Vol.2011-GN-79 No.20
2011/3/17

# 不正操作確率を考慮したログ提示手法

西岡千文[†1]　中山佑輝[†2]　小崎真寛[†2]　岡田謙一[†1, †3]

情報化社会の進展により，情報の電子化が進み，情報漏洩が深刻化している．ゆえに，情報を扱う管理者にとって，各ホストによる機密ファイルに対する操作を把握し，不正操作が行われた場合に迅速に対応することが重要となっている．既存のログ提示手法は，時系列に沿ってログを提示するものがほとんどである．この手法では，操作の前後関係を把握するのは容易であるが，不正操作を示すログがログデータの後方に含まれると，不正操作の発見に遅延が生じる．そこで本稿では，不正操作確率を考慮したログ提示手法を提案する．不正操作確率とは，そのログが不正操作である確率である．不正操作確率は，組織の過去のログデータの統計的な値から取得することが可能である．提案手法では，不正操作確率が高いログを前方へ提示することにより，不正操作の早期発見に貢献する．評価実験として，不正操作を発見するタスクを被験者に対して行った．解析時間，解析精度を評価項目とした．その結果，解析時間において 39.5%の向上が見られた．

# Visualization of Log Data with Probabilities of Fraudulent Manipulation

Chifumi Nishioka[†], Yuki Nakayama[††]
Masahiro Kozaki[††] and Kenichi Okada[†]

## 1. はじめに

情報化社会の発展に伴い，近年，情報の電子化が進められている[1]．同時に，情報漏洩事故の件数も増加している[2][3]．情報漏洩は恒常化しているといえる．情報漏洩を防ぐため，企業・組織内では，情報漏洩につながる恐れのある操作を独自に不正操作と定め，それを禁じている．管理者は，一定期間ごとに，クライアントより収集したログデータを解析し，不正操作が行われていないか確認作業を行う．この解析作業で，セキュリティインシデントを発見し，対応策をとる．早急な対応策をとることで，企業の信頼の低下を防ぐことが可能であり，セキュリティインシデントの早期認識の必要性がある[4][5]．以上のことより，ログ解析時における迅速な不正操作の発見が重要といえる．また，情報漏洩の原因は，ほとんどがヒューマンエラーによるものであり，事前に防ぐことが困難である[2]．このことからも，ログ解析によりセキュリティインシデントを発見し，早期に対応することの必要性がある．

本論文は，不正操作確率を導入することにより，不正操作を迅速に発見する手法を提案する．不正操作操作とは，そのログが不正操作である確率である．不正操作確率は，組織の過去のログデータの統計的な値から取得する．ログデータを管理者へ提示する際，従来では，ログは時間順に提示される．後方に不正操作が存在した場合，不正操作の発見が遅くなる可能性がある．不正操作発見の遅れは，セキュリティインシデント対応の遅れにもつながり，対応が遅くなるほど企業の信頼が低下の恐れを生ずる．そこで，不正操作確率が高いログより管理者へ提示することにより，不正操作の迅速な発見に貢献する．

以降，第 2 章で背景，第 3 章で関連研究について述べる．第 4 章で提案手法について詳述する．第 5 章で提案手法の評価実験を行い，第 6 章をまとめとする．

## 2. 関連研究

クライアント操作のログ解析手法に関して，これまで多くの研究がなされ，実際に多種多様なツールが導入されている．

MOTEX 社 LanScope[6]は，組織・企業を対象とした，ネットワーク総合管理ツールである．資産管理，Web アクセス監視等の多くの機能を有し，その一つに操作プロ

†1　慶應義塾大学理工学部
Faculty of Science and Technology, Keio University
†2　慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University
†3　独立行政法人　科学技術振興機構
JST CREST

セス管理機能が存在する．操作プロセス管理では，各クライアントにおけるファイル操作，アプリケーション操作など多種のログを取得し，ユーザへ提示する．しかし，取得した時間が古いログから，すなわち時系列に沿ってログを提示しているため，後方に不正操作を記録したログが存在した場合，不正操作発見の遅延につながる．

他にも様々なログ解析ツールが存在するが，いずれも時系列に沿ってログを提示するため，不正操作発見が遅延する可能性を孕む[7][8][9].

また NEC 社の InfoTrace[9]をはじめとするログ解析ツールは，フィルタリング機能をもつ．フィルタリング機能は，ログデータより，指定したユーザ，ファイルについて記録したログを抽出する．「list.xls を copy」といった単純な不正操作であれば，フィルタリング機能により不正操作を発見することが可能である．しかし「file.jpg を edit 後, upload」というような不正操作は, フィルタリング機能の有用性は薄れる．この不正操作確率を，フィルタリング機能を使用して発見する場合，まず file.jpg の upload を記録したログを抽出する. そして抽出されたログが, edit されてから upload されたものかどうか一行一行確認する必要がある．確認する中には不正操作を記録したものではないログも多く含まれる．このようなことから，不正操作を記録したログである確率の高いものより順にログを提示すれば，不正操作の早期発見につながるといえる．

## 3. 不正確率を考慮したログ提示手法

本論文では，一定期間に取得されたログを解析する場面で，不正操作をより迅速に発見する手法を提案する．企業内に存在する全ファイルを監視・追跡することは，ログ情報量の肥大化，コストの観点から現実的ではない．本提案では，予め機密ファイルを設定し，設定された機密ファイルに対する操作のログを解析する．

前提として, 過去に行われたログ解析により, 不正操作確率が得られるものとする．不正操作確率とは，ログが不正操作である確率であり，過去におけるログデータの解析結果の統計より取得する．取得した不正操作確率が高いログを優先的に管理者へ提示することにより，迅速な不正操作の発見を支援する．

### 3.1 不正操作確率と前提条件

本論文で使用するログは，表 1 に示す形式のものを想定する．

表 1 ログの属性

| Time | User | Operation | File | Destination/From |
|---|---|---|---|---|
| ログの取得時間 | 操作主 | 操作内容<br>(例)send，edit | 操作対象<br>ファイル名 | 操作先 |

属性 Destination/From は，属性 Operation の内容により，変化する．例えば，属性 Operation が send（メールでの送信）であれば，属性 Destination/From は送信先を示したものとなり，属性 Operation が receive であれば，送信元を示していることとなる．なお，表 1 に示す形式のログでなくとも，フォーマットを変更することにより，本論文の手法を導入することは可能である．

続いて，不正操作確率について述べる．不正操作確率とは，そのログが不正操作である確率である．ここで，ログとその不正操作確率の例を表 2 に示す.

表 2 ログ・不正操作確率の例

| Time | User | Operation | File | Destination/From |
|---|---|---|---|---|
| 9:32 | tanaka | upload | list.doc | ServerA |
| | 10.0% | 1.0% | 3.0% | 4.0% |

表 2 に示すログは.「9:32 に，ユーザ tanaka がファイル list.doc を ServerA に upload した」ということを示している．表 2 の最下行は，不正操作確率を示す.

属性 User に着目し，tanaka が過去 1000 回の操作の中で，100 回が不正操作であったとすると，このログのユーザの不正操作確率は，100/1000 により 10.0%と求められる．また，属性 File に着目すると，list.doc に関するログが，2000 回記録された中で，60 回が不正操作であったとすると，このログのファイルの不正操作確率は 3.0%となる．属性 Operation，Destination/From もそれぞれの属性に着目して，過去の操作における不正操作の割合を求めることにより，求めることが可能である．

本論文では，不正操作確率が高いログより，ユーザへ提示することで，迅速な不正操作発見に貢献する．

### 3.2 ソート方法

ログの各属性がそれぞれ不正操作確率を保有する．そこで，利用者がログの属性に優先度を付与することにより，ソートを行うこととした．ソートの例を以下の図 1~3 へ示す.

図 1 では，時間順にログが提示されている．図 2 では，最優先属性として，File を設定する．最優先属性として File と設定すると，属性 File の不正操作確率にのみを考慮して, ソートが行われる. ここで属性 File の不正操作確率は, File.doc, List.doc, Sec.pdf の順に高いため，図 2 のようにソートが行われる．不正操作確率が同じログについては，より古いログより表示される．図 3 では，第 2 優先属性として，User を設定する．属性 User の不正操作確率は，高いものから，佐藤，山田，田中，鈴木，高橋である．ここで，最優先属性として，属性 File が設定されている状態である．図 2，図 3 中の属性 File 列のみに着目すると，順序が変わっていないことがわかる．しかし，属性 File が同じ値をもつログの中では，属性 User の不正操作確率を考慮して

ソートされていることがわかる．

| Time | User | Ope. | File | Des./From |
|---|---|---|---|---|
| 9:32 | 田中 | Copy | List.doc | 山田 |
| 9:50 | 山田 | Upload | Sec.pdf | サーバA |
| 10:00 | 佐藤 | Print | Sec.pdf | プリンタA |
| 11:30 | 山田 | Print | List.doc | プリンタB |
| 14:50 | 田中 | Send | List.doc | 高橋 |
| 14:57 | 高橋 | Receive | Sec.pdf | 田中 |
| 15:43 | 鈴木 | Copy | Sec.pdf | 山田 |
| 16:02 | 鈴木 | Upload | File.doc | 鈴木 |
| 17:44 | 山田 | Upload | Sec.pdf | サーバA |

図 1 ソート例　デフォルト

| Time | User | Ope. | File | Des./From |
|---|---|---|---|---|
| 16:02 | 鈴木 | Upload | File.doc | 鈴木 |
| 9:32 | 田中 | Copy | List.doc | 山田 |
| 11:30 | 山田 | Print | List.doc | プリンタB |
| 14:50 | 田中 | Send | List.doc | 高橋 |
| 9:50 | 山田 | Upload | Sec.pdf | サーバA |
| 10:00 | 佐藤 | Print | Sec.pdf | プリンタA |
| 14:57 | 高橋 | Receive | Sec.pdf | 田中 |
| 15:43 | 鈴木 | Copy | Sec.pdf | 山田 |
| 17:44 | 山田 | Upload | Sec.pdf | サーバA |

図 2 ソート例　最優先属性: File

| Time | User | Ope. | File | Des./From |
|---|---|---|---|---|
| 16:02 | 鈴木 | Upload | File.doc | 鈴木 |
| 11:30 | 山田 | Print | List.doc | プリンタB |
| 9:32 | 田中 | Copy | List.doc | 山田 |
| 14:50 | 田中 | Send | List.doc | 高橋 |
| 10:00 | 佐藤 | Print | Sec.pdf | プリンタA |
| 9:50 | 山田 | Upload | Sec.pdf | サーバA |
| 17:44 | 山田 | Upload | Sec.pdf | サーバA |
| 15:43 | 鈴木 | Copy | Sec.pdf | 山田 |
| 14:57 | 高橋 | Receive | Sec.pdf | 田中 |

図 3 ソート例　最優先属性: File & 第 2 優先属性: User

以上のように，ログの属性に優先度を付与することで，利用者は，不正操作の内容に応じて，フレキシブルなソートが可能になると考える．

### 3.3 事前準備と運用

実際に，提案手法を導入するプロセスについて述べる．

① 機密ファイルを決定

全ファイルに対する操作のログを取得すると，ログデータの大きさも，解析に費やされる時間も膨大なものとなる．そのため，情報漏洩が起きてはならないファイルを機密ファイルとして定める作業が必要となる．

② 不正操作を定義

情報漏洩につながる恐れのある操作を不正操作として設定する．

③ 各ホストへ，ログを取得する監視プログラムをインストール

監視対象であるホストより，ログを取得できなければ，解析作業は行えない．ユーザが機密ファイルに対して，何らかの操作を行った際，その操作をログとして記録するソフトウェアをインストールする必要がある．

④ ログ解析

情報漏洩の発見のためにも，一定期間ごとに取得したログデータの解析作業を行う必要がある．

以上のプロセスに沿って，実際に提案手法は利用される．

## 4. アプリケーション

不正操作確率を考慮したログ提示手法を実現するために，アプリケーションの実装を行った．アプリケーション全体図を図 4 に示す．アプリケーションは，左右 2 画面に分割することが可能である．

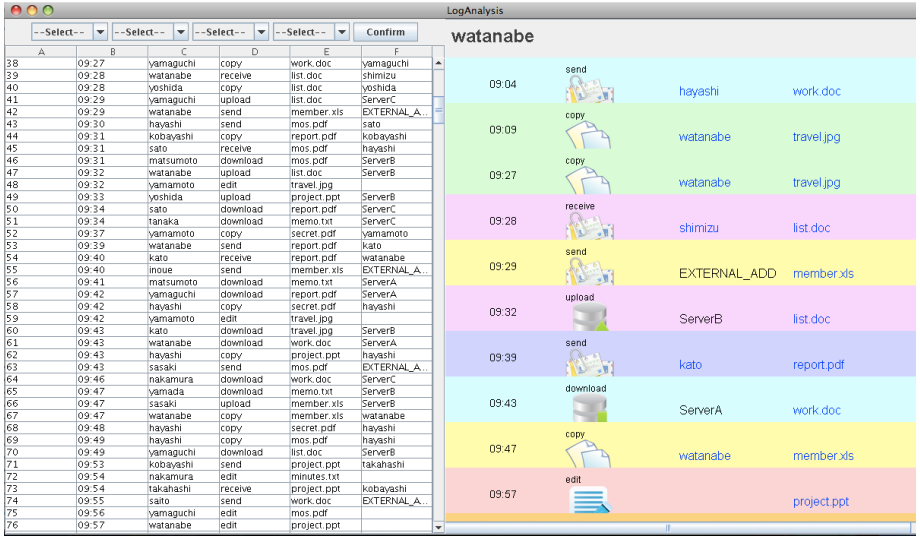


図 4 アプリケーション全体図

左画面は，テキスト表形式でログを提示する部分である．デフォルトの状態では，時間に沿った順序で提示されている．左画面上部のプルダウンボックスで，ログの属性を選択することで，ログは不正操作確率を考慮した順序にソートされる．

右画面は，ログの視覚化を行う部分である．左画面でソートを使用した場合，ログは時間順を崩した状態で表示される．よって左画面のみでログの前後関係を把握することは困難である．「edit されたあとに外部へ send」という前後関係の把握が必要な不正操作を発見する場面では，有用性を発揮できない．この問題点を解決するために，右画面でログの視覚化を行う．左画面で選択されているログのユーザのみのログデータを時間順に視覚化を行う．以下，左画面をソート部分，右画面を視覚化部分と呼ぶ．

### 4.1 ソート部分

アプリケーションの左半分を，不正操作確率を考慮したログ提示手法により実装された部分が占める．この画面では，上部の 4 つのプルダウンボックスで，優先的にソートを行うログの属性を決定する．

### 4.2 視覚化部分

左画面で選択しているログのユーザに着目したログの流れを視覚化して提示する．図 5 に画面を示す．


図 5 ユーザに着目した視覚化画面

ログは時間順に表示される．左より，各ログの属性 Time，Operation，Destination/From，File が示される．背景色は，属性 File により決定する．属性 File が同じログについては，同じ背景色となっている．青文字で表示されている Destination/From 部，File 部はクリック可能である．Destination/From 部をクリックすると，視覚化部分はクリックしたユーザに着目したログの流れを同様に提示した視覚化部分に切り替わる．File 部をクリックすると，選択したファイルに関するログを時間順に視覚化した画面に切り替わる．

File に着目した視覚化画面は，図 6 である．この画面では，左より，各ログの属性 Time，Operation，Destination/From，User が示される．背景色は，属性 User により決定する．

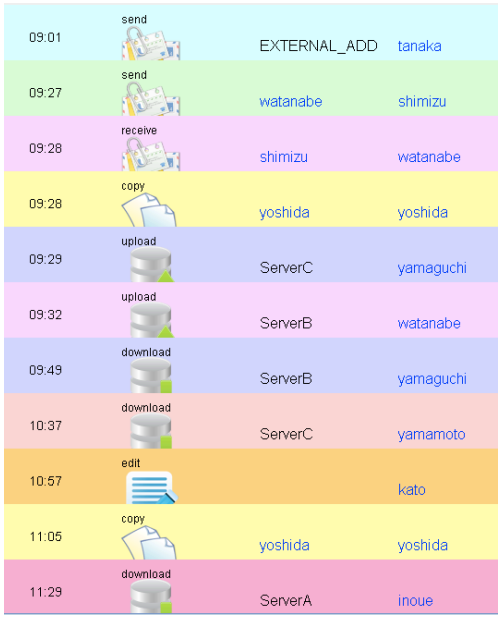


図 6 ファイルに着目した視覚化画面

## 5. 評価

不正操作確率を考慮したログの提示手法が，迅速な不正操作の発見に貢献している

ことを確かめるため，評価実験を行う．評価実験として，ログデータより不正操作をできるだけ多く発見するというタスクを，被験者に対して行った．タスクは，前後関係の把握が必要なタスクを 1 つ，不必要なタスクを 2 つ，全 3 タスク用意した．不正操作確率を考慮した提案手法と，時間順によるログ提示手法の 2 手法で比較評価を行った．それぞれの手法に対して，解析精度，解析時間について評価を行った．

### 5.1 タスク

被験者には，不正操作確率を考慮したログ提示手法を用いた解析（以下，提案手法解析）と，時系列によるログ提示手法の解析（以下，従来手法解析）を行ってもらう．提案手法解析のツールとして，4 章で紹介した実装アプリケーションを利用した．従来手法解析のツールとしては，実装アプリケーションより，不正操作確率によるソート機能をなくしたツールを利用した．

被験者が行うタスクは，用意したログから不正操作を制限時間内に出来る限り多く発見するというタスクである．評価実験で使用するタスクは，全 3 タスクで構成される．表 3 に，各タスク内容，制限時間を示す．

表 3 タスク

| | タスク内容 | 制限時間 |
|---|---|---|
| タスク 1 | secret.pdf の copy という不正操作を発見 | 1 分 |
| タスク 2 | ServerA より download された ファイルを EXTERNAL_ADD に送信するという不正操作を発見 | 5 分 |
| タスク 3 | edit 後, receive されたファイルを upload するという不正操作を発見 | 5 分 |

タスク 1 は, ログの前後関係の把握を必要としない不正操作を設定した. すなわち, 従来手法，提案手法ともに，視覚化部分を利用しなくても達成することが可能なタスクである．タスク 2 は，前後関係の把握が必要なタスクである．視覚化部分を利用することで，効率良く発見できる不正操作を想定した．タスク 3 は，視覚化部分を利用かつ，視覚化部分中での切り替え（青文字のクリック）を利用することで，効率良く発見できる不正操作を想定した．タスク 1 から 3 へ進むほど，不正操作の定義は複雑となり，タスクも被験者にとって困難なものとなってくる．

評価実験では，被験者へのアプリケーションの利用方法の説明後，タスクを開始した．被験者の慣れを考慮して，半分の被験者は従来手法より，また半分の被験者は提案手法より実験を進める．従来手法が先である場合，従来手法によるタスク 1，提案手法によるタスク 1，従来手法によるタスク 2，提案手法によるタスク 2 といった順序である．

### 5.2 評価条件

被験者は，情報工学を専攻する大学生・大学院生 10 名である．先に述べた不正操作 1~3 を各制限時間内に出来る限り多く発見してもらった．

想定した環境は，ホスト数 10 のネットワークである．ネットワーク全体で，機密ファイルを 10 ファイル保有している．期間は 8 時間であり，各ホストは 1 時間に機密ファイルに対する操作を平均 4 回行うものとした．そのため，ログデータの行数は 10*8*4 によって，320 行となった．このログはシミュレーションにより作成したものであり，ランダムに不正操作のログを発生させている．このログデータを提案手法解析・従来手法解析に使用した．

不正操作確率は，作成したログデータより計算を行うことにより，設定した．タスク 1~3 にそれぞれの不正操作が存在するので，それぞれに対応した不正操作確率 1~3 を計算した．タスク 1 を行う場合は不正操作確率 1 を，タスク 2 を発見する場合は不正操作確率 2 を，ソートに反映させることのより，提案手法に導入した．なお，不正操作確率と, 実際にログデータに含まれる不正操作の割合の誤差は, 10%以内である．

### 5.3 評価項目

解析精度，解析時間の 2 項目で，既存手法と提案手法を比較した．

解析精度は，（被験者が回答した中における実際の不正操作数）/（被験者が回答した不正操作数）とした．

解析時間は，（制限時間）/（被験者が発見した不正操作数）より求めた．すなわち，不正操作を 1 つ発見するために費やす時間である．なお，設定した制限時間内に，被験者より不正操作を全て発見したと申告があった場合は，式中の（制限時間）を（被験者の申告までの時間）に置き換えた．

### 5.4 実験結果

解析精度の結果を，表 4 に示す．

表 4 結果　解析精度

| | 従来手法（%） | 提案手法（%） |
|---|---|---|
| タスク 1 | 100.0 | 100.0 |
| タスク 2 | 93.7 | 97.7 |
| タスク 3 | 81.1 | 90.3 |
| 平均 | 91.6 | 96.0 |

全タスクにおいて，有意水準 $\alpha$ =0.05 における t 検定を行った．提案手法がより優れるという結果が得られているものの，どのタスクにおいても，有意差は得られなかった．これは，従来手法，提案手法ともに，ログデータの提示順序に差は存在するものの，インタフェースについては，両者ともに差がなかったためであるといえる．

また，解析時間の結果を，表 5 に示す．

表 5 結果　解析時間

| | 従来手法（s） | 提案手法（s） | 改善率（%） |
|---|---|---|---|
| タスク 1 | 12.6 | 7.3 | 37.9 |
| タスク 2 | 47.4 | 24.9 | 38.7 |
| タスク 3 | 191.3 | 111.1 | 41.9 |
| 平均 | 83.6 | 47.8 | 39.5 |

1−（提案手法の解析時間）/（従来手法の解析時間）を解析時間の改善率とし，表 5 中に示した．また，解析精度と同様，全タスクにおいて，それぞれ有意水準 $\alpha$ =0.05 の t 検定を行った．結果，全タスクにおいて，有意差が見られた．全タスクを平均すると，全体で 39.5%解析時間が向上したといえる．

解析時間改善率の結果より，タスク 1 から 3 へかけて，解析時間の短縮率は大きくなる．よって，前後関係の把握を必要とする不正操作ほど，または，不正操作の定義が複雑になるほど，提案手法の有用性が高いといえる．また，不正操作確率を導入することは，ログの前後関係の把握を困難にするという問題点があった．しかしこの結果により，左画面に視覚化画面を導入することで，この問題点を覆うことができたといえる．

以上より，不正操作確率を考慮したログ提示手法により，不正操作発見にかかる時間は向上したといえる．

## 6. まとめ

情報化社会の発展に伴い，近年，情報漏洩の問題は恒常化しており，減少する気配はない．情報漏洩を防ぐため，企業・組織内では，情報漏洩につながる恐れのある操作を独自に不正操作と定め，それを禁じている．管理者は，一定期間ごとに，クライアントより収集したログデータを解析し，不正操作を発見するログ解析作業を行う．ログ解析により，セキュリティインシデントを早期に発見することで，セキュリティインシデントに対する迅速な対応策をとることを可能とし，情報漏洩による企業の信頼低下を防ぐ．以上のことより，不正操作の早期発見が必要とされる．現在，ログ解析時のログ提示手法は，時間順による提示が一般的である．このとき，不正操作がログデータの後方に存在した場合，不正操作発見が遅延する可能性を孕むという問題が発生する．

本論文では，不正操作確率を考慮したログ提示手法を提案した．不正操作確率とは，そのログが不正操作である確率である．企業内が保持する過去のログデータの統計により，不正操作確率を算出し，各ログの属性に不正操作確率をもたせる．不正操作確率は，ログの各属性が保有する．管理者は，発見する不正操作に応じて，ログの各属性に優先度を付与してソートを行う．ソートにより，不正操作確率が高いログから提示することにより，不正操作の早期発見を支援する．ここで，不正操作確率によるログ提示手法では，時間順に提示されていたログデータの順序を崩してしまうため，ログの前後関係の把握が困難になるという問題点が生じた．この問題点は，ディスプレイを左右に 2 分割し，左画面で不正操作確率を考慮したログの提示を行う一方，右画面では左画面で選択しているログのユーザ内における操作の流れを視覚化して表示することにより解決を図った．右画面ではユーザに着目した視覚化，ファイルに着目した視覚化を行うことが可能である．

本手法の有用性を実証するために，不正操作を制限時間内に出来る限り多く発見する被験者実験を実施した．被験者実験は，全 3 タスクより成る．解析精度，解析時間に関して評価を行った．解析精度については，いずれのタスクにおいても有意差が見られなかった．一方，解析時間は全 3 タスク平均で 39.5%向上した．特に，ログの前後関係の把握が必要な不正操作を発見するタスクにおいて，改善率が高く，提案手法の有用性が現れた．

以上のことより，不正操作確率を考慮したログ提示手法は，不正操作の迅速な発見に貢献しているといえる．

## 参考文献

1)　総務省　平成 21 年通信利用動向調査の結果,
http://www.soumu.go.jp/main_content/000064217.pdf
2)　NPO 日本ネットワークセキュリティ協会　セキュリティインシデントにおける調査報告書,
http://www.jnsa.org/result/incident/data/2009incident survey v1.1.pdf
3)　Security NEXT　個人情報漏洩事件一覧, http://www.security-next.com/cat cat25.html
4)　独立行政法人　情報処理推進機構　情報漏えいインシデント対応方策に関する調査報告書,
http://www.ipa.go.jp/security/awareness/johorouei/report.pdf
5)　独立行政法人　情報処理推進機構　情報漏えい発生時の対応ポイント集,
http://www.ipa.go.jp/security/awareness/johorouei/rouei taiou.pdf
6)　MOTEX 社　LanScope Cat6, http://www.motex.co.jp/cat6/index.html
7)　シマンテック社　Vontu Data Loss Prevention8,
http://www.symantec.com/region/jp/vontu/products/
8)　NEC 社　InfoCage, http://www.nec.co.jp/cced/infocage/
9)　NEC 社　InfoTrace, http://www.nec.co.jp/cced/InfoTrace/