デジタルデータの分散バックアップ方式の提案

半田 富己男*　　矢野 義博

大日本印刷株式会社ビジネスフォーム事業部ビジネスソリューション開発本部

〒162-8472 東京都新宿区榎町 7 番地

論文要旨

IT 革命の進展はデジタルデータの飛躍的増大を生み，ストレージ・サービス・プロバイダ(SSP)ビジネスが生まれた。これらの SSP はインターネットデータセンター(iDC)設備を利用し，大容量・高可用性を訴求している。

本稿ではデジタルデータを分割し、複数の iDC に分散保管することにより、単独の iDC に保管する場合に比べて、セキュリティ及び利便性を向上させることができる“分散バックアップ方式”を提案する。

キーワード

iDC, インターネットデータセンター, SSP, ストレージ・サービス・プロバイダ

1. はじめに

1.1. 背景

e-Japan 戦略のもとでのブロードバンド・インフラの整備と並行した IT の進展により，電子商取引などの e ビジネス，更には社内システムのアウトソーシングなどさまざまなニーズが生まれた。これに対応してサーバ・システムの構築，管理，運営を請け負うインターネットデータセンター(iDC)事業者が急速に増加した。iDC は堅牢な設備を提供し，耐障害性，可用性を高めたり，インターネットへの高速大容量バックボーン接続を提供することを特徴としている。

これらの iDC は需給ギャップの中で，アプリケーション・サービス・プロバイダ(ASP)事業者へのホスティングやコロケーションなどの付加価値サービスを提供して差別化を図っている。

e-Japan 戦略第一フェーズを受けて制定された「元気・安心・感動・便利」社会を目指す e-Japan 戦略Ⅱのもとで，安全・安心なデジタルコンテンツ流通体制の確立が求められている。

1.2. ストレージ・サービス・プロバイダ(SSP)

このような背景からストレージ需要が高まり，時を同じくして，ファイバチャネル技術に基づいた SAN(Storage Area Network)等の新しいストレージ技術が実用化されつつある。一方，こうしたストレージ管理は，コスト的にも技術的にも，IT 管理者にとって大きな負担となってきている。

そこで，iDC 事業の付加価値サービスの一環として，顧客に新しい技術によるストレージを提供するとともに，その管理を請け負うストレージ・サービス・プロバイダ(SSP)ビジネスが生まれ，成長しつつある。SSP は，利用者が必要な時に必要な場所で，簡単にストレージが利用できるよう，ストレージを貸与して管理を代行するビジネスである。

1.3. 問題点の整理

ここで，SSP サービスに求められる要件を考察する。SSP は，その性質上，複数の顧客データを預かるので，SSP 事業者は，個々の顧客のストレージ利用目的やデータ内容を知る立場にあってはならない。データ漏洩や改ざんにつながる要因を排除することが重要である。

デジタルデータが元来持つ性質として，原本とまったく同じ複製を容易に作成できることがある。さらに複製されたことの証跡を確保するためには，さまざまな技術的課題がある。デジタルデータの原本を安全に保管・利用する仕組みが求められる。

2. 分散バックアップ方式の提案

ここでは1.3節で述べた問題を解決する方法として，分散バックアップ方式を提案する。

2.1. 構成要素

分散バックアップ方式の構成要素を図 1 に示す。

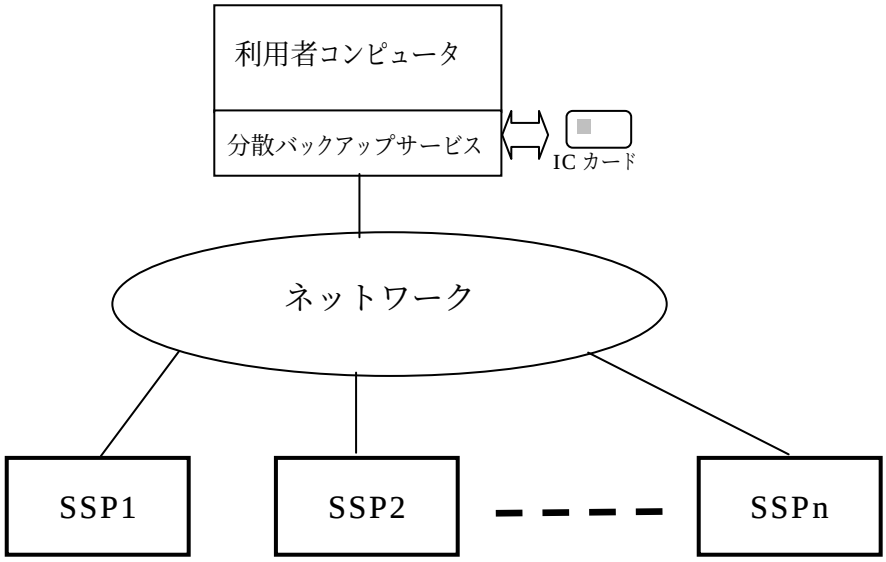

図 1 分散バックアップ方式の構成

ストレージサービスの利用者コンピュータには，分散バックアップ方式を利用するための分散バックアップサービスプログラム(以下，分散バックアップサービスと呼ぶ)がインストールされる。分散バックアップサービスは, 耐タンパー性を持った IC カード(以下, アーカイブカードと呼ぶ)に, 分散保管したファイルに関する管理データを記録する。

アーカイブカード内の不揮発性記憶領域には，分散バックアップサービス用の管理フォルダを設定し，この管理フォルダに分散バックアップ方式で保管したファイル毎にデータ保管場所およびデータ保管手順を示す管理データが格納される。アーカイブカードは，カード保持者として認証された利用者以外には管理フォルダへのアクセスを許可しない。

図 2 に，管理データの一例を示す。ここでは 3 つのファイル F1, F2, F3 を分散バックアップサービスを利用して保管した例を示す。各管理データは，各ファイルを構成するデータの保管場所を示すデータ保管場所情報と，データの保管手順を示すデータ保管手順情報とから構成されている。提案方式では, 保管対象となる 1 つのデータファイルは, 複数に分割され，複数の SSP に分散して保管される。データ保管場所情報は，保管先となっている複数の SSP の場所を示す情報である。

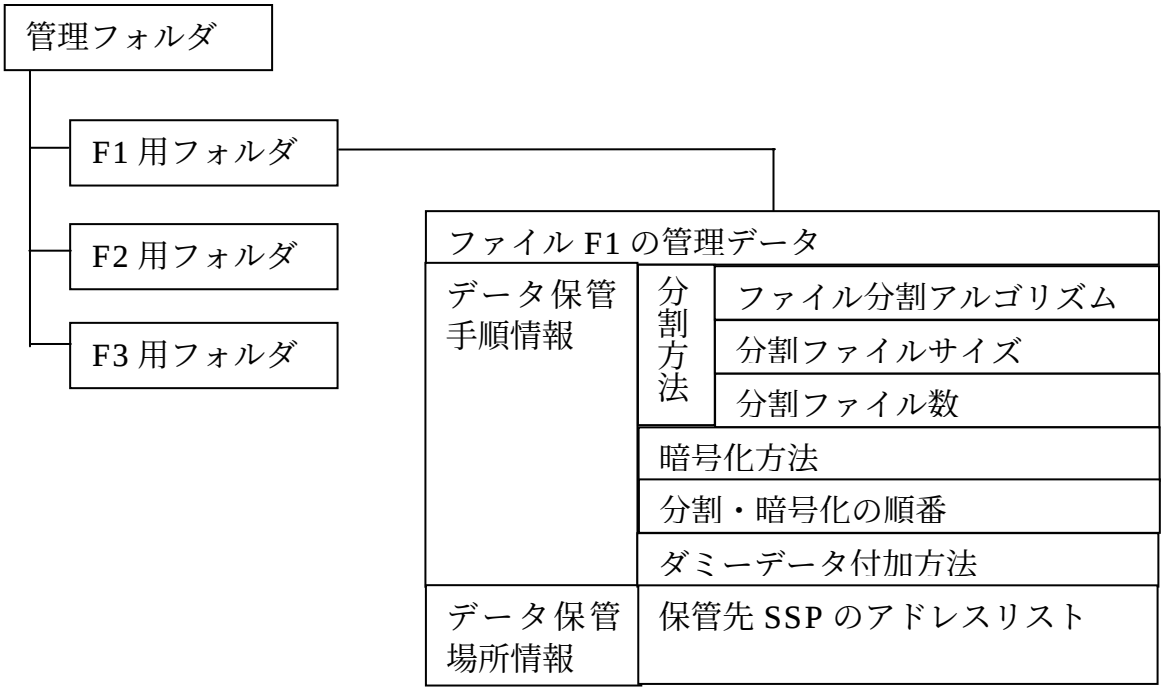

図 2 管理データの一例

データ保管手順情報は，図 2 の例の場合，「分割方法」，「暗号化方法」，「分割・暗号化の順番」，「ダミーデータ付加方法」を示す情報によって構成される。「分割方法」については更に，「ファイル分割アルゴリズム」，「分割ファイルサイズ」，「分割ファイル数」から構成される。たとえば，保管対象となる 1 つのデータファイル F1 を提案方式で保管する場合，このデータファイル F1 を複数のファイルに分割することになるが，どのような方法で分割を行うかという情報が「分割方法」に管理データとして格納される。すなわち，どのような「ファイル分割アルゴリズム」を用いて分割を行い，個々の分割ファイルのサイズと分割ファイル数が「分割方法」の細目として管理データに格納される。「暗号化方法」，「分割・暗号化の順番」，「ダミーデータ付加方法」は提案方式の実施形態におけるオプションであり，これらについては 2.2 節(1)で説明する。

2.2. 分散バックアップ方式

この節では，分散バックアップサービスの処理手順を説明する。分散バックアップサービスは，起動するとアーカイブカードとの間で認証に必要なデータのやりとりを行う。ここでは，PKI トークン用途の IC カードが一般的に備える相互認証機能を用いて，分散バックアップサービスがアーカイブカードの正当性を認証し，アーカイブカードは分散バックアップサービスの正当性を認証する。次に，アーカイブカードは，PIN(Personal Identification Number)や生体認証のメカニズムを用いて，カード保持者の本人認証を行う。アーカイブカードの持つアクセス制御機能により，カード保持者として認証された者のみが管理フォルダへアクセスできる。

カード保持者の本人認証に成功すると，分散バックアップサービスはアーカイブカードの管理フォルダの内容を読み出し，利用者に操作メニューを表示する。この操作メニ

ューには，保管対象データを新規に保管する保管処理，既に保管されているデータを取り出す取出処理がある。以下に，保管処理，取出処理の順で説明する。

(1) 保管処理

保管処理を選択した場合，利用者はまず保管対象ファイルを指定する。

次に，「ファイル分割方法」が決定される。すなわち，どのような方法で(「ファイル分割アルゴリズム」)，どのようなファイル長をもった(「分割ファイルサイズ」)，いくつのファイル(「分割ファイル数」)に分割するか，という条件を定める。

ファイル分割アルゴリズムとしては，「ファイルの先頭から等しいファイル長の連続領域に等分割する」といった trivial なアルゴリズムだけでなく，さまざまなアルゴリズムを採用することが可能である。例えば，保管対象データファイル F1 を 3 つのファイル F11, F12, F13 に分割する(分割ファイル数=3)場合，F1 の先頭から i バイト目の保管先として，$i \equiv 1 \pmod 3$ の場合は F11，$i \equiv 2 \pmod 3$ の場合は F12，$i \equiv 0 \pmod 3$ の場合は F13，へ保管するといった分割アルゴリズムも有効である。どのようなファイル分割アルゴリズムを採用するかによって，セキュリティ・レベルのの選択が可能なことが，提案方式の特徴の一つである。

このあと，分割バックアップサービスは「暗号化方法」,「分割・暗号化の順番」,「ダミーデータ付加方法」の決定を行う。これらの条件および「ファイル分割方法」の決定は，利用者によって指定させてもよいし，分割バックアップサービスがデフォルト値を決定してもよい。

「暗号化方法」の決定では，どのような暗号アルゴリズムを用いてファイル暗号化を行うか，どのような暗号鍵を用いるか，を決定する。「分割・暗号化の順番」には，各分割ファイルごとに暗号化を行うか否かといった事項や，分割処理後に個々の分割ファイルに対して暗号化を行うのか，あるいは，保管対象ファイルを暗号化した後に，これを分割するのか，といった事項を決定する。

保管対象ファイルを分割するプロセスにおいて，保管対象ファイルとは無関係なダミーデータを付加する処理を行った場合には，どのような方法でダミーデータを付加したかを示す情報を「ダミーデータ付加方法」に格納する。このような処理の例として，ランダムな任意のデータを発生させて，これをダミーデータとして利用することもできるし，あらかじめ用意しておいた何らかのデータをダミーデータとして利用してもよい。ダミーデータの付加は，分割前の保管対象ファイルに対して付加してもよいし，分割後の暗号化前のファイルに付加してもよいし，暗号化後のファイルに付加してもよい。このようなダミーデータを付加しておけば，万一，保管されているデータを盗み見られたり復号されたとしても，ダミーデータが介在しているために完全な復元には至らないので，セキュリティが更に向上することになる。

以上のようなデータ保管手順情報が決定されると，分散バックアップサービスは保管対象ファイルに対して実際に分割処理，暗号化処理，ダミーデータの付加処理を行い，複数の分割ファイルが作成される。

続いて，個々の分割ファイルが保管先となる SSP へ転送される。全ての分割ファイルの転送が終了すると，アーカイブカードの管理フォルダに，保管対象データファイル F1 に関する管理データが記録される。

(2) 取出処理

アーカイブカードの管理フォルダには，このアーカイブカードを使って，これまでに分散バックアップしたファイルに関する管理データが記録されているので，利用者はこのアーカイブカードを使って復元することができるファイル名を知ることができる。図

2 の例の場合 F1, F2, F3 の 3 つのファイル名が表示される。ここで利用者が F1 の取出を指定すると，アーカイブカードの管理フォルダから「ファイル F1 の管理データ」が読み込まれる。分散バックアップサービスは，この管理データ内のデータ保管場所情報から, 取出対象ファイル F1 を構成する個々の分割ファイルが保管されている SSP のアドレス情報を認識することができ，データ保管手順情報を参照すれば，保管時にどのような分割処理，暗号化処理，ダミーデータ付加処理が実効されたかを認識することができる。

分散バックアップサービスは，データ保管場所情報に基づいて，SSPs から取出対象ファイル F1 を構成する個々の分割ファイルを取り寄せ，データ保管手順情報に基づいて，個々の分割ファイルに対する復号化および統合処理が実行される。また，データの保管時に，ダミーデータを付加していた場合には，ダミーデータ付加方法に基づいてダミーデータを除去する。最後に，復元された取出対象ファイル F1 を，利用者が指定した記録場所に保管する。

2.3. 提案方式の特徴

この節では，提案方式の特徴をまとめる。

静的デジタルデータを複数のファイルに分割し，暗号化したりデジタル署名を付与したりする。ファイル分割アルゴリズムの選択，暗号化アルゴリズムの選択，デジタル署名の有無などに自由度があるので，データの重要度に応じて利用者によるセキュリティ・レベルの選択が可能である。

個々の保管先 SSP では, 分割されたファイル断片を預かるだけなので, 元のデータの内容を知ることができない。

データ保管場所および保管手順を示す管理データは，利用者が管理するアーカイブカードの中に格納されているので，個々の保管先 SSP では，残りの分割ファイルがどこの SSP に預けられているかを知ることができない。

利用者は，分散バックアップサービスを利用できる場所であれば，どこでも，アーカイブカードを持参するだけで，元のデータファイル原本を取出すことができる。

3. 考察

3.1. 安全性の考察

一般に暗号解読を試みるものが，暗号文から平文が得られたことを知るのは，意味のある文が復元できた場合である。もとの平文が文字列として意味をなさない場合，暗号鍵の総当り攻撃で，たとえ解読に成功していても，それに気づかない。電子メールなどで，頻繁に使われる語句，例えば,「株式会社」や「御中」などが文中に含まれていると，暗号解読者に手がかりを与えることになる。提案方式では，ファイル分割アルゴリズムを適用することにより，平文自体が意味をもたない文になるので，このような解読の危険性を減らすことができる。図 3 に示すのは，ファイル分割アルゴリズムとして，ファイルの先頭から 1 バイト，3 バイトの順に取出し，3 個のファイルに分割した例である。

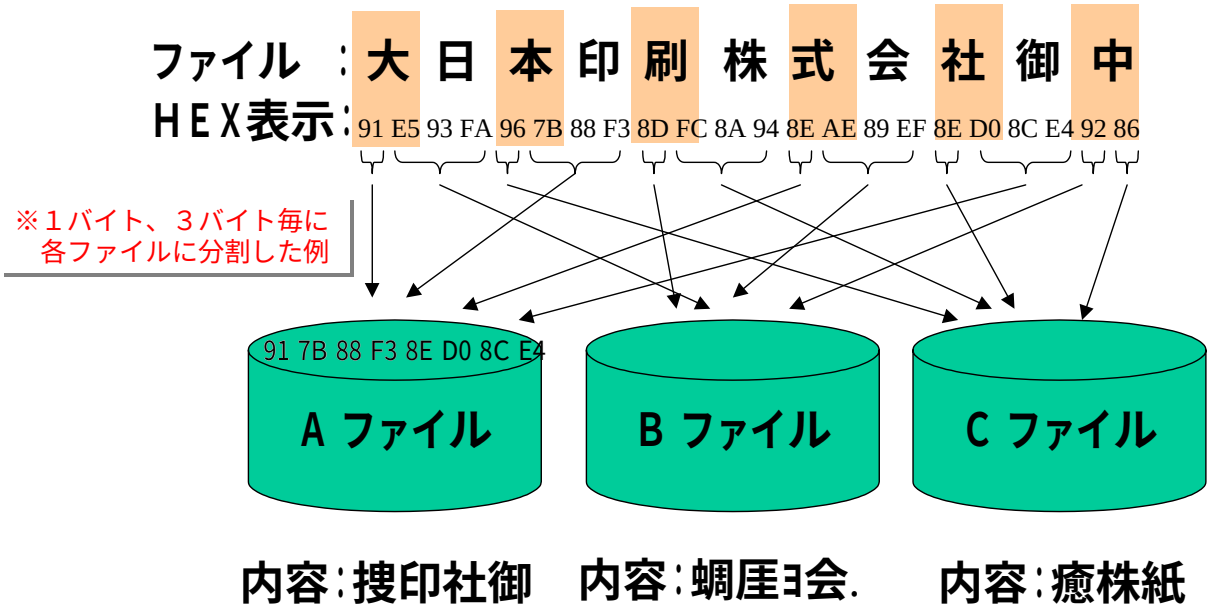

図 3 ファイル分割アルゴリズムの一例

4. まとめ

インターネットデータセンターを有効に活用し，ストレージ・サービスに関する課題を解決する方法として，“分散バックアップ方式”を提案し，その有効性を議論した。

提案方式では、データファイルの原本そのものを分散保管し、その保管位置情報のみを利用することにより,データファイル原本を損なうことなく、いつでも利用することが可能となる。

参考文献

[1] 平田 慎一郎, “ストレージサービスプロバイダの技術基盤”, ユニシス技報 72 号, 2002 年 2 月, http://www.unisys.co.jp/tec_info/tr72/72home.htm