

ログ分析による情報漏洩監視

榊原裕之[†] 桜井鐘治[†]

近年、企業などの組織において、従業員による機密情報の不正送出が脅威となっている。対策の一つとして、様々なログを監視し機密ファイルが組織外へ送出されたか確認する方法がある。ログの監視方法として機密ファイルがどのように扱われたか、その変遷を端末操作ログ、ファイルサーバログ、メール監視ログ等を用いてトレースする方法がある。筆者らの想定するログ監視環境では、端末操作ログにファイル間でのデータのコピー&ペーストが記録されないため、当操作が発生した場合はファイルの変遷をトレースできない課題がある。この課題を解決するため、コピー&ペーストの操作がログに記録されない環境において、変遷をトレースする対象のファイルを開いている期間に、別途開いていたファイルをコピー/ペースト対象のファイル候補としてトレース対象に追加する方式を検討した。また、トレース対象候補が複数発生した場合に絞込みを行う追加処理について検討・考察した。

Information Leakage Monitoring By Log Analysis

Hiroyuki Sakakibara[†] and Shoji Sakurai[†]

Recently in an organization such as an enterprise, information leakage by employees has become a threat. One of countermeasures against the threat is detection that an employee sends a confidential file from an organization to outside by monitoring various logs. One of log monitoring methods is tracing operation to a confidential file by analyzing multiple logs such as a terminal operation log, a file server log, a mail monitoring log and so on. However, in a log monitoring environment where we target, data copy & paste operation between files is not written on a terminal operation log. Therefore when a data copy & paste operation with a confidential file is performed, the operation can not be traced. For solving this issue we propose a method that opened files during a trace target file being opened is estimated as data copy source or data paste destination and treated as a new trace target file. When multiple new trace files exist, processing load of tracing will be heavy, therefore we propose an additional method to narrow down the files and describe their observation.

1. はじめに

近年、企業などの組織において、従業員による機密情報の不正持ち出しが問題となっている。例えば、従業員が機密情報の保存されたシステムから機密ファイルを取得し、添付メール等で組織外へ送信したり CD 等で持ち出したりするケースがある。このような内部犯行による情報漏洩対策の一つとして、様々なログを監視し機密情報が組織外へ持ち出されたか確認する方法がある。例えば、ファイルサーバログ、メール監視ログ等について、アクセス回数や送信量などの基準を個別に設けて監視することで、従業員による情報漏洩と疑わしい行動を検知する試みがある。さらにログの個別の監視に加え、複数のログをつき合わせて、従業員により組織外へ送出されたファイルがどのようなデータの由来か遡り確認するトレースバックや、従業員により取得された機密ファイルがその後どのように扱われたか確認するトレースフォワードによる監視がある。

トレースバック/フォワードでは、主に端末操作ログにおけるファイルのリネーム、コピー、移動などの記録からファイルの変遷を追跡する。ところが、筆者らが想定するログ監視環境では、ファイル間のデータのコピー&ペーストはログに記録されないため、コピー&ペーストが行われた場合は追跡が困難となる課題が判明した。コピー&ペーストを追跡するためにはコピー&ペーストを記録する専用の仕組みを適用する必要があるが、想定するログ監視環境では、この様な仕組みを新規に追加せず、既に取得しているログでコピー&ペーストの追跡に対応する制限があった。

本稿は、コピー&ペーストの記録を含まないログにおいて、ファイル間でコピー&ペーストが行われた場合でもトレースバック/フォワードを行う方式について検討したものである。本稿は、以下の構成である。2章では、内部犯行による情報漏洩とログ監視による対策について述べる。3章では、トレースバック/フォワードによるファイルの追跡について説明する。4章では、筆者らが想定するログ監視環境におけるトレース時のコピー&ペーストの課題について説明する。5章では、課題への対策方式を述べる。6章で考察と今後の課題について述べ、7章でまとめる。

2. 内部犯行による情報漏洩とログ監視による対策

2.1 内部犯行による情報漏洩

米国の CERT の報告 2)では、組織における従業員による内部犯行の脅威について分析しており、調査対象とした 250 の内部犯行中、118 件が情報漏洩、残りは業務妨害という分析結果が示されている。情報漏洩の理由として、個人の利益のために組織の機密情報を売却して金を得る金銭目的や、転職先の業績に貢献するための機密情報の

[†] 三菱電機株式会社 情報技術総合研究所
Mitsubishi Electric Corporation, Information Technology R&D Center

悪用が挙げられている。その他、ストーキングや身元調査の目的で、顧客情報を窃盗する可能性が考えられる。

国内においては企業における顧客情報の漏洩が報告されており 3)、企業の正社員、契約社員などが漏洩させている事例が多い。顧客情報の漏洩は、企業への信用を失墜させ、顧客への賠償対応などにより、企業へ大きなダメージを与える。

2.2 ログ監視による対策

このような内部犯行による情報漏洩への対策として、専用ソフトウェアによる添付メール等の通信の監視や、媒体によるファイルの持ち出し制御を行い、リアルタイムに漏洩防止を実現する方法がある 2)。専用ソフトウェアによる監視は、その設定により検知漏れが発生することがある。漏洩監視の設定を厳格にすると、誤検知が多くなることもあり、これを回避するために設定を緩めると検知漏れが発生する。このため、運用の現場ではノウハウに依存した漏洩監視の設定が行われており、検知漏れの可能性が残っている。

これを補完する対策として、ファイルサーバログ、端末操作ログ、メール監視ログなどのファイルに関わるログを分析し、情報漏洩の痕跡が無いのか、定期的に確認する方法が考えられる。関連技術として、ログやイベント記録を用いてデータの取り扱いを追跡する文献がある 1) 4) 5)。ログの分析により漏洩が確認された場合、情報漏洩自体はブロックできなくとも、事後対応が迅速になるメリットがある。

3. トレースバック/フォワードによるファイルの追跡

ログ分析による情報漏洩監視の方法として、ログ個別に情報漏洩の監視基準を設け、これを逸脱する行為について詳細に調査する方法がある 2)。例えば、添付メールのファイルがある大きさを越えた場合には、添付ファイルによる漏洩を疑うという方法である。しかし、これだけでは添付ファイルの内容が組織外へ漏洩してはならない機密情報か否か分からないため、複数のログ分析により、組織外へ送出したファイルが機密情報由来か調査する方法がある。本稿では、これをログによるファイルのトレースと呼ぶが、ログの分析の仕方によりトレースバックとフォワードがある。本章では、ファイルのトレースバックとフォワードについて説明する。

3.1 ファイルのトレースバック

メール等で組織外へ送出したファイルの由来が、機密情報を含んだファイルであるかログから確認する方法である。トレースバックはファイルの送出後の事後確認であるが、定期的を実施することで、万が一機密情報由来のファイルが送出されたことが判明した場合でも、送出先に対してファイルの削除を依頼するなど事後対応を速やかに実施可能となるメリットがある。

具体的なトレースバック方法を図 1 に示す。図 1 では、メールサーバはメールの送

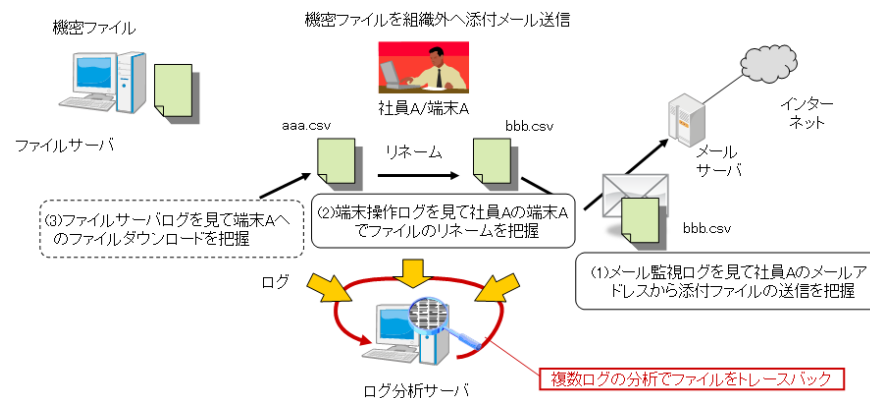


図 1 トレースバック

信履歴をメール監視ログに記録し、ファイルサーバにおいては機密ファイルのダウンロードの履歴をファイルサーバログに記録する。さらに、社員の端末においてはファイルをどの様に操作したかを端末操作ログに記録する。ファイルのトレースバックは以下の(1)(2)(3)のログの記録を遡りファイルの変遷を追跡する作業である。

(1)メール監視ログに記録された「社員 A のメールアドレスからがファイル bbb.csv をメール添付・送信」という記録。これは、添付ファイルのサイズが基準値を超えた場合に検知したエントリである。

(2)社員 A の使用端末である端末 A の端末操作ログに記録された「ファイル aaa.csv をファイル bbb.csv にリネーム」という記録

(3)ファイルサーバログに記録された「ファイルサーバから端末 A へ aaa.csv をダウンロード」という記録

これらの記録をメール送信から遡って追跡すると、社員 A が送信した添付ファイル bbb.csv は機密ファイルである aaa.csv が由来であることが分かる。この様にログを時間的に遡ってファイルへの取り扱いを調査する方法をトレースバックという。図 1 では、メールに添付された bbb.csv をトレース対象ファイルとする。各ログはログ分析サーバに集められトレースバックされる。トレースバックは(1)の時点から予め定めた一定期間を遡り実施する。

3.2 ファイルのトレースフォワード

トレースフォワードは、トレースバックとは逆に、時系列的に未来に向かってファイルの変遷を追跡する。トレースフォワードによる監視では、ファイルサーバからダウンロードしたファイルのコピーやリネームなどを追跡し、最終的に添付メールで送

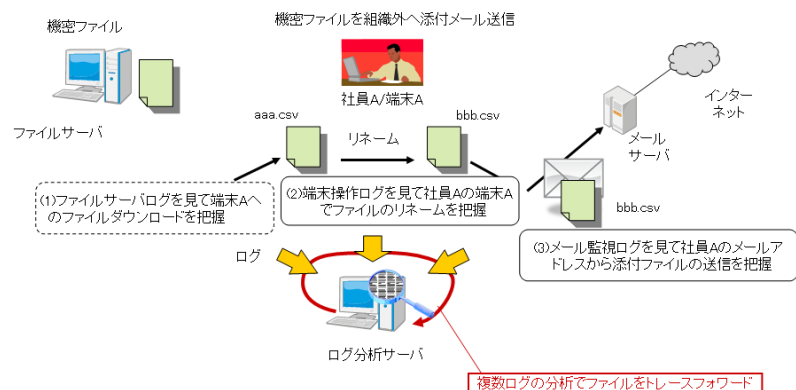


図 2 トレースフォワード

信されたか監視する。

図 2 では、トレースフォワードの例として、端末 A によるファイルサーバからの aaa.csv のダウンロード、aaa.csv の bbb.csv へのリネーム、bbb.csv のメール送信を各ログから調査する例を挙げている。トレースフォワードはファイルサーバからのダウンロードを行った時点から予め定めた一定期間について実施する。

4. 想定するログ監視環境におけるトレースの課題

4.1 想定するログ監視環境

筆者らが想定するログ監視環境には、メール監視ログ、ファイルサーバログ、端末操作ログが記録される。

- ・メール監視ログ
送信メールの from/to アドレス、送信日時、件名、添付ファイル名が記録される。
- ・ファイルサーバログ
ファイルをダウンロードした端末の識別子、ダウンロード日時、ダウンロードファイル名が記録される。
- ・端末操作ログ
ファイルの生成、削除、リネーム、移動、コピー、ファイルを開いたアプリケーション名、操作対象ファイル名、ファイルサイズが記録される。

ログ監視環境では、これらのログの内容を分析することが可能である。

4.2 想定するログ監視環境におけるトレースの課題

メール監視ログ、ファイルサーバログ、端末操作ログを突き合わせる事で、ファイ

ルのメール送信に関するトレースバック/フォワードが可能である。しかし、端末操作ログに記録されるファイル操作が、ファイルの生成、削除、リネーム、移動、コピーのみであるため、トレース対象ファイルと別ファイル間で発生したデータのコピー&ペーストをトレースすることができない課題がある。あるファイルからデータをコピーしトレース対象ファイルにペーストしていた場合は、このコピー&ペーストがログに記録されていないため、コピー元のファイルとペースト先のファイル（トレース対象ファイル）を関係付けることができず、トレースバックが中断されてしまう。トレースフォワードにおいても同様の課題がある。

コピー&ペーストの発生時もトレースを実現するための方法の 1 つとして、コピー&ペーストの操作を記録するソフトウェアを端末で稼働させる方法が考えられる。しかし、このようなソフトウェアの追加インストールを行うことなく、現在記録されているログを分析することでコピー&ペーストに対してもトレースを実現する制約が挙げられた。

5. コピー&ペーストに対応したトレース方式の検討

4.2 に示した制約のもと、既を取得しているログでコピー&ペーストに対応したトレースを実現する方式を検討した。最初に、トレースバックとトレースフォワードにおけるコピー&ペーストへ対応した基本方式について説明する。次に、基本方式では条件によりトレース対象が増える課題があるため、これを解決する追加方式を説明する。

5.1 トレースバックにおけるコピー&ペースト対応基本方式

図 3 はトレースバックの一例であり、トレース対象ファイルはメールに添付されたファイル①である。端末操作ログから、ファイル①はファイル②をリネームしたものであることが分かっており、ファイル②をトレース対象ファイルとする。また、ファイルサーバには機密ファイルが格納されているが、ファイル②はファイルサーバログに記録された端末にダウンロードされたファイル名と一致しないことが分かっている。ここで 2 つのケースが考えられる。1 つ目は、ファイル②はファイルサーバからダウンロードされた機密ファイル由来ではないケースである。2 つ目は、ファイルサーバからダウンロードした機密ファイルのデータをファイル②へコピー&ペーストしており、ファイル②はファイルサーバの機密ファイル由来となるケースである。

本方式では、コピー&ペーストを判断するために端末操作ログを使用する。記録内容が、ファイルの生成、削除、リネーム、移動、コピー、ファイルを開いたアプリケーション名、操作対象ファイル名、ファイルサイズであり、ファイル間のデータのコピー&ペーストに関する情報は無い。そこで、ファイル間におけるデータのコピー&ペースト時は、端末の操作者はコピー元のファイルをアプリケーションでオープンし、同時にペースト先のファイルをアプリケーションでオープンし、これらのファイル間

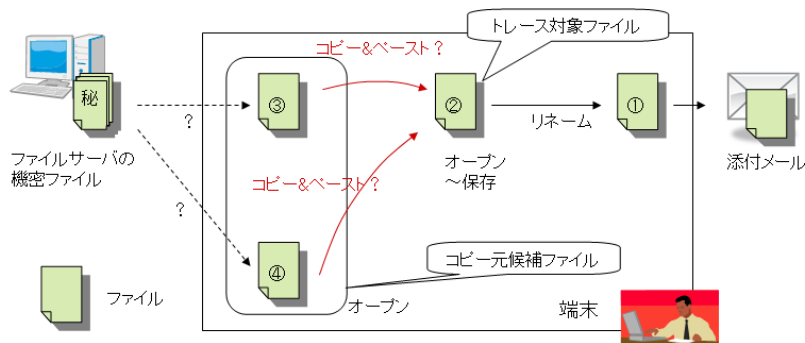


図 3 トレースバック時のコピー元候補ファイルの判断

でデータのコピー&ペーストを実施する点に着目した。トレース対象ファイルをペースト先と考えた場合、トレース対象ファイルをアプリケーションでオープンしてから保存するまでの期間に別途オープンしたファイルをコピー元の候補となるファイル（コピー元候補ファイル）として識別する。図 3 においては、ファイル③、④がコピー元候補ファイルである。

図 4 は、トレース対象であるファイル②をオープンしてから保存する間に別途オープンしていたファイル③、④の関係を示したものである。ファイル③、④はコピー元となりうるため、コピー元候補ファイルとして識別する。次に、コピー元候補ファイル③、④がファイルサーバからダウンロードしたファイルに辿り着くか否か調べるために、これらを新たにトレース対象ファイルとして、端末操作ログとファイルサーバログを用いてトレースを行う。最終的に、ファイルサーバからダウンロードしたファイル名に辿り着いた場合には、メール送信したファイルはファイルサーバ由来であると判断する。

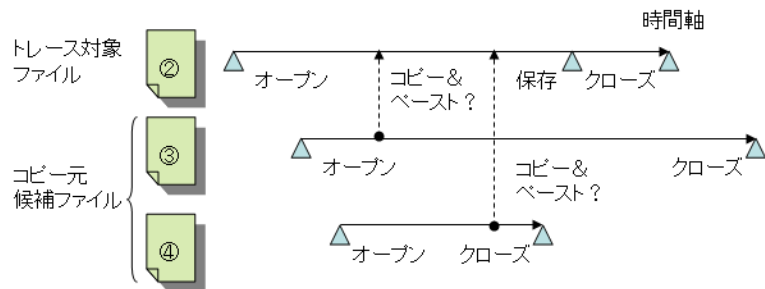


図 4 トレース対象ファイルとコピー元候補ファイルの関係

5.2 トレースフォワードにおけるコピー&ペースト対応基本方式

トレースバックの場合と同じ考え方で、ペースト先の候補となるファイルを識別する。図 5 において、ファイルサーバからダウンロードしたファイルをリネームしたものがファイル①であり、トレース対象ファイルである。また、ファイル①は添付メールで送信されていないことがメール監視ログから分かっている。ここで 2 つのケースが考えられる。1 つ目は、ファイル①は実際に添付メールで送信されていないケースである。2 つ目は、ファイル①のデータを別ファイルへコピー&ペーストし、そのファイルが添付メールで送信されたファイル④にリネームなどを経て辿り着くケースである。本方式においては、トレース対象ファイルをコピー元と考えた場合、トレース対象ファイルをアプリケーションでオープンしてからクローズするまでの期間に別途オープンしたファイルをペースト先の候補となるファイル（ペースト先候補ファイル）として識別する。

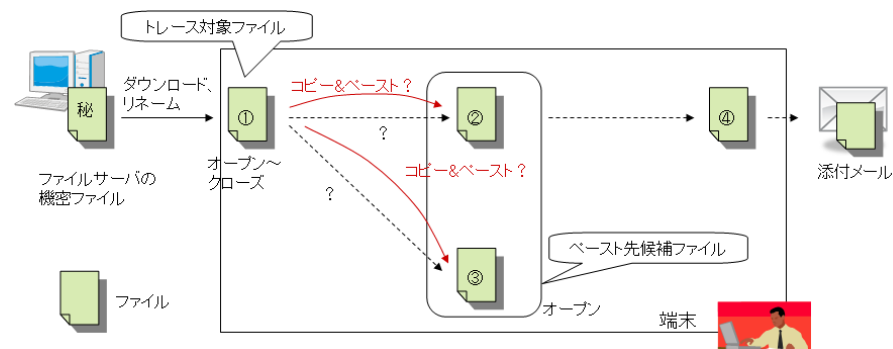


図 5 トレースフォワード時のペースト先候補ファイルの判断

図 6 はファイル①をオープンしてからクローズする間に別途オープンしていたファイル②、③の関係を示したものである。

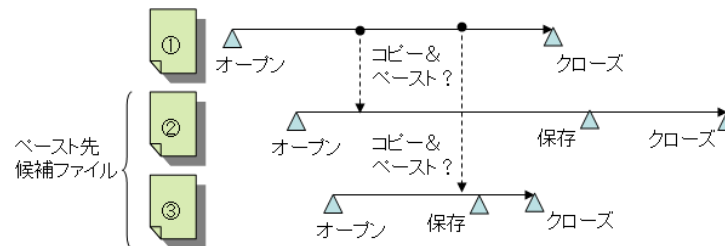


図 6 トレース対象ファイルとペースト先候補ファイルの関係

ファイル②, ③はペースト先となりうるため, ペースト先候補ファイルとして識別する. 次に, ペースト先候補ファイル②, ③が添付メールで送信したファイルであるファイル④に辿り着くか否か調べるために, これらを新たにトレース対象ファイルとして端末操作ログとメール監視ログを用いてトレースを行う. 最終的に, 添付メールで送信したファイルに辿り着いた場合には, 添付メールで送信したファイル④はファイルサーバ由来であると判断する.

5.3 検討した基本方式の課題

5.1 に示した基本方式では, トレース対象ファイルをオープンしてから保存する間に開いたファイルをコピー元候補ファイルとし, 5.2 に示した基本方式では, トレース対象ファイルをオープンしてからクローズする間に開いていたファイルをペースト先候補ファイルとした. また, コピー元候補ファイル, ペースト先候補ファイルを新たにトレース対象ファイルとしてトレースを行うことでファイルの変遷を追跡し, 結果としてメール送信されたファイルがファイルサーバ由来か調査する方式とした.

これらの基本方式では, コピー元候補ファイル, ペースト先候補ファイルが複数存在した場合はトレースするファイル数が増加する課題がある. 従って, 複数のコピー元候補ファイル/ペースト先候補ファイルがある場合は他の手段でファイル数を絞りトレースを継続する必要がある.

5.4 トレース元/ペースト先候補ファイルの絞込みのための追加方式

5.1, 5.2 の基本方式において, 複数のコピー元候補ファイル/ペースト先候補ファイルが存在した場合, これらにはトレース対象ファイルとの間でコピー&ペーストが発生していないファイルも含まれている可能性がありトレースした場合はトレース誤りの原因となる. 従って, コピー&ペーストが発生している可能性の高いファイルをさらに絞込むために以下の追加方式を検討した. これらは, 5.1 により識別されたトレース元候補ファイル, 5.2 により識別されたペースト先候補ファイルに対して適用される.

① ファイル間の類似性により絞込む方式

トレースバックの場合は, トレース対象ファイルとコピー元候補ファイルの間に類似性があるかファイルを調査する. トレースフォワードの場合は, トレース対象ファイルとペースト先候補ファイルの間に類似性があるかファイルを調査する (類似性の比較方法は既存の類似性比較技術を利用する. 例えばコードクローンを検出する技術の適用が考えられる 6)). トレース対象ファイルと類似性のあるコピー元/ペースト先候補ファイルを新たなトレース対象ファイルとして絞込む.

② 機密のキーワードを含むファイルに絞込む方式

ファイルサーバにおけるファイル内に存在する機密を表すキーワードを予め抽出しておく. トレースバックの場合は, コピー元候補ファイルにキーワードが含まれているか調べる. トレースフォワードの場合は, ペースト先候補ファイルにキーワード

が含まれているか調べる. キーワードが含まれるコピー元/ペースト先候補ファイルを新たなトレース対象ファイルとして絞込む.

③ ファイルサイズの増加により絞込む方式

トレースバックの場合は, トレース対象ファイルのオープン日時 t_1 のサイズと保存日時 t_2 のサイズの差分を調べる. この差分を Δs とする. $\Delta t = t_2 - t_1$ とし, $\Delta s / \Delta t$ がある閾値を超えた場合にコピー&ペーストが発生したと判断する. 当方式は, 短時間にファイルサイズが増加した場合, キーボード手入力ではなくコピー&ペーストによりサイズが増加したと判断する考え方に基づく. 例えば, テキストファイルが 1 秒あたり 1K バイト増えた場合は (閾値 1K バイト/秒), キーボード手入力では通常発生しない増加レートであるため, コピー&ペーストでデータが増えたと見なす.

トレースフォワードの場合は, トレース対象ファイルのオープン日時以降にペースト先候補ファイルを保存した日時を t_2' とし, t_2' から遡りペースト先候補ファイルをオープンした日時 t_1' を調べる. $\Delta t' = t_2' - t_1'$ とし, この期間のファイルの増分を $\Delta s'$ とし, $\Delta s' / \Delta t'$ が閾値を超えたか確認する.

④ 拡張子により絞込む方式

経験的に, コピー&ペーストが起こりやすい拡張子の組み合わせがある. 例えば, コピー元がテキストファイルの拡張子であれば, 同じテキストファイルやワープロファイルへのペーストが起こりやすい. そこで, 予めコピー&ペーストが発生する可能性の高い拡張子の組み合わせとスコアを定義する. トレース対象ファイルとコピー元/ペースト先候補ファイルの拡張子の組み合わせについて定義からスコアを調べ, スコアが別途定めた閾値未満であれば, コピー元/ペースト先候補ファイルから除外することで絞込みを行う. 例えば, 表 1 の様な拡張子の組み合わせに対するスコアを定義する. この例では, 同種の拡張子を組み合わせたスコアは 3 であり, それ以外の場合のスコアは 2 または 1 である. さらに, 閾値を 2, トレース対象ファイルの拡張子をワープロファイル拡張子とする.

表 1 拡張子の組み合わせ

ペースト先 コピー元	テキストファイル 拡張子	ワープロファイル 拡張子	プレゼンテーション ファイル拡張子
テキストファイル 拡張子	3	2	1
ワープロファイル 拡張子	2	3	1
プレゼンテーション ファイル拡張子	1	1	3

トレースバックにおいては, トレース対象ファイルをペースト先とするので, 閾値

2 以上となるコピー元の拡張子はテキストファイル拡張子とワープロファイル拡張子である。従って、コピー元候補ファイルが複数ある場合、テキストファイル拡張子とワープロファイル拡張子を持つファイル以外は除外してトレースを行う。トレースフォワードにおいては、トレース対象ファイルをコピー元とするので、閾値 2 以上となるペースト先の拡張子はテキストファイル拡張子とワープロファイル拡張子である。従って、ペースト先候補ファイルが複数ある場合、テキストファイル拡張子とワープロファイル拡張子を持つファイル以外は除外してトレースを行う。

6. 考察と今後の課題

6.1 方式の比較

5.1, 5.2 の基本方式のみと、5.4 の追加方式①～④を併用した場合の比較を表 2 に示す。比較項目として、ログ以外に調査する対象の有無、トレース誤りの可能性、トレース漏れの可能性、トレース時にコピー元/ペースト先候補ファイルが既に削除された場合の処理の可否を挙げた。

表 2 追加処理の比較

	基本方式	方式①併用	方式②併用	方式③併用	方式④併用
ログ以外の調査	不要	必要	必要	不要	不要
トレース誤り	高	基本方式以下 アルゴリズム依存	基本方式以下 キーワード依存	基本方式以下 閾値依存	基本方式以下 閾値依存
トレース漏れ	低	基本方式以上 アルゴリズム依存	基本方式以上 キーワード依存	基本方式以上 閾値依存	基本方式以上 閾値依存
ファイル削除時	処理可	処理不可	処理不可	処理可	処理可

追加方式①～④は、基本方式で得られたコピー元候補/ペースト先候補ファイルに対して各方式の条件に合致しないものを破棄し候補ファイル数を減らす方式のため、トレース誤りは基本方式より増加しないが、追加方式が適切に機能せず本来トレースすべきファイルを破棄した場合にトレース漏れが基本方式より増加する可能性がある。追加方式が適切に機能するか否かはアルゴリズム、キーワード、閾値に依存する。

追加方式①, ②は、ファイルが必須となるため、対象ファイルが削除された場合は処理が不可となる。また、これらの適用はファイル調査が許諾される環境に限定され

るデメリットがある。

追加方式③, ④はログのみの調査である。従ってファイル削除時においても処理が可能である。③, ④は同時に実施することが可能であるため、併用によりトレース誤りをさらに低減することが可能となるが、トレース漏れが増える可能性もあるため、バランスを考慮した併用が必要と考察する。

6.2 今後の予定

本稿では、コピー&ペーストの記録の無い端末操作ログのみからコピー&ペーストの発生を推測しトレースを継続する方式を検討した。今後は、これらの方式の効果について実ログを用いて確認する予定である。その際に、トレースを実施する期間を適切に決定する必要がある。トレースする期間を長くすればトレース漏れが減る可能性があるが、同時に、コピー元/ペースト先候補ファイルが増加する。トレース期間を短くした場合はその逆である。

検討方式によるトレースの結果、機密ファイル由来のファイルの組織外への送があったと判断された場合は、トレース誤りが発生する可能性があるため、送を実施した従業員へヒアリングを行い機密ファイルの送の事実確認を実施する必要がある。

7. おわりに

筆者らの想定するログ監視環境では、端末操作ログにコピー&ペーストの記録が無い場合、コピー&ペースト発生時のトレースが行えない課題があった。本稿では、コピー&ペーストを行う際に、コピー元ファイルとペースト先ファイルを同時に開く場合があることに着目したトレースの基本方式を検討した。当方式では、端末操作ログにコピー&ペーストの記録が無い環境でも、コピー元/ペースト先の候補のファイルを決してトレースを実行する。基本方式のみではコピー元/ペースト先の候補のファイルが複数決定される可能性がありトレース誤りに繋がるため、これらを絞込む追加方式を検討した。今後は、実ログの監視に適用し検討方式の効果を確認する予定である。

参考文献

- 1) メディア、組織を越えた情報来歴管理, http://www.hitachi.co.jp/rd/sdl/people/info_history/02.html
- 2) Common Sense Guide to Prevention and Detection of Insider Threats 3rd Edition – Version 3.1, CERT
- 3) 個人情報漏洩, <http://ja.wikipedia.org/wiki/個人情報漏洩>
- 4) 片山, 高, 小櫻, 津田: クラウドにおけるデータ秘匿・追跡技術とその応用, CSS2010
- 5) 高 杰, 園田, 片山, 津田: メール添付ファイルのトレースシステムの試作, 情報処理学会 第 72 回全国大会
- 6) the archive of CCFinder Official Site, <http://www.ccfinder.net/>