Vol.2010-IOT-9 No.6
2010/5/13

組織内電子文書の追跡手法の提案

薄田昌広[†]　　上原哲太郎[††]

組織内の電子文書の流出事故は無くなる傾向にない。流出が発生した際、発見が遅れたてしまうと被害が拡大する恐れが非常に強い。また、流出経路を特定できないと、再発の恐れもある。本研究では、組織内で共有される文書に対して、組織や個人固有の識別子を自動で埋め込むことで、流出の早期発見や流出経路の特定を行う手法を提案し、システムを試作して動作を検証した。

A Proposal of Tracking Method of Digital Documents in the Organization

MASAHIRO SUSUKITA[†]　TETSUTARO UEHARA[††]

Leak incidents of internal digital documents from many organizations do not tend to stop. In the event of leakage, and discover later, the damage of the organization will be stronger. And, another incident may be happen again unless anyone identifies the leak path. In this paper, we proposed a method of early discovering the fact of leakage and identify the leak path by information embedding in digital documents automatically. And we developed a proto-system to verify the method.

1. はじめに

組織内電子文書の流出は情報セキュリティ事故の中でも大きなウェイトを占めており、さまざまな対策が取られているが、大きく減少はしていない。原因のひとつとして、組織のメンバが、組織文書を外部へ持ち出し、私物の PC などで利用する際に、ウィルスやワームなどの悪意のあるソフトウェアによりインターネットへ流出するということがあげられる。2008 年度の個人情報漏えい調査によると、Web・Net による漏えいの約 40%をウィルスやワームが占めている[1]。利用者には自覚も悪意もないが、PC に感染したウィルスがローカルの保存データを P2P ネットワークへ流出させ、それが悪意のある一部の人間によって Web に公開されて広範囲に広がるという可能性は常に考慮する必要がある。また感染した PC を直接 Web サーバとしてインターネットに公開するという悪質なウィルスも存在する。

多くの組織では、イントラネット内で業務文書が共有されており、組織のポリシーに応じたセキュリティ対策がなされている。ただし、対策は組織外部からの内部文書へのアクセス防御に重点が置かれており、イントラネット内部へのログイン認証の仕組みや、組織メンバ用 PC のウィルス対策、ファイルの暗号化などが中心である。この場合、外部からの文書へのアクセスは難しいが、信頼できる組織メンバであれば、文書を外部へ持ち出すことは比較的容易である。

しかし例えば、業務量の増大により、組織内部のメンバが仕事を持ち帰ることによって文書が組織外へ持ち出され、さらに自宅の PC からインターネットへ流出した事例も数多く存在する。そして、このような持ち出しに対し、内部ユーザに対してアクセスや持ち出し手続きを過剰に厳格にした場合、必ずしも効果が得られるとは限らない。それだけではなく、手続きの煩雑さによる作業効率の低下を避けようとルールを破ることが常態化するという事態が発生する可能性も高くなる。結論として、完全に流出を避ける絶対的な方法は無く、常に流出の可能性があると想定しておく事が妥当である。

流出を完全に防ぐことはできないという前提では、外部に持ち出される文書を追跡できるかどうかが被害拡大の防止やその後の対策コストに大きく影響する。ここで、文書の追跡という問題は、流出が発生している事実を早期に発見することと、流出の経路を特定することの二種類の重要な要素に分割することができる。

筆者らは、画像ファイルに対して識別子を電子透かしとして実時間で埋め込み、持ち出し者を記録するシステムを試作しているが[2]、本稿では、この考えを発展させ、

† 関西電力株式会社
Kansai Electric Power Co. Inc.,
†† 京都大学
Kyoto University

組織内で共有管理する電子文書に対し、文書内部に識別子を埋め込むことでこれらの問題に対応する手法を提案した。さらに、小規模な組織イントラネットを想定して試作システムを設計・実装し、動作させて機能検証を行った。

2. 追跡の課題と手法の提案

追跡の課題を流出の発見と経路の特定の二つの問題にわけ、それぞれの対策について検討した。前提としては、組織内の共有文書はファイルサーバで共有管理されており、各メンバはログインすることで文書にアクセスすることができ、メンバ PC へダウンロードが可能となるというものである。

2.1 流出の発見(組織 ID による検索)

ある組織の内部文書が流出していることを早期に発見するためには、インターネットを何らかの形で巡回し、対象とする文書を検索する必要がある。組織の内部文書をすべて記録し、インターネット上に存在する文書と照合することができれば、理論上は検索が可能であるが、きわめて処理サイズが膨大であり、非現実的である。内部文書のすべての文書名を記録しておき、文書名で検索するようにすれば、処理サイズは減少するが精度は高くならない。同一の文書名は多数存在する上に文書名の変更は容易であるため誤ったものに一致したり、対象文書に一致しなかったりということが考えられる。

組織の内部文書によく利用される固有名詞などを検索キーとし、文書を全文検索すると、文書名の変更などに影響を受けないが、どのような固有名詞であっても多くの文書に使用されていることが多いため、キーをどう選択しても精度はある程度以上上げることはできない。

そこで、本研究では、既成の文字列と重ならない、組織に固有な充分に長い識別子(組織 ID)を新たに生成し、組織内で共有する文書本体に埋め込むことを提案する。この組織 ID を検索キーとして全文検索を行うことによってファイル名が変更されたり、ファイル内容の一部が書き換わったりした場合でも高い精度での検索が可能となる。

この場合、現在のインターネットの規模では、Web 全体を独自に巡回し、検索を行うことは多大なコストがかかるため、既存の検索サービスの利用が現実的である。そこで、組織 ID の生成ルールや埋め込む場所については文書の種類ごとに実際の検索サービスで検証を行う必要がある。検証結果については 3 章に記述した。

また、埋め込むタイミングについては、ID は組織全体で長期間共通であるとすると、組織文書のテンプレートなどへ事前に埋め込んでおくことが可能である。

2.2 経路の特定（個人 ID の埋め込み)

流出経路を特定するためにもっとも重要なことは、組織内のどのメンバが関わったかということである。流出ファイルが特定のメンバにのみ関連するもので、そのメンバの PC でのみ利用されていた場合は経路の特定は容易であるが、組織で共有している文書であれば多くのメンバが利用する可能性があるため特定が困難である。アクセス記録などから関連メンバに対して内部調査を実施することが考えられるが、それだけでは不確実であり、また、ウィルスなどによる流出の場合は本人に認識がない場合もある。

本研究では、経路の特定についても、検索に使用したのと同じくメンバ固有の識別子(個人 ID)を文書へ埋め込むことで関連したメンバを特定することを提案する。個人 ID は組織内で各メンバに一意に割当てる。

この際、共有管理のファイルサーバからメンバの PC へダウンロードを行った時点でファイルへの管理ができなくなるため、ダウンロード時点で ID を実時間で埋め込むこととする。個人 ID は、組織 ID と異なり、直接検索の対象とはしないので Web 検索を考慮する必要はない。ただし、実時間で埋め込みを行うため、埋め込みに複雑な操作を伴うなら実用的なシステムにはならない。本研究では、文書の管理領域に個人 ID を埋め込むこととした。

3. 試作システムの設計と実装

提案した手法を元にして試験システムを設計し、小規模な組織の標準的なイントラネットを想定して実装を行った。

3.1 設計の前提

システムを試作するにあたり、対象として、小規模組織の標準的なイントラネットの構成を想定した(図 1)。組織のメンバは個人の PC を利用し、Web アプリケーションにより情報共有を行っている。ファイル共有も Web アプリケーションで行っており、サーバのファイルにアクセスするためには、サーバに登録したアカウントによるログインが必要である。文書ファイルの編集は PC のアプリケーションで行う。

また、文書ファイルの種類として、一般的に広く利用されている Microsoft Office 形式の三種類(Word、Excel、PowerPoint)を対象とした。

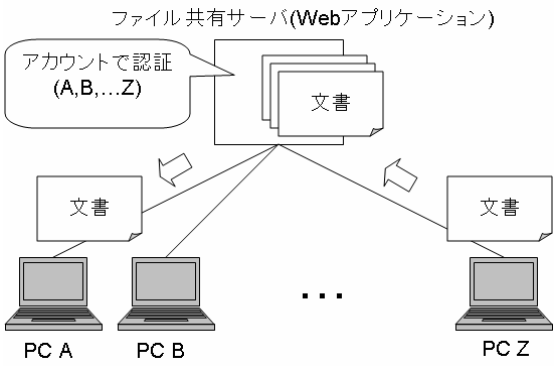

図 1　Web アプリケーションによる標準的なファイル共有
Figure 1　Typical File Sharing by Web Application

3.2 埋め込み機能の設計

本システムでは、ファイル共有サーバに保存されている元ファイルを、メンバがダウンロードした際に、メンバごとにその個人 ID が埋め込まれた少しずつ異なる文書に変換する必要がある。そこで、ファイル共有アプリケーションに変更を加え、メンバ PC からファイル共有サーバへのダウンロード要求が送信された時点で、対象ファイルに個人 ID を埋め込んだ複製ファイルを一時的に作成するとともに、リンク先を複製ファイルに差し替えることとした。

今回のシステムでは、ファイル共有に Web アプリケーションを想定しているが、Web アプリケーションでは動的に Web ページ作成を行うため、ダウンロードページ作成部分のプログラムを修正することで埋め込み機能を実現することができる。

3.3 検索機能の設計

検索機能は、インターネットと接続した検索サーバにより、Web サービスを利用して実現するよう設計を行った。

検索サーバは組織 ID をオンラインあるいはオフラインで受け取り、組織 ID を検索キーワードとして、一定時間ごとに Web サービスを呼び出して発見した場合は自動でダウンロードを行うこととした。サービス提供サイトがネット巡回によって流出ファイルを発見していれば、その文書を得ることができる。

ダウンロードしたファイルについては埋め込まれている個人 ID を調査し、ID のリストと照合を行うこととした。

3.4 埋め込み機能の実装

ログインによりアクセス管理ができるファイル共有アプリケーションで、ソースが公開されているものを選び、個人 ID の埋め込み機能を追加実装した。ダウンロード要求時に実行される部分のプログラムを変更し、要求されたファイルへ個人 ID を埋め込むプログラムを呼び出す処理を追加した。さらにダウンロード指定されたファイルを埋め込みファイルと差し替えて PC へ渡すよう変更した。

ID 埋め込みプログラムは Java で記述した。ファイル名と個人 ID を引数として渡すとファイルへの埋め込み動作を行う。埋め込み場所は、三種類のオフィス文書に共通して存在するプロパティ領域を選択した。ファイルへの埋め込み操作は Microsoft Office の互換ソフトウェアである OpenOffice.org の API を埋め込みプログラムから呼び出すことで実現した。

3.5 検索機能の実装

事前調査として、実際の Web の検索サービスに対し、どの程度の文字列が検索対象となるかということと、文書ファイルのどの部分が検索対象となっているかということについて検証を行った。結果として、32 文字の英数字からなる文字列を利用しても検索が実行できることや、本文以外に Word ファイルの透かし領域やヘッダ・フッタ領域なども検索対象となることを確認した。よって、組織 ID はかなり柔軟に生成でき、埋め込む場所の選択肢も多いことがわかった。

また、検索に使用する Web の検索サービスを比較するために、複数文書を複数のサイトで公開し、一定期間後に検索したところ、Google が最も多くのファイルを発見することが確認できたため、今回の実装では Google の Web 検索 API を採用することとした。

検索プログラムは組織 ID 検索機能部と個人 ID 検査機能部に分けて構成した。検索機能部は定期的に組織 ID を検索キーとして Web 検索を行い、組織 ID を含む文書を発見した場合はダウンロードする。検査機能部は文書のプロパティ領域を検査して個人 ID を特定する。このプログラムは Java で記述した。

また、今回の試作システムでは、組織 ID は事前に組織文書の雛形に埋め込んでおくという前提としたため、組織 ID の埋め込みに関してはプログラム実装を行っていない。

4. 機能検証と性能評価

試作システム全体を構成し、個別の機能が正しく動作することを確認するとともに、実時間で行う必要のある個人 ID の埋め込みについては埋め込み時間を測定して性能を評価した。

4.1 機能検証

検証に用いたシステム構成を(図 2)に示す。また、埋め込み機能を持つファイル共有サーバのスペックは以下の通りである。

・CPU - AMD Athron 1640B 2.7GHz
・Memory - 4GBytes
・OS - Linux

このシステムでサイズの異なる試験用の三種の文書ファイルを用意して共有し、埋め込み機能と検索機能の検証を行った。

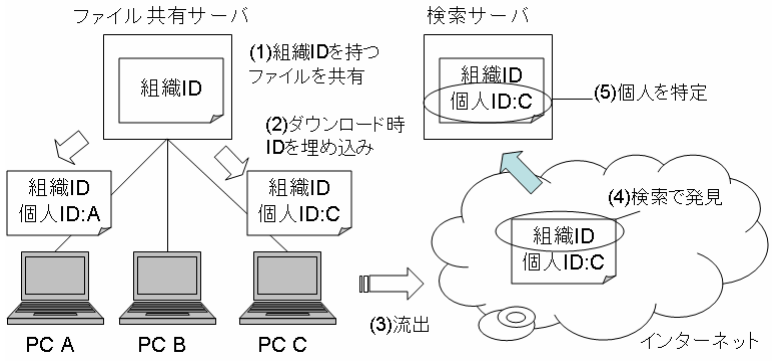

図 2　試作システムの構成
Figure 2　System Construction of Proto-system

このサーバに対して、別アカウントでログインしてダウンロードを行うことでそれぞれのファイルに個人 ID が埋め込まれていることが確認できた。

また、試験用のダミー文書に組織 ID と個人 ID を埋め込み、Web で公開して検索サーバで検索を行ったところ、対象のファイルを発見でき、ダウンロードして個人を特定することができた。

4.2 性能評価

埋め込みによるダウンロード時間の伸びは 100 キロバイト程度のファイルであれば文書の種類によらず 1 秒程度に収まっているが、ファイルサイズ 1 メガバイトの Word ファイルのみ埋め込み時間が 10 秒を超えている(表 1)。

表 1　三種類の文書への埋め込み速度
Table 1　Embedding rates in Three Types of Documents.

filesize	Word	Excel	PowerPoint
100kBytes	1.0 sec	1.1 sec	0.9 sec
1MBytes	13.9 sec	2.0 sec	2.1 sec

5. まとめ

本研究では、組織内で共有する電子文書に対し、組織固有の ID と組織メンバ固有の ID の二種の ID を埋め込むことでインターネットへの流出を早期に発見し、経路を特定することのできる手法を提案した。また、この手法を適用した試験システムを設計・実装し、実際に動作させて機能検証と性能評価を行い、手法の有効性を確認した。

ただし、今回の試験システムの対象は小規模組織であり、文書ファイルの種類によっては処理時間が長くなる場合もあるため、実用性については高速化も課題となっている

今後は小規模組織のイントラネットで試験用システムを動作させ、検証を続ける予定である。

参考文献

1)　2008 年情報セキュリティインシデントに関する調査報告書 Ver. 1.3
http://www.jnsa.org/result/2008/surv/incident/2008incident_survey_v1.3.pdf
2)　薄田昌広、上原哲太郎、岡田満雄: 電子透かしによるコンテンツ流出抑止システムの試作、FIT2008 講演論文集第 4 分冊, pp.163-164(2008)