MyCloud: 複数ベンダのクラウドを用いて構成する秘密分散ストレージ

堀内 公平
horiuchi.kouhei@un.hc.uec.ac.jp
電気通信大学

近年，クラウドコンピューティングが隆盛を極めつつある．しかしその技術は完全なものでなく，ユーザの立場に立った際に信頼性の観点からクラウド・ベンダによる囲い込み（ロックイン）などの問題が残っている．そこで本稿では，複数ベンダのクラウドを並列的に取り扱うことにより，何れのベンダにもロックインされずにより高い信頼性を得られる，新たなデータクラウド MyCloud を提案する．

MyCloud: A Secret Sharing Storage System Combining Several Vendors' Data Cloud Services

KOHEI HORIUCHI
horiuchi.kouhei@un.hc.uec.ac.jp
University of Electro-Communications

In recent years, Cloud Computing plays an ever-increasing role in the whole world. Although the service is useful, there are such problems about invasions of users' privacy and lock-in by Cloud vendors. This paper is to propose an improved scheme called "MyCloud" for resolving these problems. Using some vendors' Cloud Computing in parallel, "MyCloud" supplies the high reliability and liberates all the users from locking by the vendors

1. はじめに

1.1. 背景

近年，アプリケーション，システム，ハードウェアをネットワーク経由で利用するサービスの一形態であるクラウドコンピューティングが注目されている．コンシューマの立場からのクラウドコンピューティングのメリットとして，アクセス透過性，導入の容易さ，リソースが無制限である点などが挙げられる．一方，コンシューマ視点からの課題として，クラウドコンピューティングサービス提供者（以降ベンダと記述）に対する信頼性の問題，プライバシー問題，ベンダが企業戦略の一環としてコンシューマを囲い込むロックイン問題などが指摘されている[1]．

1.2. 目的

本稿では，これらの課題に対処するひとつの方法として，複数ベンダのクラウドを用いて構成する秘密分散ストレージ MyCloud[2]を提案する．

本提案の目的は，（1）システムの一部であるクラウドサービスが一定数以下停止しても，継続して利用できるシステムを構成すること（高信頼性），（2）クラウドからデータが漏洩，あるいはベンダがクラウド内のデータを利用しようとしても，意味ある情報を読み取れないシステム（強秘匿性），そして（3）高いユーザビリティを実現することとする．

2. 関連研究・関連技術

関連研究として，要素ノードの信頼性に一定以上の値を期待できない状態で高信頼なス

トレージを構築するという点において，井口氏による「信頼性を考慮したグリッド向け自律分散ストレージシステム」[3]が挙げられる．この提案の特徴として，ストレージノードの信頼性（稼働率）を常に監視し，その信頼性に応じてリード・ソロモン符号[4]の冗長度を動的に操作する手法が提案されている．本提案では自前のストレージの代わりに一般のデータクラウドを用いるため，コンピュータリソースの増減が非常に柔軟である点や，NAT などの存在を全く意識することなく機能を利用できる点において異なる．また要素技術としては，ベンダに意味のあるデータを渡さないための分散手法や，クラウドとの通信を管理するための機構などが必要となる．

また関連技術として日本ユニシスグループの「真性乱数を使用した秘密分散データストアサービス」[5]が挙げられる．これは顧客データを秘密分散して複数のデータセンタに置くサービスであり，秘密分散によってクラウド上のデータの安全性を高めるという点で本提案と関連性がある．しかしこのサービスはあくまで日本ユニシスグループ単体で完結しており，コンシューマはベンダからのロックインを逃れることはできない．

2.1. 提案概要

本稿では，複数ベンダのクラウドを用いて構成する秘密分散ストレージ MyCloud を提案する（図１）. MyCloud のフロントエンドは，ユーザ（ファイル操作を伴う上位システム）に対して複数のクラウドの存在を隠蔽し，あたかも単一のストレージのように見える．一方バックエンドでは，ユーザのデータに対して分散化・復号化の処理を施し，複数のクラウドとデータを分散化したまま通信し，ユーザデータの管理を行う．クラウド上ではデータは分散化状態でしか存在しないため，何れ

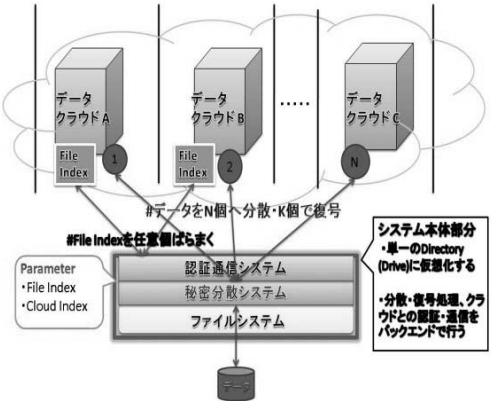

図 1　MyCloud の概要図

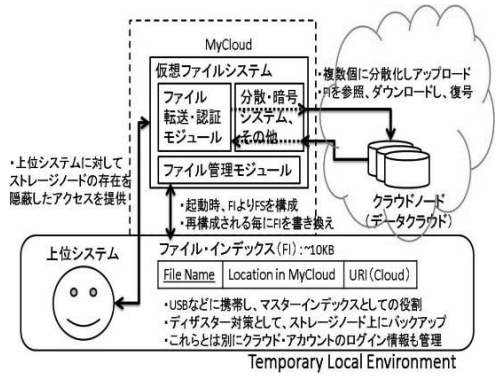

図 2　MyCloud の構成図

かのクラウドからデータが漏洩しても個人情報が漏れることがない．また分散手法については閾値秘密分散法のような冗長性を持つ方法を用いることにより，幾つかのクラウドがサービスを停止しても他のクラウドからデータを収集し元のデータを復号することができる．つまり MyCloud を使うことのメリットは，ユーザが何れのベンダにも依存せずに済むのと同時に，クラウドの通常利用時よりも高い信頼性と秘匿性を獲得できることにある．

2.2. 詳細設計

図２に MyCloud 全体をイメージ化したものを表す．以下より各要素に説明を加える．

2.2.1. MyCloud 本体

MyCloud の本体部分は以下に示すモジュールで構成される．

- 認証通信モジュール：
  各クラウドとの認証及び，分散データのアップロード・ダウンロードを行う．また，各クラウドノードにテストパケット（ping）を送信することによってそれらの生存確認を行う．
- 分散（暗号）モジュール：
  上位システム（ファイル操作を伴うアプリケーション）から渡されたデータを分散する．またクラウドノードからダウンロードしたデータを復号化する．なお分散手法には，データの機密性に応じて分散する個数 k と復元に必要な個数 n を容易に操作できる閾値秘密分散共有法[6]を想定している．また，分散法の代わりに共通鍵暗号方式による実装も並行して行っている．
- ファイル管理モジュール：
  ファイル・インデックスの情報から MyCloud のディレクトリ構造を再現する．
- 仮想ファイルシステムモジュール：
  ファイル管理モジュールの再現したディレクトリ構造から仮想的なファイルシステムを構築し，上位システムに対してファイルアクセスを提供する．また開発には Windows 上で仮想的なファイルシステムを構築できる Dokan Library[7]を用いた．

2.2.2. ファイル・インデックス

ファイルを MyCloud にアップロードする際，そのファイルの情報をファイル・インデックスに格納する．格納する情報は，ファイル名をキーとして，そのファイルの存在するクラウドの URI，MyCloud のディレクトリ構造内でのロケーションを情報として持つ．

このファイル・インデックスは，常に USB メモリに入れて持ち歩くことを想定し，絶対唯一のマスターインデックスとしての役割を果たす．つまり単一の MyCloud のサービスを２箇所以上で２人同時に操作するような事態は想定しない．以上の方法によりデータの整合性を保証する．

2.2.3. クラウドノード

分散したデータを預けるストレージ群．対象としては，一般にサービス展開しているデータクラウドを想定している．また，今回システムの有効性を拡張するため，自前の FTP Server もデータを預けるストレージとして利用する．

2.3. 機能説明

2.3.1. アップロード

フロントエンドでは，ユーザは上位システムから MyCloud にアップロードしたいデータを渡し，その機密性に応じて分散する個数 k と復元に必要な個数 n を指定する．

バックエンドの処理として，システムは受け取ったデータをローカル環境で k 個に分散し，それぞれをネットワーク上のクラウドノードにアップロードする．

2.3.2. ダウンロード

フロントエンドでは，ユーザはダウンロードしたい MyCloud 上のファイルを指定する．

バックエンドの処理として，システムから分散データを所有する k 個のクラウドノードに ping を送信し生存確認を行う．その後，最も TTL の短かったクラウドノード n 個から分散データをダウンロードし，ローカル環境でそれらを復号する．

2.3.3. ディザスタリカバリ

ファイル・インデックスが失われると，預けてあるファイルの名前，クラウドノードの

種類などが分からなくなるため，システムが利用できなくなる．そこでファイル・インデックスを，一定期間ごとに全クラウドノードに分散させておく．そしてクラウドノードからファイル・インデックスを復元する機構をMyCloud 本体に組み込んでおくことによって，メモリ内のファイル・インデックスを消失した場合も自身の MyCloud 環境を復元できる．

3. おわりに

本論文では複数ベンダのデータクラウドを用いて構成する秘密分散ストレージの構成法について述べた．

本提案の新規性は，クラウドを有効に活用することと，一切のサービス提供者からのロックインを逃れコンシューマ主体のストレージを構築することを両立した点にある．

システムの信頼性については，多くのクラウド・ベンダを利用するほど信頼性が向上することが自明である．例えば Amazon S3 の信頼性は 99.9%を謳っているが，仮に信頼性 99.9%のクラウドを 5 つ用いて復元必要数 2 で MyCloud を構築した場合，MyCloud の信頼性は 99.999999999%となる．これは基幹系のシステムの運用にも十分に耐えうる値である．

現在の進捗としては，個人データの隠蔽方法に AES 暗号[8]を採用し，クラウドノードとして単一のデータクラウドと，２箇所以上のデータセンタを用いた MyCloud を構築した（図３）．この場合のシステムは分散数ｋ・復号必要数１ということになる（但し復号には鍵が必要）．共通鍵暗号を使うことのメリットは（１）計算量が少ない，（２）実装が容易，の２点が挙げられる．

今後の予定としては，対応するクラウドサービスを増やしていくこと，秘密分散法による実装を行うこと．またそれらと同時に，ユーザビリティ向上のためにユーザインタフェース，キャッシュシステムの実装などを進めていく．また近い内にオープンソースソフトウェアとして MyCloud β版を公開し，ユーザのレビューを募っていく予定である．

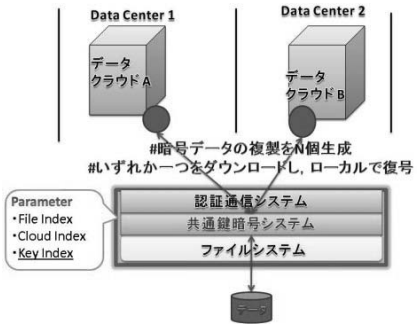

図 3　共通鍵暗号による実装

謝辞

本提案は独立行政法人情報処理推進機構により，未踏 IT 人材発掘・育成事業 2009 年度上期 未踏ユースに採択され，この支援を受けて開発を行っています．

参考文献

[1] M. Armbrust, et. al., “Above the Cloud: A Berkley View of Cloud Computing,” Technical Report, no.UCB/EECS-2009-28, Feb. 2009.
[2] 堀内，“複数ベンダのクラウドを用いた秘密分散ストレージ「MyCloud」の開発,” 2009 年度上期未踏 IT 人材発掘・育成事業（未踏ユース）採択案件, July. 2009.
[3] 井口他, “信頼性を考慮したグリッド向け自律分散ストレージシステム,” 情報処理学会論文誌：コンピューティングシステム, Vol. 47 No. SIG 7(ACS14), May. 2006.
[4] リード・ソロモン符号, http://www.siglead.com/technology.html
[5] 真性乱数を使用した秘密分散データストアサービス，http://www.unisys.co.jp/news/nr_091007_cloud.html, Oct. 2009.
[6] 保坂他, “秘密分散法とその応用,” 東芝レビュー, Vol.62, No.7 pp23-26, 2007.
[7] Dokan, http://dokan-dev.net/
[8] AES, http://csrc.nist.gov/archive/aes/index.html