検討を行う．これは，平成 15 年度ー平成 17 年度萌芽研究「コンピュータウィルス撃退のためのセキュリティモデルの開発と性能評価」で長期的な予測に対して効果のあった非線形回帰モデルや非同次ポアソン過程モデルを基礎とする．しかしながら，以下の手順で示すように少数サンプルからの推定が基本となるため，パラメータ自由度が少なくなるような工夫を施す．

（４）経験ベイズによる少数データを用いた繁殖の推定

ここでは，初期認知期間で得られる情報から先の手順で構築したモデルのパラメータ推定を行うことを考える．初期認知期間における短期予測モデルと長期予測モデルを区別したことから，ここでも同様に各モデルに対する推定手法の検討を行う．推定の基本的な技術は経験ベイズによる推定である．ワームの発生率や繁殖率と言ったパラメータを

算出する枠組みは構築されているが，実際のモデル適用状況で高精度の推定を行うには，現在得られているワーム発生データを用いた発生率や繁殖率の調整が必要である．本研究では，この調整をベイズ推定によって行う．つまり，少数サンプルを用いた推定に対して事前情報を用いて高精度な推定を行う．具体的には統計モデルから発生率と繁殖率の基本パラメータを算出する．次に統計モデルにおけるパラメータの確信度（重回帰モデルの重相関係数やベイジアンネットワークにおける確率）を用いて発生率と繁殖率に関する確率分布を作成する．これを事前分布とした経験ベイズによる推定を実行することで初期認知期間における推定の高精度化をねらう．そのため，その基礎となる事前分布の類推アルゴリズムと事前情報を用いたパラメータ推定アルゴリズムの開発が主な作業となる．

これらの成果も国内の研究会において発表する．また，国際会議での発表を行う．そこでは意見を集約し必要な検証を行った上で，学術雑誌（IEEE/ACM Trans. on Networking など）への投稿を行う．

４．研究成果

平成１９年度では，脆弱性評価の基礎となるモデル開発と，モデルに基づいたデータ分類に関する研究を行った．具体的には，非同次ポアソン過程に基づいたモデルのブラッシュアップと，マルコフ到着過程に基づいた評価モデルの構築，およびＷｅｂより収集したウィルス感染データ（過去２年間，１１６種類）のモデルへの適合度検定を行った．以下，それぞれの成果に関して述べる．

（１）非同次ポアソン過程モデルでは，混合系分布に従う感染を想定し，そのようなデータに対応できるようにモデルの拡張を行った．数理的観点から見た拡張モデルの優位点は，モデルの表現力が格段に向上したことに加えて，モデル選択が混合比によって容易に行える点にある．これはデータの背後にある数理的な関係をモデルよって説明することを容易にする．拡張モデルのデータフィッティングに対して EM アルゴリズムを提案し，そのアルゴリズムおよび実際に観測したデータに拡張したモデルを適用した事例について IEEE が主催するソフトウェア信頼性に関するトップクラスの国際会議 International Symposium on Software Reliability Engineering にて発表した．

（２）非同次ポアソン過程を超える枠組みとして，マルコフ到着過程の利用可能性を模索した．一般にウィルス感染データは日毎の感染数で与えられる（グループデータ）が，従来のフィッティング手法はこのような形式のデータに対応していなかった．そこで，従来の EM アルゴリズムの改良を行い，グループデータへの対応を行った．また，マルコフ到着過程において最も汎用的なクラスであるマルコフ変調複合ポアソン過程の提案およびそのフィッティングアルゴリズムを考案し，ACM が主催する国際会議で発表を行った．

平成 20 年度は，主として数理的なモデル化および統計手法の開発を行った．平成 19 年度の成果をうけて，非同次ポアソン過程モデルおよびより一般的な枠組みであるマルコフ到着過程を用いたウィルス増殖を予測・評価するための確率モデルを提案した．特に，ベイズ原理を用いた各モデルの推定手法の拡張，扱えるデータ形式の拡張を行った．具体的にマルコフ到着過程ではグループデータを効率的に扱うための推定手法を提案した．ベイズ推定の枠組みでは変分原理を適用した変分ベイズをもちいた推定アルゴリズムを構築した．これにより，無情報事前情報の利用と同時に，他のウィルス特徴量から推定される経験的な値をあつかう経験ベイズへの応用に展開することができる．同時に，多角的なウィルスデータ利用の可能性を有しており，従来の無情報事前情報から大きく予測精度を向上することが期待される．

提案した手法（推定アルゴリズムおよび評価アルゴリズム）を C++言語により実装し，UNIX 系 OS 上で利用可能なツールのプロトタイプを開発した．さらに，より大量のデータを高速にあつかうための工夫として平均場近似を援用した近似手法である一般化 EM アルゴリズムの適用可能性について検討した．これにより，膨大なビットパターンから統計的類似性を検出するアルゴリズムの基礎を構築した．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１０件）

1. Hiroyuki Okamura and Tadashi Dohi, Software reliability modeling based on capture-recapture sampling, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (A), E92-A, 2009 (掲載決定，査読有り)
2. Hiroyuki Okamura, Tadashi Dohi and Kishor S. Trivedi, Markovian arrival process parameter estimation with group data, IEEE/ACM Transactions on

Networking, 17, 2009 (掲載決定, 査読有り)

3. Koji Ohishi, Hiroyuki Okamura and Tadashi Dohi, Gompertz software reliability model: estimation algorithm and empirical validation, Journal of Systems and Software, 82, 535-543, 2009 (査読有り).
4. Hiroyuki Okamura and Tadashi Dohi, Software reliability modeling based on mixed Poisson distributions, International Journal of Reliability, Quality and Safety Engineering, 15(1), 9-32, 2008 (査読有り).
5. Hiroyuki Okamura and Tadashi Dohi, Hyper-Erlang software reliability model, Proc. 14th Pacific Rim Int. Symp. Dependable Computing (PRDC'08), 232-239, 2008 (査読有り).
6. Hiroyuki Okamura, Michael Grottke, Tadashi Dohi and Kishor S. Trivedi, Variational Bayesian approach for interval estimation of NHPP-based software reliability models, Proceedings of 2007 International Conference on Dependable Systems and Networks (DSN-2007), 698-707, 2007 (査読有り).
7. Hiroyuki Okamura, Kazuya Tateishi and Tadashi Dohi, Statistical inference of computer virus propagation using non-homogeneous Poisson processes, Proceedings of The 18th International Symposium on Software Reliability Engineering (ISSRE'07), 149-158, 2007 (査読有り).
8. Hiroyuki Okamura, Yuya Kamahara and Tadashi Dohi, Estimating Markov-modulated compound Poisson process, Proc. of the 2nd international conference on Performance evaluation methodologies and tools, 1-8, 2007 (査読有り).
9. Tomotaka Ishii, Tadashi Dohi and Hiroyuki Okamura, Bivariate software fault-detection models, Proc. of 31st Annual International Computer Software and Applications Conference (COMPSAC2007), 535-538, 2007 (査読有り).
10. Kazuya Shibata, Koichiro Rinsaka, Tadashi Dohi and Hiroyuki Okamura, Quantifying software maintainability based on a fault-detection/correction model, Proc. of 13th Pacific Rim International Symposium on Dependable Computing (PRDC'07), 35-42, 2007 (査読有り).

〔学会発表〕(計２件)

1. Ryutaro Fujimoto, Hiroyuki Okamura and Tadashi Dohi, Security evaluation of an intrusion tolerant system with MRSPNs, 4th International Conference on Availability, Reliability and Security, 2009 年 3 月 16-19 日, 福岡市
2. Yusuke Yamaguchi, Hiroyuki Okamura and Tadashi Dohi, Estimating a mixture of Erlang distributions based on variational Bayes, 2008 Asian International Workshop on Advanced Reliability Modeling, 2008 年 10 月 23-25 日, 台湾

６．研究組織

(1)研究代表者

岡村　寛之 (HIROYUKI OKAMURA)

広島大学・大学院工学研究科・准教授

研究者番号：１０３１１８１２

(2)研究分担者

(3)連携研究者