特集論文

統合ログ管理技術と履歴追跡型ログデータベース

平井規郎*
森山令子*
森川修一*

Integrated Log Management Technology and High Traceable Log Database

Norio Hirai, Ryoko Moriyama, Shuichi Morikawa

要　旨

内部統制やコンプライアンスへの対応が企業の不可欠課題となり，至るところで様々な業務記録(ログ)を蓄積・管理することが求められている。しかし，ログの大規模化，多様化及び複雑化によって，従来の技術ではログの活用が困難になりつつある。このような問題を解決するために，三菱電機では統合ログ管理技術の研究開発に取り組んでいる。

統合ログ管理技術の特長は次のとおりである。

(1)　高速処理

当社独自の高性能並列情報検索技術によって，大規模ログの高速処理を実現した。

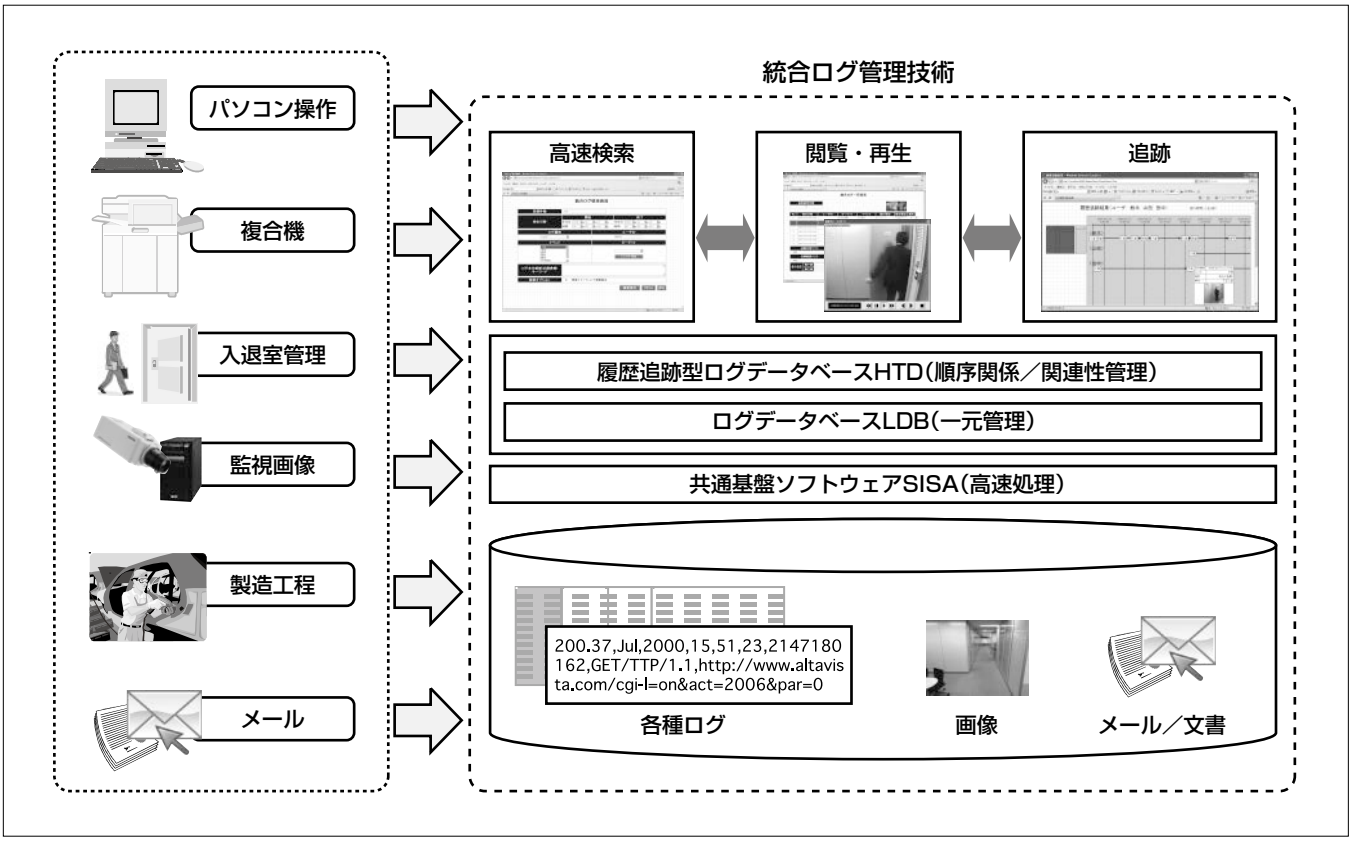
(2)　一元管理

ログデータベースLDB(Log DataBase)によって，様々な形式を持つログやメール，文書，画像などのデータをそのままの形式で蓄積し，あらゆるデータの一元管理を可能にした。

(3)　順序関係／関連性管理

履歴追跡型ログデータベースHTD(Highly Traceable log Database)によって，RDBMS(Relational Database Management System)に代表される従来型データベースでは不可能であったイベントの間の時系列的な順序関係や関連性の管理を可能にした。

本稿ではまずログ管理の現状と課題を述べ，次に当社が開発した統合ログ管理技術について述べる。また統合ログ管理技術の応用事例として，セキュリティログ管理システムや製造トレーサビリティシステムについて述べる。



統合ログ管理システムの概念図

統合ログ管理システムは，大規模なログの高速処理，多種多様なログの一元管理，さらにイベントの間の順序関係や関連性の管理によって，ログの統合管理と活用を実現する。

*情報技術総合研究所

1. まえがき

内部統制やコンプライアンスへの対応が企業の不可欠課題となり，至るところで様々な業務記録(ログ)を蓄積・管理することが求められている。例えば，セキュリティログ管理システムや製造トレーサビリティシステムなどでは，情報漏洩(ろうえい)事故や不良品の発生時の原因調査にログは欠かせない。しかし，蓄積されるログの大規模化，多様化及び複雑化が進むにつれて，従来の技術ではログの活用が困難になりつつある。当社ではこの問題を解決する統合ログ管理技術の研究開発に取り組んでいる。

2. 従来のログ管理の課題

従来のログ管理では，多くの場合RDBMSが利用されてきた。しかし，RDBMSを用いたログ管理には次の課題があった。

(1) 高速化

ログの規模は1日当たり1億件，年間で数十テラバイトに及ぶ事例もある。しかし，RDBMSは時系列的に追加される大規模ログの処理に必ずしも適していない。

(2) 一元管理

RDBMSで異なる形式を持つログを一元的に取り扱うためには，多種多様なログの形式をあらかじめ統一する必要がある。また，事前に形式を特定する必要があるため，システム構築時に想定していない新たな形式のログへの対応が難しい。

(3) 順序関係／関連性管理

ログの利用時には，様々なログの中に含まれるイベントの順序関係や関連性をたどって追跡，調査することがある。しかし，RDBMSではこのような順序関係や関連性を管理することが困難である。

3. 統合ログ管理技術

当社では**2章**で述べた課題を解決する統合ログ管理技術を開発した。

3.1 高性能並列情報検索技術による高速処理

高性能並列情報検索技術[1]を構成する次の技術によって，大規模ログの高速処理を実現した。

(1) 多数のプロセッサやディスク装置を効率的に利用する並列処理技術

(2) 効率的データ配置とデータ圧縮によるストレージアクセス技術

(3) 複雑な検索条件を高速処理するテキストフィルタリング技術

(4) 様々なデータを一元管理する異種データ統合管理技術

統合ログ管理システムは高性能並列情報検索技術を実装した共通基盤ソフトウェアSISA(Scalable Intelligent Storage Architecture)，及びSISAの機能を利用するログデータベースLDBと履歴追跡型ログデータベースHTDから構成される(**図1**)

3.2 ログデータベースLDBによる一元管理[2][3]

ログデータベースLDBは，データ形式にかかわらず格納可能な“本文”と，本文に付随する任意の“属性”の組によってデータを管理する。この構造によって，多様な形式を持つログを加工することなくそのまま蓄積可能にした(**図2**)。

ログデータベースLDBは，ログを蓄積時に形式を特定する代わりに，ログ蓄積後に正規表現[注1]で指定した条件によってログ形式を判別しながら利用可能であるという特長を備えている。従来の文字列照合方式では，ログ“本文”の形式判別で十分な速度が得られなかったが，ログデータベースLDBでは複雑な条件でも1億文字/秒の速度で照合可能な高速文字列照合技術sDFA[4]によってこの問題を解決した。また，ログ形式の判別のための正規表現を自動的に生成する機能を利用することによって，CSV(Comma Separated Values)形式のような代表的なログ形式や，値の範囲指定などの頻繁に利用される正規表現を容易に作成できる。

ログデータベースLDBは，画像やメール・文書のような様々なデータの保管庫として利用できる。この特長を利用し，ログデータベースLDBは，映像監視システムと連携した統合ログ管理システムにおける監視画像(入退室時スナップショット画像)の保存[2][3]や，メールアーカイブシステム[5]におけるメール本文や添付ファイルの格納に利用されている。

(注1) 文字列のパターンを表記する方法

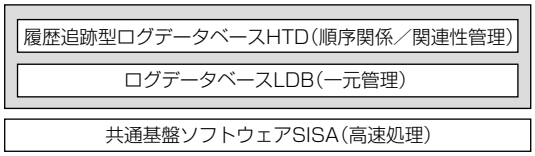


図1．統合ログ管理技術

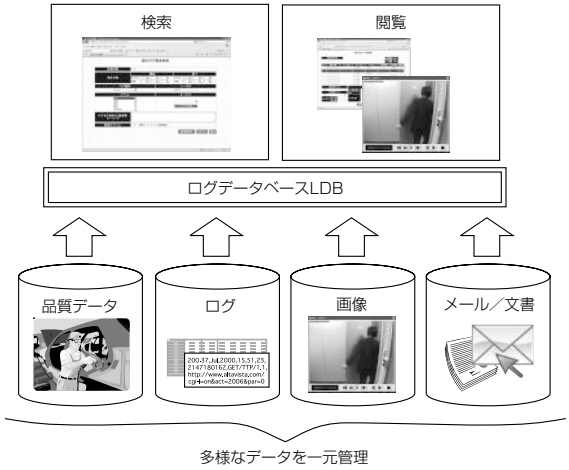


図2．ログデータベースLDB

さらにログデータベースLDBは，蓄積するデータを自動的に圧縮する機能を備えており，典型的なログでは，ストレージ容量を約1/10以下に縮小できる。

ログデータベースLDBは三菱電機インフォメーションテクノロジー㈱（MDIT）のLogAuditor(注2)の中核的機能として利用されている。

3.3 履歴追跡型ログデータベースHTD(6)による順序関係／関連性管理

ログが威力を発揮する場面の一つに，事故など，ログに記録されたイベントの原因や影響範囲の調査がある。しかし従来のRDBMSでは，ログに含まれる時系列的な順序関係や関連性を管理できないため，効率的な調査が困難であった。例えば，“ログイン”自体は正当な操作であっても，“入室”の操作を行わずに“ログイン”を行うのは不審な行動とみなせる場合がある。しかし従来のRDBMSでは，ログの中から“入室記録なしにログイン記録のある人”を直接検索することはできないため，ログインしたすべての人のログイン前の行動を人手で確認する必要があった。

履歴追跡型ログデータベースはこの問題を解決するために開発した技術であり，**図3**に示すように様々なログに含まれる情報の順序関係や関連性を管理し，順序関係を考慮した検索を可能にする。

図4は入退室ログや複合機ログに履歴追跡型ログデータベースを適用した例である。このように，順序関係や関連性をグラフ構造として管理することによって，人物などの行動の追跡や，個々のイベントを照合し特定のイベントが発生した順序を考慮した検索を可能にする。これによって例えば“入室操作をすることなくログインした”不審人物などを検出する。

（注2） LogAuditorは，三菱電機インフォメーションテクノロジー㈱の登録商標である。

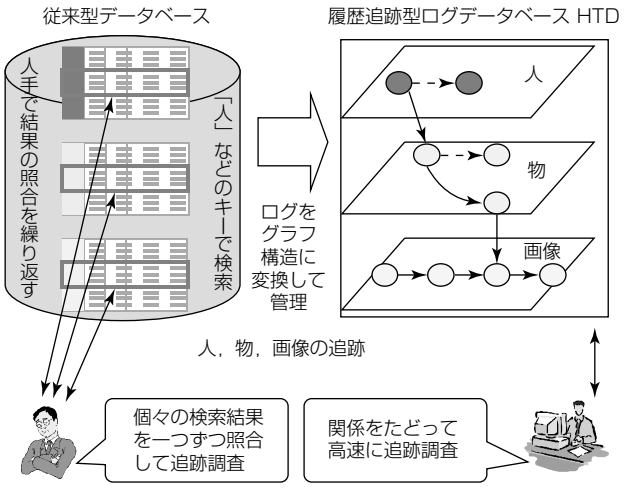


図3．履歴追跡型ログデータベースHTD

4. 応用事例

統合ログ管理技術は，様々な応用に適用することが可能である。ここでは，情報システムや入退室管理システムなどのログや監視映像を統合管理し活用するセキュリティログ管理システムへの応用事例，及び製造業で各工程のログを収集蓄積し活用する製造トレーサビリティシステムへの応用事例について述べる。

4.1 セキュリティログ管理システムへの応用事例

図5はオフィス内のパソコン，複合機，入退室管理装置などのログとこれらの機器を監視する映像を蓄積・管理し，人の行動や物の移動を監視するセキュリティログ管理システムである。

一例としてこのシステムで，すべての操作を管理するユーザーIDカードを紛失した場合，次の調査が想定される。

(1) 膨大なログから，紛失したユーザーIDカードがその後どのように使用されたかを調査し，状況を把握
(2) ログ以外に，監視カメラによる映像から人物を確認
(3) 不正に使用した人物の行動範囲を追跡し，カードの紛失による影響範囲を調査
(4) 類似の不正が疑われる不審行動の監視

図6に統合ログ管理技術を適用したセキュリティログ管理システムの調査例を示す。

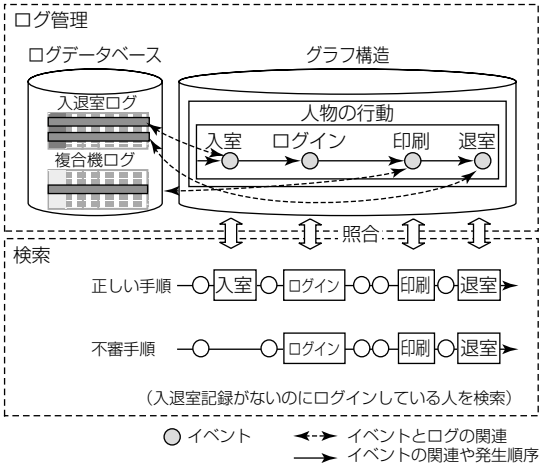


図4．ログ管理技術と検索技術

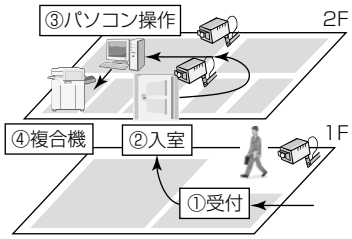


図5．セキュリテイログ管理システム応用事例

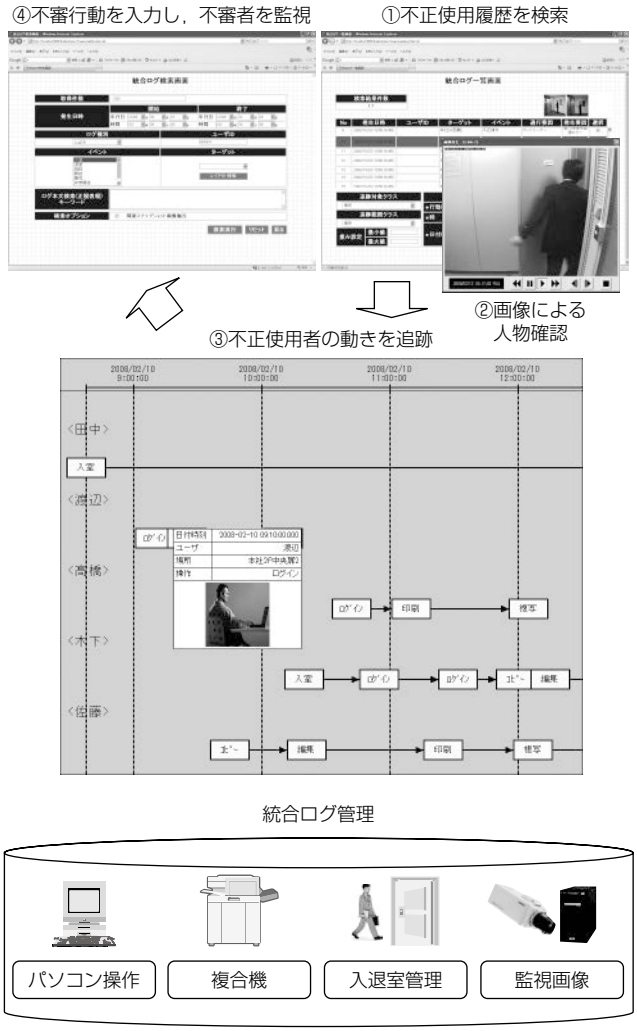


図6．セキュリティログ管理システムにおける調査例

(1) 紛失したユーザーIDカードの使用履歴を調査するために収集した様々なログ(入退室ログ，パソコン操作ログ，複合機ログ)から高速に履歴を検索し(**図6①**)，状況を把握する。

(2) ログと画像の一元管理によって，紛失したユーザーIDカードで入室した人物の画像を確認する(**図6②**)。

(3) 様々なログから，順序関係や関連性をたどることによって，ユーザーIDカードを不正に使用した人物の行動を追跡し，入室先，使用したパソコンなどを特定し，影響範囲を調査する(**図6③**)。

(4) “入室することなくパソコンを操作した人”など特定の行動順序を不審行動として定義し検索／監視することによって，事故発生前に不審行動を検出する(**図6④**)。

このように，この技術は様々なログによる事故原因の調査などに有効である。

4.2 製造トレーサビリティシステムへの応用事例

この応用事例では，製造トレーサビリティシステムに統合ログ管理技術を適用した例について述べる。

図7に示すような製造トレーサビリティシステムを想定する。このシステムでは原料仕入れ，加工工程，包装工程，

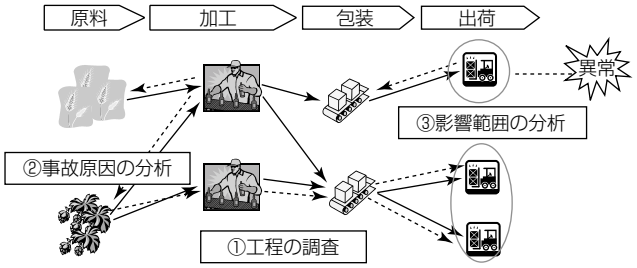


図7．製造トレーサビリティシステム応用事例

出荷の各工程でログが収集・蓄積されている。出荷後に不良品が発生した場合は，次のような調査を実施する。

(1) 不良品がどの工程で発生したかを，全工程のログから検索し状況を把握する。

(2) 各工程のログから不良品の発生原因となった設備を確認する(トレースバック)。

(3) 不良品の発生原因となった設備で処理された製品の出荷先を追跡し，影響範囲を調査する(トレースフォワード)。

なおこれまで述べた応用事例以外にも業務システムの障害解析など，多種多様なログやデータを収集しているあらゆる分野のログ管理システムに適用することが可能である。

5. む す び

本稿では当社が開発した統合ログ管理技術について述べた。この技術は，多種多様な大規模ログを効率よく活用することを目的として開発され，これまで個別に管理されてきたログやデータを関連付けて統合管理することを可能にした。今後もこれらのログ管理技術を発展させることによって適用分野を拡大し，ログ活用の幅を広げていく予定である。

参 考 文 献

(1) 郡　光則，ほか：高性能並列情報検索技術，三菱電機技報，**83**，No.12，705～708（2009）

(2) 山岸義徳，ほか：入退管理・映像監視システム向け統合ログ管理方式，情報処理学会FIT2008第7回情報科学技術フォーラム（2008）

(3) 小山明伸，ほか：物理セキュリティ情報の統合管理を実現した“LogAuditor”，三菱電機技報，**83**，No.9，563～566（2009）

(4) 中村隆顕，ほか：大規模正規表現の高速照合方式，第67回情報処理学会全国大会（2004）

(5) 大塚哲史，ほか：1000万件のメールを1秒で検索する“LogAuditor Mail Saver”，三菱電機技報，**82**，No.7，461～464（2008）

(6) 平井規郎，ほか：履歴追跡型データモデルの評価，日本データベース学会論文誌，**7**，No.3，73～78（2008）