

ログ管理による情報流通の追跡・可視化方式の検討

小高 祐樹[†] 高田 慎也[‡] 石本 英隆[‡]

警察大学校 警察情報通信研究センター[†]

NTT情報流通プラットフォーム研究所[‡]

1. はじめに

情報の流通の過程を把握することは、重要情報の拡散の検知による情報漏えいの抑止や、漏えいに至るまでの流通経路把握による漏えい後の原因調査支援といった効果が期待できる。

また、既存のPC監視・管理技術を用いることで得られるファイル操作ログを集めてファイルの移動を追跡することにより、情報の流通を可視化することができる。本発表では、情報漏えい対策支援のための、ファイル操作ログによる情報流通の追跡・可視化方式を検討する。

2. 既存技術と問題点

情報の流通を可視化する製品が発表されているが、一度に可視化される範囲は個々のPC内における流通のみであり、ファイルが組織内の複数のPC間にわたって流通する場合には、人によってPC毎に可視化を繰り返す必要がある[1][2]。

しかしながら、情報漏えいへの対処には迅速性が要求されるため、個別PC毎の分析やそのつながり方を順次繰り返し追っていくことは分析上の問題となる。

組織全体にわたる情報流通は一度に関連性をもって可視化される必要がある。

3. 目的

本発表では既存技術の問題点を踏まえて、組織全体にわたる情報流通を迅速に追跡・可視化するための方式の確立を目的とする。

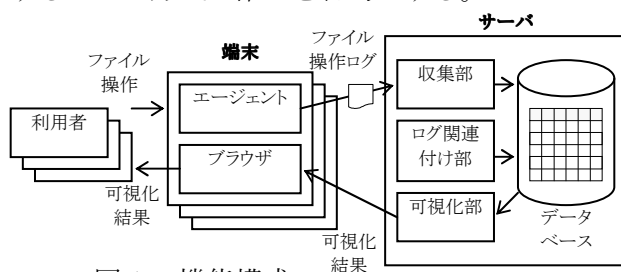


図1 機能構成

「A study on the tracking and visualizing methods of information distribution by log management.」

[†] National Police Academy,

Police Info-Communications Research Center

[‡] NTT Information Sharing Platform Laboratories

4. 解決すべき課題

組織内すべてのPCからファイル操作ログ（以下、「ログ」という）を1箇所に集めて保存し、流通を把握したいファイルについて、そのファイルの流通過程で出力された一連のログを保存されたログの中から抽出し、図によって可視化することで組織全体にわたる情報流通の把握が可能となる。

一連のログの抽出にあたっては、関連するログの追跡が必要となる。追跡とは、あるログに記載されている情報をkeyに検索し、続いてkeyを変化させながら次々に他のログ検索を繰り返す、関連するログを順次抽出していく処理である。例えば、起点ログを定め、それに記載の日時を起点として過去方向と未来方向に、ファイルの名称や保存場所などを検索条件に追跡を行うことで一連のログを抽出することができる。

追跡にかかる時間は、それを構成する検索時間の累積であるため、検索時間の短縮が課題となる。また様々なkeyで検索を続けるため、キャッシュの効果も低いと想定される。

5. 実現方式

図1に機能構成を示す。

本方式は組織内の各端末と1台のサーバとなり、各々の端末にはエージェントが組み込まれる。また、サーバは収集部、データベース、ログ関連付け部、可視化部から構成される。以下に処理の流れを述べる。

- (1) 利用者が端末を用いてファイル操作を行うと、エージェントが当該操作情報を記載したログをサーバに送信する。
- (2) サーバの収集部はログを受信し、これに一意なログIDを付与したうえでデータベースに保存する。
- (3) サーバのログ関連付け部はデータベースに新たなログが保存されると、直ちに当該ログを起点として過去方向に追跡を1回行い、抽出したログ（以後、「親ログ」という）のログIDを当該ログに親ログIDとして追加する。
- (4) 利用者が可視化要求を行った場合には、可視化部は親ログIDを基にデータベース内を

追跡し、抽出したログを基にブラウザを通じて情報流通を図で可視化する。

6. 本方式の特徴

本方式の特徴は、可視化要求の前に、処理(3)において予めすべてのログについて親ログ ID 追記による親ログとの関連付けを行っておくことで、処理(4)の追跡時間を低減することにある。

処理(3)で追加される親ログ ID が無い場合、処理(4)の追跡において検索条件として操作日時、操作端末名、ファイルの名称、保存場所などを用いる必要があり、検索条件の複数化は検索時間の増大をもたらす。

そこで、処理(3)において全てのログに対して1回追跡を行っておき、得られた親ログのログ ID を各々のログに追記する。これによって処理(4)の追跡においては検索条件が親ログ ID のみと単一化できる。この検索条件単純化には検索時間の低減が期待できる。

7. 提案方式の試作

提案方式を評価するため、試作を行った。

端末の OS は Windows2003Server とし、エージェントとして市販製品を用いた。

サーバの OS は Linux とし、蓄積部には RDBMS である PostgreSQL を用いた。

8. 評価・考察

8. 1 可視化範囲

組織全体から集めたログを追跡対象とすることで、組織全体にわたる情報流通を一度に可視化できることを確認した。

8. 2 検索にかかる時間

関連するログを順次抽出する（追跡する）際の、ログを1個抽出するのに要する検索時間を評価した。対象ログの総量は200万件とした。これは100人の利用者が1日80ファイル操作を行った場合の1年間分のログ量を想定している。

表1は親ログ ID を利用しない検索方法と親ログ ID を利用する検索方法（本方式）の2つの検索方法について、検索時間とデータベースのデータ量を比較したものである。

表1 検索時間とデータ量

検索方法	検索時間(秒)	データ量(MB)
親ログ ID 利用なし	0.821 (100%)	421 (100%)
親ログ ID 利用あり	0.628 (76%)	533 (127%)

本方式によって、検索時間を24%低減できることを確認した。また本方式で用いるログ ID や親ログ ID の付加により、データベースのデータ量は本方式の方が27%大きくなった。

次にデータベースのインデックス機能を利用した場合の、表1と同様の比較を表2に示す。

表2 検索時間とデータ量(インデックス機能利用)

検索方法	検索時間(秒)	データ量(MB)
親ログ ID 利用なし	0.340	693
親ログ ID 利用あり	0.001	709

インデックス機能利用により、2つの検索方法とも検索時間が低減されたが、本方式の方が低減割合が大きく、2つの検索方法の検索時間を比較すると、本方式によって1/340に低減されている。また、データ量はインデックス用データ付加のためにそれぞれ増加している。

以上より、本方式は検索時間の低減にあたって有効であり、その効果はデータベースのインデックス機能との併用でより大きくなることが確認できた(図2)。

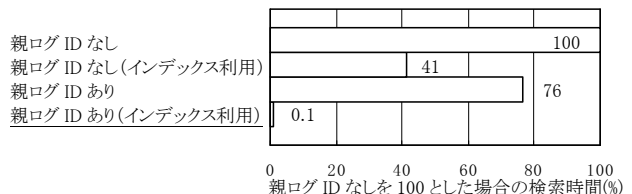


図2 検索時間低減効果

しかしながら、データベースのデータ量の増分も無視できない量である。よって本方式は、ストレージリソースのコスト増加よりも、情報流通把握にかかる人件費などのコスト低減が優先されるユースケースに適用すると思われる。

9. まとめ

本方式によって、組織全体における情報流通を追跡・可視化できることを示した。また、本方式における親ログ ID によるログ関連付けは、ログ追跡時間低減に効果があり、その効果はデータベースのインデックス機能利用によって、より大きくなることを示した。

本方式は情報流通把握の迅速化に効果があるが、データベースのデータ量増加も招くため、データベースのストレージコストよりも情報流通把握にかかるコスト削減が優先されるユースケースに適用すると思われる。

参考文献

- [1]「JP1」, <http://www.hitachi.co.jp/Prod/comp/soft1/jp1/index.html>(2012/1/12 アクセス)
- [2] 榊原裕之、桜井鐘治「ログ分析による情報漏洩監視」『情報処理学会研究報告』Vol.2011-DPS-146 No.23