情報処理学会論文誌 Vol. 50 No. 3 1002–1011 (Mar. 2009)

# デジタルフォレンジックのための ワーム感染経路特定手法

稲場太郎[†1] 田原慎也[†1] 川口信隆[†1]
塩澤秀和[†2] 重野寛[†3] 岡田謙一[†3]

近年，ネットワークワームによる被害が多数報告されてきた．これに対し，脆弱性の補強や事後の法的手段，すなわちデジタルフォレンジックのためにワームの感染経路を特定する需要が高まっている．しかし，その感染経路を自動で特定する手法には多くの誤検知が存在してしまうのが現状であり，特に false positive と false negative とのトレードオフに悩まされている．そこで本論文では，視覚化システムを用いて自動アルゴリズムと人の手による解析の融合を実現し，感染経路特定を行う手法を提案する．自動アルゴリズムでは false negative がゼロとなるように解析を行い，その後解析者が視覚化システムを用いて false positive を削除する解析を行う．この 2 つの解析を融合することによりトレードオフを解消し，精度の高い経路特定を目指す．本提案手法の評価を行うためにプロトタイプシステムを実装し，ユーザ実験を行ったところ，自動アルゴリズムによって残された false positive が 90%削減され，視覚化システムによる解析の有効性が示された．

# Worm Path Identification for Digital Forensics

Taro Inaba,[†1] Shinya Tahara,[†1]
Nobutaka Kawaguchi,[†1] Hidekazu Shiozawa,[†2]
Hiroshi Shigeno[†3] and Kenichi Okada[†3]

In this paper, we propose a visualization system for worm investigation, which finds worm origins and worm paths. Although investigation of worms are very important for forensic use and further prevention, it is quite difficult for automatic systems to identify worm origins or paths due to the trade-off between false positives and false negatives. Therefore, we focused on interaction between analysts and connection logs. At first, an automated algorithm is run so that there are no false negatives, and then analysts investigate the result to reduce false positives by visualized system. We aim to solve the trade-off by conducting these two steps. We implemented a prototype and conducted a user experiment to evaluate our system. The results show our system enabled subjects to reduce 90% of false detection by an automated algorithm. Although the results depend on parameters or conditions, we show the effectiveness of our idea.

## 1. はじめに

近年，デジタルフォレンジックについて頻繁に議論されている[1),2)]．デジタルフォレンジックでは膨大なログを解析する必要や，裁判等の法的な場において解析結果を専門知識のない人々に提示する必要がある．そこで，作業効率の向上や分かりやすい解析結果の提示手法が必要になる．しかし，デジタルフォレンジックでの解析手順や証拠の提示方法等は体系化されていないのが現状である[1),3)]．一方で，近年ワームの被害が増大しており，ワームの感染手法は高度なものとなってきている[4)–6)]．したがって，デジタルフォレンジックはワームによる被害の分野でも必要となってくると考えられる．さらに，デジタルフォレンジックだけでなく，脆弱性を補強し，事後の被害を防止するためにもワームの感染経路を特定する需要は高まってきている[7)]．それに対し，感染経路を自動検知する手法の研究は多く行われているが，自動検知における誤検知の問題は依然として解決されていない[8),9)]．false positive（感染していないものを感染していると判断する誤検知）と false negative（感染しているものを感染していないと判断する誤検知）とのトレードオフに悩まされているのが現状である．一方，人の手による解析では誤検知の少ない高精度な結果を出すことが可能であるが，扱うことのできるログの量が限られ，実際のネットワークログをすべて人の手で解析することは現実的ではない．

そこで本論文では，ワーム感染経路を視覚化することによって自動アルゴリズムと人の手による解析の融合を実現し，経路特定を支援する手法を提案する．具体的には，まずワーム被害のあったネットワークログに自動アルゴリズムを適用し，false negative がゼロとなるような解析を行った後に視覚化システムを用いて人の手による解析作業を行う．この際，自動アルゴリズム適用後のログには false negative が含まれていないため，解析者は false

†1 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University
†2 玉川大学工学部
Faculty of Technology, Tamagawa University
†3 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University

positive のみを発見する作業を行えばよい．人の手によって false positive と false negative の両方を発見するのは非常に負担が大きい作業となるが，本提案手法では false positive の削除に限定することで精度の高い解析を可能としている．

また，解析作業終了後，最終的な解析結果は証拠として第三者に提示する必要があるが，視覚化された感染経路は誰が見ても一目で分かりやすいものとなっており，この点においても本手法はデジタルフォレンジックに貢献する．

以降の本論文の構成は以下のようである．2 章ではワームの被害におけるデジタルフォレンジックの位置づけについて述べ，3 章ではワーム感染経路の視覚化手法の提案を行う．4 章で本提案手法のプロトタイプシステムについて述べ，5 章では評価実験について述べる．6 章は本論文のまとめとする．

## 2. ワーム被害におけるデジタルフォレンジック

デジタルフォレンジックとは，「インシデント・レスポンスや法的紛争・訴訟に対し，電磁的記録の証拠保全および調査・分析を行うとともに，電磁的記録の改ざん・毀損等についての分析・情報収集等を行う一連の科学的調査手法・技術」と定義することができる[1)]．ワームの感染経路特定におけるデジタルフォレンジックでの作業手順は以下に示す 3 つのフェーズに分けることができる．

(1)　ログの保全
(2)　ログの解析
(3)　解析結果（証拠）の提示

まず，ワームによる被害が発覚したときに，ログの保全を行う．次に，感染ホストや感染経路の特定をするためにログの解析を行い，その結果を証拠として提示する．本研究は，この流れにおけるログの解析から解析結果の提示のフェーズを対象としている．

ログ解析のフェーズにおいては，自動検知アルゴリズムを用いてワームの感染ホスト，感染経路の検知が行われることが多い．しかしながら，自動検知は多くの誤検知を含み，特に false negative と false positive とのトレードオフに悩まされているのが現状である．

自動アルゴリズム以外の有効なログ解析手法の 1 つとして，視覚化ツールを用いた解析があげられる．視覚化は直感的に理解しやすく，そこに対して操作を行うことで，テキストベースのログを解析するよりも効率の良い解析作業が可能となる．既存のログの視覚化手法としては，グラフ化やマトリックス上に表示といった手法が主にあげられる[10),11)]．また，ワームの感染活動に特化した手法として，ノード探索の視覚化手法[12)] がある．この手法のようにワームの感染活動に着目した視覚化ツールを用いることによって，ワームの感染経路特定の解析作業の効率を上げることができる．

解析結果を提示するフェーズでは，解析結果の感染ホストや感染経路を専門知識のない人々に提示する必要がある．

以上のことをふまえ，本研究ではワームの感染活動の特徴に着目した視覚化手法を用いることにより，ワームの感染経路を特定し，デジタルフォレンジックに貢献することを目指す．

## 3. 視覚化による感染経路特定手法

### 3.1 概　　要

ネットワークログ解析を通じて自動でワーム感染経路を特定する手法の研究は多く行われているが，大量の誤検知がともなうのが現状である．特に false positive と false negative の間にはトレードオフの関係があり，両方を削減することは困難である．そこで我々は，自動アルゴリズムと人の手による解析の融合に注目した．自動化されたログ解析は低コスト，短時間で多量のログを解析できるという利点がある一方で，解析精度が良くないという欠点がある．高性能な自動アルゴリズムを搭載することで精度を向上させることはできるが，ネットワークログには多くの例外やノイズが含まれており，そのすべてを網羅したアルゴリズムを設計することは不可能である．これに対し，人の手による解析では解析できるログの数には制限がある一方で，自動アルゴリズムより高精度な解析を行えるという利点がある．例外やノイズを発見しやすいことに加え，インタビューやユーザの性格等を考慮した解析を行う，といったことは自動アルゴリズムには不可能なことである．これら双方の特徴を生かし，我々は次のような手順でワーム解析を行うことを提案する．

(1)　すべてのコネクションを感染疑惑コネクションとする．
(2)　自動アルゴリズムを用いて感染疑惑のないコネクションを検出し，削除する．
(3)　視覚化システムを用い，人の手によって感染疑惑のないコネクションを検出し，削除する．
(4)　視覚化システムを用い，解析結果をワーム感染経路として提示する．

手順 (1) の段階ではすべてのコネクションが陽性と判断されているため，true positive（感染しているコネクションを感染していると正しく判断）と大量の false positive が混在しているということになる．手順 (2) では，ここから false positive のみが削られる．この際，true positive のコネクションは絶対に削除してはならない．これは，後の人の手による解析を false positive の削除のみに限定し，容易にするためである．もし手順 (2) で true

positive を削除してしまった場合，false negative が発生することになる．本提案手法では，解析者の作業を false positive の削除のみに限定することによって精度の向上を図っているので，解析者は false negative 発見の作業は行わない．したがって，ここで false negative が発生してしまうとそのまま誤検知となり，精度が低下してしまう．false positeve 数が多少増加してもよいので，false negative が 1 つも出ないように自動アルゴリズムのパラメータを調整する必要がある．これが完全に達成された場合，手順 (2) 終了後のログにはすべての true positive と比較的少量の false positive が含まれていることとなる．手順 (3) では手順 (2) 終了後のログを視覚化し，人の手による解析が行われる．自動アルゴリズムによりログの絶対量は減少しているため，解析者が直接扱える程度の量になっている．ここでも false positive 削除の作業が行われ，最終的にはワームコネクションのみを残すことを目指す．手順 (4) では解析結果となるワームコネクションを視覚化することにより，第三者でも容易に感染経路が分かるようにする．

では自動アルゴリズムと視覚化手法について以下で述べる．

### 3.2　自動アルゴリズム

本提案手法に適用するアルゴリズムには，以下のことが求められる．

- 自動でワームの存在を検知できる．
- ログ解析の際，true positive は削除せず，false positive のみを削除する．

これを満たすアルゴリズムとしてはさまざまなものが考えられるが，その 1 つとして「ダミーアドレスを用いたワームの検知手法」[13] がある．このアルゴリズムは感染ホスト内部にあるアドレスリストから次なるターゲットを発見することによって感染活動を行うワームを対象とし，各ホストのアドレスリストの中にあらかじめ挿入されたダミーアドレスにコネクションが張られることによってワームの存在を検知する．さらにダミーアドレスに張られたコネクションからログを遡ることによって感染疑惑のあるホストを発見し，それらによってワームツリーを作り上げることができる．ここで遡る時間，回数を十分大きく設定することで，感染疑惑のあるコネクションをすべて網羅する，すなわち false negative コネクションをゼロとすることが可能になる．後述するプロトタイプにおいてはこのアルゴリズムを用いて実装を行った．

### 3.3　視覚化システム

本提案手法における視覚化システムは，解析者とネットワークログとのインタラクションを容易にするものでなくてはならない．また，最終的には第三者に対して結果を提示する必要があるので，誰の目にも分かりやすい視覚化手法が必要となる．

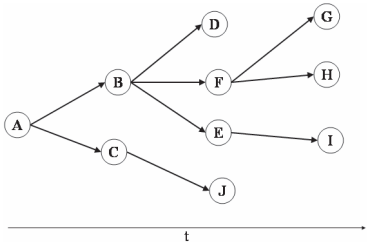


**図 1**　ワームツリー
Fig. 1　Worm tree.

**表 1**　コネクション例
Table 1　Examples of connections.

| No. | ソースホスト | ディスティネーションホスト | 時刻 |
|---|---|---|---|
| 1 | A | B | 1 |
| 2 | A | C | 2 |
| 3 | A | B | 3 |
| 4 | C | A | 4 |

これらを満たすよう，**図 1** に示すような「ワームツリー」の形で視覚化を行う．

多くのワームは 1 つの感染源から木構造の形で感染していくため，ホストをノード，コネクションをエッジとするグラフで感染活動を表現することで全体像の把握が容易となる．また，時間軸を横軸にとることで時間とともに感染の広がる速度や様子が一目で分かるようになっている．

しかし，ワームツリーは感染の広がりにのみ着目した表示手法であり，すべてのログを表示することはできない．なぜなら，1 つのホストは 1 度しか表示されないためである．たとえば，**表 1** のようなコネクションログがあったとする．この場合，ワームツリーで表されるコネクションは No.1 と No.2 の 2 つのみであり，No.3 と No.4 は表示されない．なぜなら，時刻 2 の段階でホスト A，B，C はすべて感染しており，これらのホストをディスティネーションとするコネクションは感染の広がりという着眼点からは必要ないためである．したがって，すべてのコネクションを解析するためには，ワームツリー以外の表示手法も必要となる．そこで，**図 2** に示すような単純な「ロググラフ」の形も採用することにする．横軸に時間，縦軸にホストをとり，矢印でコネクションを表すことによってすべてのコネクションが表示可能となっている．

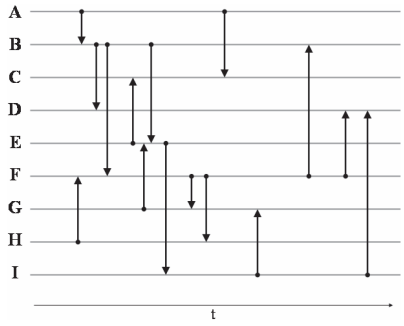


**図 2**　ロググラフ
Fig. 2　Log graph.

解析者はこの「ワームツリー」と「ロググラフ」の双方に対してインタラクションを行いながら false positive コネクションを削除していく作業を行う．たとえば，あるホストの使用者が席をはずしていた時間にコネクションが張られていたとしたらそのコネクションはワームによる可能性が高い，といったことや，よく通信を行う相手とコネクションの交換が多くなされていた場合はワームである可能性が低い，といったことが考えられる．また，ワームの感染活動の特徴に着目した解析も行うことができる．たとえばワームの感染速度がある程度判明した場合，その速度よりも明らかに速い速度で次々とコネクションを張っているホストはワームとは別の挙動を示しているものと考えられ，false positive としてワームツリーから削除することができる．このような作業を繰り返し，解析者は真のワームツリー構築を目指す．

## 4. プロトタイプシステム

本研究では，提案手法のコンセプトをもとにプロトタイプシステムの実装を行った．

### 4.1 対象とするワーム

本プロトタイプではホストが持つ内部アドレスリストを利用して感染活動を行うワームを対象とする．ここでいう内部アドレスリストとは，E メールソフトのアドレス帳，インスタントメッセンジャの登録アドレス，ARP キャッシュ，インターネットの接続履歴等，ホストマシンの内部に保存されているアドレスのことを指す．このワームは以下のような特徴を持つと仮定する．

- **感染速度**
  このワームに感染したホストは通常通信によるコネクション頻度と同程度の間隔で感染コネクションを張る．このようにすることにより，感染したホストの通常コネクションの中に感染コネクションを紛れ込ませ，感染コネクションを検知し難くする．
- **ターゲット**
  感染ホストの内部アドレスリストを参照し，その中からランダムに次なるターゲットホストを発見する．したがって，すでに感染したホストにコネクションを張ることもありうる．

### 4.2 自動検知アルゴリズム

本プロトタイプによる解析の前段階として適用される自動検知アルゴリズムとしては，前述の「ダミーアドレスを用いたワームの自動検知手法」を用いた．この手法ではダミーアドレスへ張られたコネクションからログを遡ることによって感染疑惑のホストを発見し，それらで構成されるワームツリーを作成することができる．このアルゴリズムでは，遡るホップ数や遡る時間等のパラメータを調整することにより false positive と false negative のバランスを整えることができ，false negative をゼロとすることも可能となる．

### 4.3 視覚化ウインドウ

本プロトタイプのウインドウは**図 3** のようになる．上部のウインドウはワームツリーを表示し，下部ではコネクションログが直接表示されている．本論文では，前者をツリーウインドウ，後者をログウインドウと呼ぶこととする．

#### 4.3.1 ツリーウインドウ

**図 4** に示されるように，ツリーウインドウでは感染疑惑ホストをノード，感染疑惑コネクションをエッジとするグラフの形でコネクションログが表示される．1 つのコネクションで結ばれるノードのうち，左に示されるのもがソースホスト，右がディスティネーションホストである．横軸は時間軸となっており，ディスティネーションホストの位置がコネクションの張られた時刻を表す．たとえば図 4 におけるコネクション A は，時刻 1,897 秒においてホスト 158 からホスト 165 に張られたコネクションであると読み取ることができる．また，ホストの中には黒い長方形がついているものがあり，これがダミーアドレスを表している．したがってこれらに張られたコネクションは確実に感染していると考えることができ，解析の際のヒントとなる．

このようにワームの感染コネクション，感染ホストを感染ツリーとして表示することで，解析対象となるワームの感染活動の特徴を直感的に認識できる．このウインドウではこの感

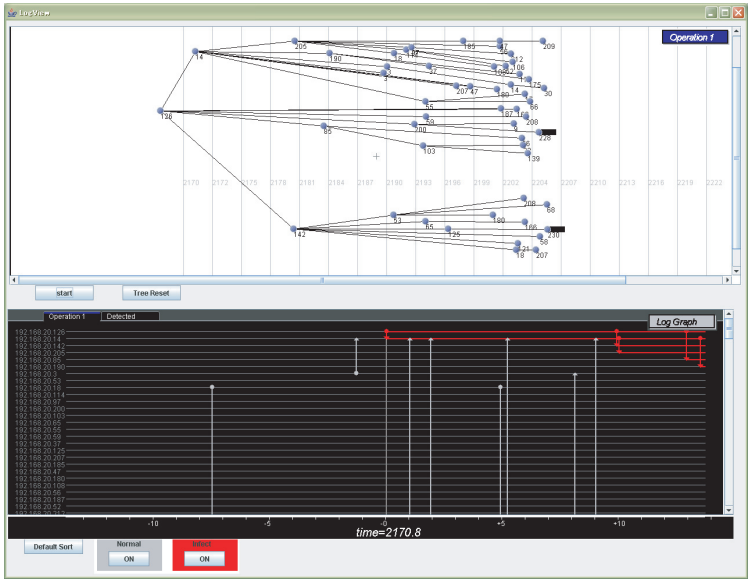


**図 3**　プロトタイプのウインドウ
Fig. 3　Window of prototype.

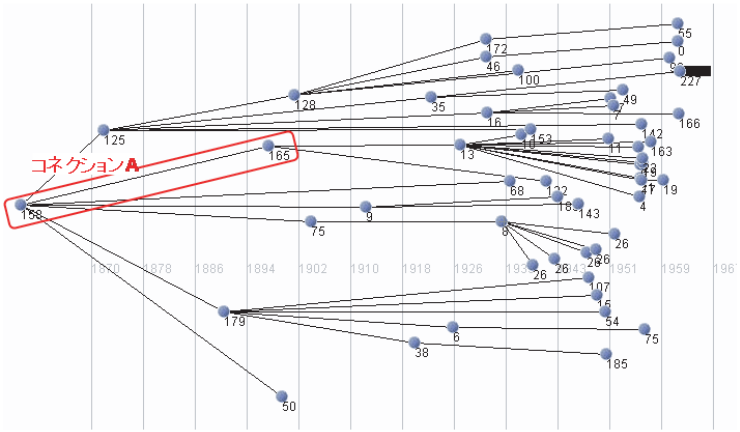


**図 4**　ツリーウインドウ
Fig. 4　Tree Window.

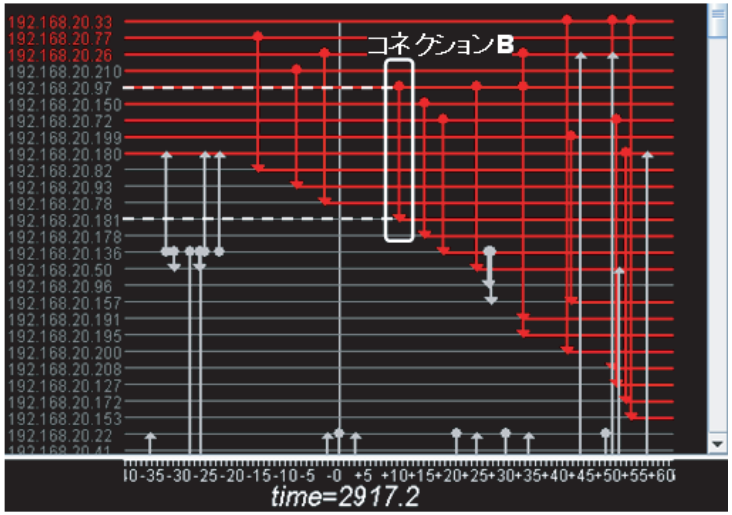


**図 5**　ログウインドウ
Fig. 5　Log Window.

染ツリーからコネクションを直接選択，削除といった直接的なインタラクションを可能としている．

また，図 4 は感染起源のホスト，感染経路，感染ホスト等が一目で認識できる形となっているため，第三者でも感染の広がりを把握しやすく，解析後の証拠提示の際にも有用な表示手法であるといえる．

### 4.3.2　ログウインドウ

本プロトタイプではワームツリーのほかにネットワークログを直接的に視覚化したウインドウも設けている（**図 5**）．このウインドウにはすべてのコネクションログが表示される．横軸は時間軸，縦軸はホストであり，左側にホスト名が表示されている．コネクションは矢印で表され，根元がソースホスト，先がディスティネーションホストとなっている．また，通常通信と判断されているコネクションは灰色，感染疑惑があるコネクションは赤で表されている．ツリーウインドウではこのうち赤いコネクションのみが表示されていることとなる．たとえば図 5 中のコネクション B に着目すると，時刻約 2,927 秒にホスト 97 からホスト 181 に張られたコネクションであることが分かる．

このウインドウではホストのソートや時間軸の伸縮が可能である．ログに対する操作はど

ちらのウインドウでも可能であり，その操作結果は両方のウインドウに反映される．したがって，解析者はこれらのウインドウを交互に用いて解析を行うことができる．

## 5. プロトタイプによる評価

本提案手法がワームの経路特定に有効であることを示すために，プロトタイプシステムを用いてユーザ実験を行い，評価を行った．

### 5.1 実験方法

はじめに被験者に対してワームの特徴や本プロトタイプの使用方法を説明し，その後，被験者が本提案手法を用いてネットワークログの解析を行った．また，本提案手法（ツリーウインドウとログウインドウ）を用いる場合とログウインドウのみ用いる場合での比較を行った．本実験は情報工学を専攻する大学生，大学院生 18 人を対象に行った．

#### 5.1.1 ネットワークログ

使用したネットワークログは，4.1 節で述べたワームがエンタプライズネットワーク内で繁殖活動を行った場合を想定したネットワークのコネクションログである．このログはコンピュータシミュレーションによって作成した．具体的には，DARPA の IDS 評価用のネットワークログ[14] を通常通信とし，そこにシミュレーションによるワームの感染コネクションを混ぜ込むことで作成した．なお，シミュレーションにおける全ホスト台数は 200 台，作成された全体のログは平均 5,000 個であった．その後このログに自動検知アルゴリズムを適用し，ログ全体を通常コネクションと感染疑惑コネクションに分類しなおした．自動検知アルゴリズムとしては 4.2 節で述べたダミーアドレスを用いた手法を適用し，false negative コネクションが 1 つも存在しないように各パラメータを設定した．

使用したコネクションログの例を**表 2** に示す．ここで “suspicious” パラメータが 1 であるということは感染疑惑コネクションを示し，“infection” パラメータは実際に感染していることを表す．もちろん，解析を行う際には被験者は “infection” パラメータは確認できないことになっている．

ワームの感染シミュレーションは複数回行い，以上のようなネットワークログセットを複数個作成した．その際には，ワームの感染活動開始時刻，感染速度等のワーム感染活動シミュレーションのパラメータはランダムに設定した．そして，その中から false positive コネクションの数が約 10 個のログ（パターン A），約 20 個のログ（B），約 40 個のログ（C）をそれぞれ 6 パターンずつ抽出し，計 18 パターンのネットワークログを本実験に用いた．なお，比較が行いやすいよう，感染疑惑コネクションの総数（ワームコネクション数＋ false positive コネクション数）が約 60 個となっているもののみ抽出した．したがって，各ログセットの感染疑惑コネクションの内訳は**表 3** のようになる．

**表 2**　コネクションログの例
Table 2　Examples of connection logs.

| No. | time | source host | destination host | infection | suspicious |
|---|---|---|---|---|---|
| 1 | 3.5 | 192.168.10.101 | 192.168.10.92 | 0 | 0 |
| 2 | 4.7 | 192.168.10.106 | 192.168.10.88 | 0 | 1 |
| 3 | 4.9 | 192.168.10.88 | 192.168.10.12 | 0 | 1 |
| 4 | 5.4 | 192.168.10.73 | 192.168.10.22 | 0 | 0 |
| 5 | 6.0 | 192.168.10.103 | 192.168.10.35 | 1 | 1 |
| 6 | 6.8 | 192.168.10.103 | 192.168.10.67 | 0 | 1 |

**表 3**　ログセットの種類
Table 3　Log sets.

| | タイプ A | タイプ B | タイプ C |
|---|---|---|---|
| FP コネクション数 | 10 | 20 | 40 |
| ワームコネクション数 | 50 | 40 | 20 |
| 感染疑惑コネクション数 | 60 | 60 | 60 |

本実験では，被験者 1 人につき 18 パターンのログセットから 6 パターンランダムに選択し，ログウインドウのみを用いる場合とツリーウインドウも同時に用いる場合のそれぞれで解析作業を行った．

#### 5.1.2 評価項目

評価項目は以下の 4 点である．

(1)　解析作業後の false positive コネクションの個数
(2)　解析作業後の false negative コネクションの個数
(3)　人の手による解析作業時間
(4)　解析作業時間と解析精度の相関関係

なお，false positive コネクションの個数は，解析作業において発見できずに残った個数である．また，false negative コネクションの個数は，解析作業において，本当は感染コネクションであるコネクションを誤って false positive コネクションと判断してしまった個数である．評価項目 (4) における解析精度は，以下の式で定義されるものであり，1 に近いほど精度が良いことを表す．

**表 4**　作業後の false positive コネクションの数
Table 4　Number of false positives after analysis.

| | FP コネクションの数（個） | | | |
|---|---|---|---|---|
| ログパターン | A | B | C | 平均 |
| 作業前 | 7.7 | 22.2 | 35.8 | 21.9 |
| ログウインドウのみ | 4.3 | 9.8 | 22.1 | 12.1 |
| ツリーウインドウ＋ログウインドウ | 1.7 | 0.6 | 3.2 | 1.8 |

$$P = 1 - \frac{FP_a + FN_a}{TP_b + FP_b} \tag{1}$$

$P$：解析精度
$FP_a$：解析後の false positive
$FN_a$：解析後の false negative
$TP_b$：解析前の true positive
$FP_b$：解析後の false positive

### 5.2　実験結果と考察

評価項目 (1)，false positive についての結果を**表 4** に示す．作業前の false positive コネクションの平均個数は 21.9 個であり，ログウインドウのみを用いた解析後は 12.1 個，ツリーウインドウも同時に用いた解析後は 1.8 個であった．このことより，ログウインドウのみを用いた場合より，ツリーウインドウも同時に用いた場合のほうが，ネットワークログの解析作業によって false positive コネクションを大幅に削減できることが分かる．両方のウインドウを用いての解析では，false positive コネクションの個数を作業前の個数の 1 割にまで削減できている．

また，実験に用いたネットワークログの false positive コネクションの個数の違いに注目する．ログウインドウのみを用いた場合では，作業前の false positive コネクションの個数が，作業後の個数に影響している．これは，ログウインドウを用いてもワームの感染活動の特徴が見えにくく，解析作業が困難であったためである．一方，両方のウインドウを用いた場合では，作業前の false positive コネクションの個数に関係なく，解析作業によって false positive コネクションを大幅に削減できている．この理由としては，ログをツリー状に表示することによって，全体のコネクションの流れが把握しやすくなっているためと考えられる．

このことより，本提案手法を用いた解析作業は，作業前の false positive コネクションの個数に関係なく，false positive コネクションの個数を大幅に削減可能であるといえる．

**表 5**　作業後の false negative コネクションの数
Table 5　Number of false negatives after analysis.

| | FN コネクションの数（個） | | | |
|---|---|---|---|---|
| ログパターン | A | B | C | 平均 |
| 作業前 | 0.0 | 0.0 | 0.0 | 0.0 |
| ログウインドウのみ | 1.4 | 0.3 | 0.7 | 0.8 |
| ツリーウインドウ＋ログウインドウ | 1.3 | 1.3 | 0.6 | 1.1 |

**表 6**　解析作業時間
Table 6　Analysis time.

| | 作業時間（秒） | | | |
|---|---|---|---|---|
| ログパターン | A | B | C | 平均 |
| ログウインドウのみ | 215.3 | 213.6 | 195.9 | 208.3 |
| ツリーウインドウ＋ログウインドウ | 431.3 | 458.8 | 547.7 | 479.3 |

次に，**表 5** に示す false negative コネクションの数についての考察を述べる．本実験で用いた自動検知アルゴリズム[13] のパラメータのチューニングにより，作業前の false negative コネクションの個数はすべて 0 個である．作業後の false negative コネクションの個数は，ログウインドウのみを用いた場合も提案手法を用いた場合も 1 個程度であった．また，作業後の false negative コネクションの数は，作業前の false positive コネクションの個数とは関係がない．

続いて，**表 6** に示す解析作業時間を比較する．各ログセットのパターンに対してはそれほど解析作業時間の変化は見られないが，ログウインドウとツリーウインドウを同時に用いた場合にはログウインドウのみの場合の 2 倍の作業時間を必要としている．しかしながら，ログウインドウのみを用いた場合の解析時間が短くなっているのは表 4 にあるように false positive コネクションを見つけきれていないためであり，解析精度を犠牲にしての時間短縮であると考えられる．また，両ウインドウを用いた場合でも平均解析時間は 500 秒，すなわち 8〜9 分程度となっており，現実的な解析にたえうる時間であるといえる．

最後に，被験者単位での作業時間と解析精度の相関関係を**図 6** に示す．なお，図中の白い丸印はログパターン B に対してログウインドウのみを用いて解析した場合であり，黒い四角印はログパターン B に対してログウインドウとツリーウインドウの両方を用いた場合

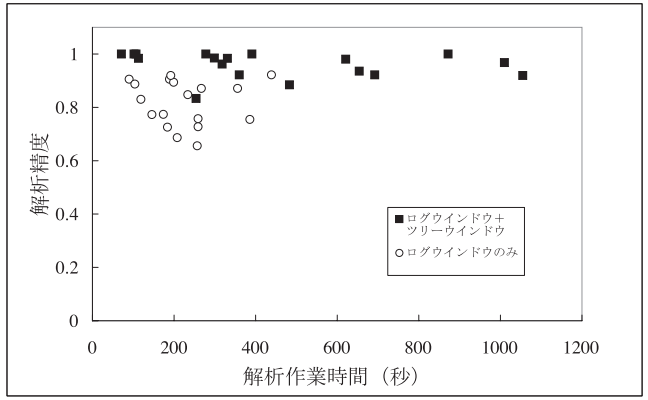


**図 6**　ログパターン B を用いた際の解析作業時間と解析精度の相関図
Fig. 6　Correlation diagram between analysis time and accuracy.

である．この図からもやはり，ログウインドウのみを用いた場合よりも両方のウインドウを用いた場合のほうが解析時間は平均的にかかっているが，解析精度が良いことが分かる．また両ウインドウを用いた場合に注目すると，解析時間は被験者によって大きなばらつきがあるが，そのばらつきが解析精度には影響を与えておらず，解析精度は高いレベルで安定していることが読み取れる．したがって，本提案手法では解析速度は作業者によってばらつくが，安定して高レベルの解析精度を得られる手法であるといえる．解析時間に関しても一番遅い被験者でも 1,200 秒，すなわち 20 分以内に解析を終了しており，本プロトタイプの実験においては現実的に解析可能なログの量であったといえる．

以上のことより，本提案手法を用いることにより，ワームの感染コネクションの特定が容易となることが分かった．感染コネクションが特定されれば，感染ホストと感染経路も明らかになる．したがって，コネクションをワームツリーとして視覚化する本提案手法はワームの感染経路特定に有用であるといえる．

## 6. 結　　論

本論文では，視覚化システムを用いたワーム感染経路特定手法を提案した．本提案手法では，自動アルゴリズムによる解析と人の手による解析を融合することにより，精度の高い経路特定を目指している．自動アルゴリズムは false negative がゼロとなるように解析を行い，その結果に対して解析者は false positive を削減する作業を行う．この 2 つを融合することにより双方の負担が減り，false positive と false negative のトレードオフの解消が可能となる．人の手による解析作業はワームツリーとロググラフが表示される視覚化システムを用いて行われ，解析者はログに対して直接的なインタラクションを行いながら false positive 削除の作業を行う．特にワームツリーは感染の広がる様子を直感的に表しており，解析の際の大きなヒントとなる．

本提案手法の有効性を示すために，プロトタイプシステムを実装し，ユーザによる評価実験を行った．自動検知アルゴリズム適用後のログを被験者が解析したところ，残された false positive の 90%を削減し，精度の高い結果が得られた．したがって，本提案手法はワームの感染経路特定に有効であることが示された．

**謝辞**　本研究の一部は，文部科学省科学研究費補助金（C）課題番号 1850063（2006 年），ならびにセコム科学技術振興財団の支援により行われた．

## 参　考　文　献

1) 辻井重男，萩原栄幸：デジタルフォレンジック辞典，デジタルフォレンジック研究会 (2006).
2) 佐々木良一：police　第 3 回セキュリティ解説　コンピュータ・フォレンジックス，警視庁（オンライン）．入手先 http://www.cyberpolice.go.jp/column/explanation03.html (参照 2008-12-17)
3) 佐々木良一，芦野祐樹，増渕孝延：デジタル・フォレンジックの体系化の試みと必要技術の提案，電子情報通信学会 SCIS2006 概要集，p.136 (2006).
4) Symantec.com: W32.Reztrictmm, Symantec.com (online). available from http://www.symantec.com/ja/jp/security_response/writeup.jsp?docid=2007-061115-3248-99 (accessed 2008-12-17)
5) ViruslistJP.com: Email-Worm.Bagle.gt, ViruslistJP.com (online). available from http://www.viruslistjp.com/viruses/encyclopedia/?virusid=145852 (accessed 2008-12-17)
6) Antonatos, S., Akritidis, P., Markatos, E.P. and Anagnostakis, K.G.: Defending against Hitlist Worms using Network Address Space Randomization, *Proc. 2005 ACM workshop on Rapid malcode*, pp.30–40 (2005).
7) Sekar, V., Xie, Y., Maltz, D.A., Reiter, M.K. and Zhang, H.: Toward a Framework for Internet Forensic Analysis, *Proc. ACM HotNets-III* (2004).
8) Xie, Y., Sekar, V., Maltz, D.A., Reiter, M.K. and Zhang, H.: Worm Origin Identification Using Random Moonwalks, *Proc. 2005 IEEE Symposium on Security and Privacy*, pp.242–256 (2005).
9) Xie, Y., Sekar, V., Reiter, M.K. and Zhang, H.: Forensic Analysis for Epidemic

Attacks in Federated Networks, *Proc. ICNP 2006*, pp.43–53 (2006).
10) Ma, K-L.: Cyber security through visualization, *Proc. Asia Pacific Symposium on Information Visualisation*, Vol.60, pp.3–7 (2006).
11) Lee, C.P. and Copeland, J.A.: Flowtag: A collaborative attack-analysis, reporting, and sharing tool for security researchers, *Proc. 3rd International Workshop on Visualization for Computer Security*, pp.103–108 (2006).
12) 仲小路博史，寺田真敏，洲崎誠一：ノード探索特性の可視化および定量化の提案，情報処理学会論文誌，Vol.48, No.9, pp.3163–3173 (2006).
13) Inaba, T., Kawaguchi, N., Tahara, S., Shigeno, H. and Okada, K.: Early containment of worms using dummy addresses and connection trace back, *Proc. 2007 International Confernce on Parallel and Distributed Systems*, Vol.2, pp.1–8 (2007).
14) MIT Lincoln Laboratory: DARPA Intrusion Detection Evaluation Data Sets, MIT (online). available from http://www.ll.mit.edu/IST/ideval/data/data_index.html (accessed 2008-12-17)

(平成 20 年 6 月 5 日受付)
(平成 20 年 12 月 5 日採録)

**稲場　太郎**（学生会員）
2007 年慶應義塾大学理工学部情報工学科卒業．現在同大学大学院理工学研究科修士課程在学中．ネットワークセキュリティ，デジタルフォレンジックスに関する研究に従事．

**田原　慎也**（学生会員）
2006 年慶應義塾大学理工学部情報工学科卒業．ネットワークセキュリティ，デジタルフォレンジックスに関する研究に従事．2008 年同大学大学院理工学研究科修士課程修了．2008 年日本 IBM（株）入社．

**川口　信隆**（正会員）
2005 年慶應義塾大学大学院理工学研究科開放環境科学専攻前期博士課程修了．2008 年同大学大学院理工学研究科開放システム科学専攻後期博士課程修了．博士（工学）．ネットワークセキュリティ，デジタルフォレンジックスの分野に興味を持つ．IEEE 会員．

**塩澤　秀和**（正会員）
1971 年生．2000 年慶應義塾大学大学大学院理工学研究科計測工学専攻博士課程修了，博士（工学）．2003 年より玉川大学工学部知能情報システム学科専任講師．情報可視化，ヒューマンインタフェース，情報セキュリティ等に興味を持つ．電子情報通信学会，ACM，IEEE-CS 各会員．

**重野　　寛**（正会員）
1990 年慶應義塾大学理工学部計測工学科卒業．1997 年同大学大学院理工学研究科博士課程修了．1998 年同大学理工学部情報工学科助手（有期）．現在，同大学理工学部情報工学科准教授．博士（工学）．計算機ネットワーク・プロトコル，モバイル・コンピューティング，ネットワーク・セキュリティ，マルチメディア・アプリケーション等の研究に従事．著書『〜ネットワーク・ユーザのための〜無線 LAN 技術講座』（ソフト・リサーチ・センター），『コンピュータネットワーク』（オーム社）等．電子情報通信学会，IEEE，ACM 各会員．

**岡田　謙一**（フェロー）

慶應義塾大学理工学部情報工学科教授，工学博士．専門は，CSCW，グループウェア，ヒューマン・コンピュータ・インタラクション．情報処理学会誌編集主査，論文誌編集主査，GW 研究会主査等を歴任．現在，情報処理学会 MBL 研究会運営委員，BCC 研究グループ主査，日本 VR 学会理事，CS 研究会委員長．情報処理学会論文賞（1996，2001，2008），情報処理学会 40 周年記念論文賞，日本 VR 学会サイバースペース研究賞，IEEE SAINT'04 最優秀論文賞を受賞．情報処理学会フェロー，IEEE，ACM，電子情報通信学会，人工知能学会各会員．