

コンシューマ向け重複排除型オンライン・ストレージ・サービス —クラウドを、みんなで共有、クーモシェア—

1. 背景

インターネット上の情報は、時として、突然無くなる。お気に入りの登録してあったページが肝心なときに消えていたり、あとで見ようとチェックしておいた動画がアップロード者によって削除されてしまっていたりする。そのような状況を繰り返すうち、無くなると困る情報はパソコンに保存しておく習慣がついてくる。

ところが、それは非常に無駄なことである。なぜならば、無くなって困る情報というのは、多くの人にとって共通のものだからである。多くのファイルが、多くの人のそれぞれのパソコンの中に、何重にも重複して保管されているのが現状である。人気の動画などは、巨大なファイルであるにも関わらず、何十万人によって保管されているかもしれない。

2. 目的

そこで、「一旦ローカルに保管したファイルを、もう一度ネットに上げて、効率的に保管し直そう」というのが本プロジェクト(*KumoShare*)の目的である。ただし、ネットに戻した先にもそのファイルの管理者がいて、その管理者がいつ消してしまうか分からない状況では意味がない。各々のユーザにとっては、単にローカルに保存する代わりにネット上の自分専用領域にファイルを保管しているだけのように見えなければならない。

3. 開発の内容

3-1. システム構成

代表的なクラウド環境である Amazon Web Services(AWS)を利用して、重複排除型オンライン・ストレージ・システムを構築した(図 1)。開発したソフトウェアは、①EC2 内にデプロイされ、S3 や SimpleDB にアクセスする HaaS 型の Web サービス、②専用端末アプリ、及び、③ブラウザアクセス用の Web ページである。

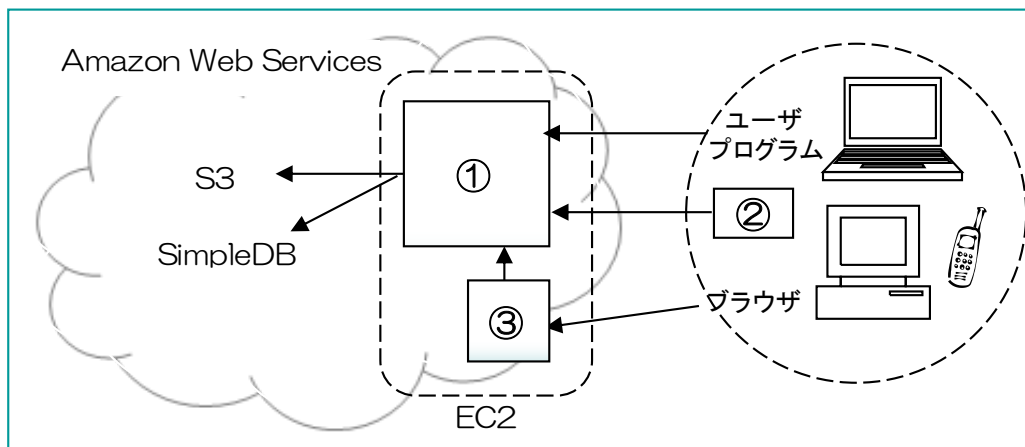


図 1 システム構成

3-2. ファイルサイズ分配機能（格納コスト分配機能）：特許第 4256397 号

本システムの基盤となる技術は、重複排除技術である。各ユーザはストレージ上の自分専用領域に個別にファイルを保管する。ストレージシステムは、各ユーザが格納しているファイルの中に同じファイルがあるかどうかを探し出し、同じファイルがあればシングルインスタンスとして管理し、各ユーザのファイルをそのインスタンスへのリンクに置き換える。

問題となるのは、このインスタンスの格納コストを誰がどのように負担するかである。本システムでは、ファイルの「実効サイズ」という概念を定義し、これに基づいて課金を行う。例えば、図 2 においては、ユーザ 1 とユーザ 2 が同じ 1GB のファイルを格納したところ、1GB のシングルインスタンスとして管理された状況である。このインスタンスの格納コストをこのインスタンスの受益者であるユーザ 1 とユーザ 2 とに公平に割り振るため、インスタンスの実サイズ(=1GB)を重複度(=2)で割ることにより、実効サイズ 0.5GB ずつをユーザ 1 とユーザ 2 にそれぞれ割り振る。ユーザ 1 とユーザ 2 は、それぞれ 1GB のファイルを格納しているにも関わらず、0.5GB のファイルを格納しているものとして扱われることになる。もちろん、重複度が 3 ならば、実効サイズは実サイズの 1/3 となり、重複度が 1000 ならば、実効サイズは実サイズの 1/1000 となる。この仕組みにより、ファイルの格納コストは多くのユーザで分け合うことになり、結果として格安で利用することができる。これは、S3 の格納コストを、インスタンス毎にユーザ間で“割り勘”することに他ならない。

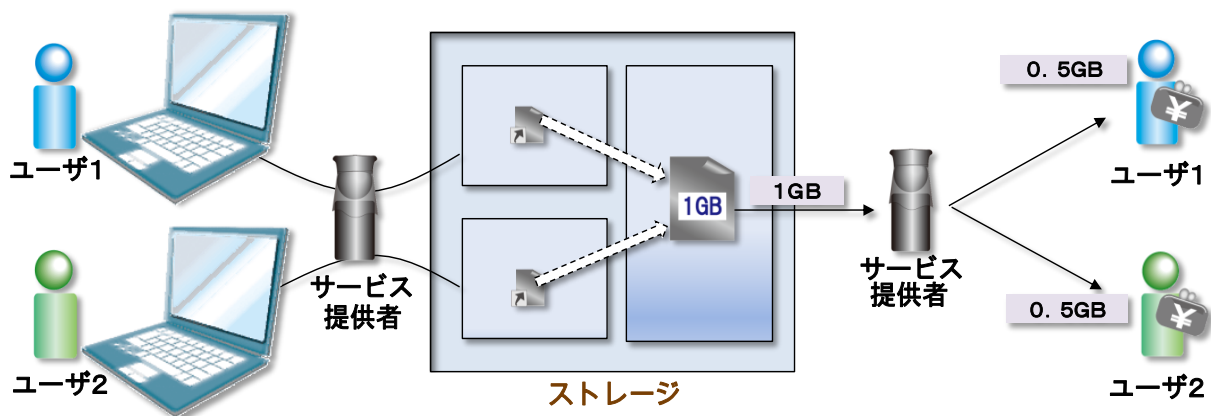


図 2 ファイルサイズ分配機能

3-3. 検証機能付き高速アップロード機能：特許第 4376960 号

通信面の重複排除技術を用いると、ファイル実体の重複した通信を避けることができる。端末からサーバにファイルを送信するときには、即座にファイルの実体を送信するのではなく、まずそのファイルのハッシュ値をサーバに送信する。サーバにそのハッシュ値の元となるファイルが既に格納されている場合は、その実体のアクセス権をハッシュ値を送信してきたユーザに与える。

ただし、不特定多数のユーザを対象にこのシステムを運用すると、不正に入手したハッシュ値を使って不当にファイルのアクセス権を得ようとする行為が危惧される。そこで、チャレンジ&レスポンス形式により、端末にそのファイルが本当にあることを検証した上でアクセス権を付与するシステムを開発した(図 3)。

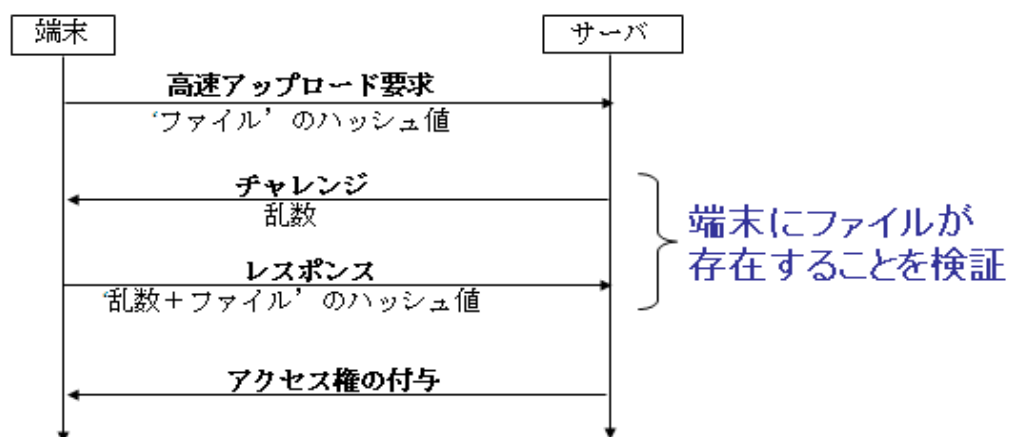


図 3 検証機能付き高速アップロード機能

3-4. 選択的同期機能：特許第 4386459 号

本システムでは、重複度の大きいファイルに関してはコスト面での競争力が高いが、重複していないファイルや重複度の小さいファイルに関しては、それほどでもない。むしろ端末のハードディスクに置いておいた方がコスト面で有利であることもある。

そこで、クラウドストレージとローカルストレージとのハイブリッド型のストレージアプリを開発した(図 4)。左端のアイコンがファイルの状態を示しており、雲マークがクラウドで管理されている状態、ハードディスクマークがローカルで管理されている状態、そして、回転矢印マークが同期中状態を表わしている。このアプリにインポートされたファイルは、システムがサーバ側に管理を移すべきか、端末側に残すべきかを自動判定し、サーバ側に管理を移す方が得な時にのみ、そのファイルをサーバにアップロードする。

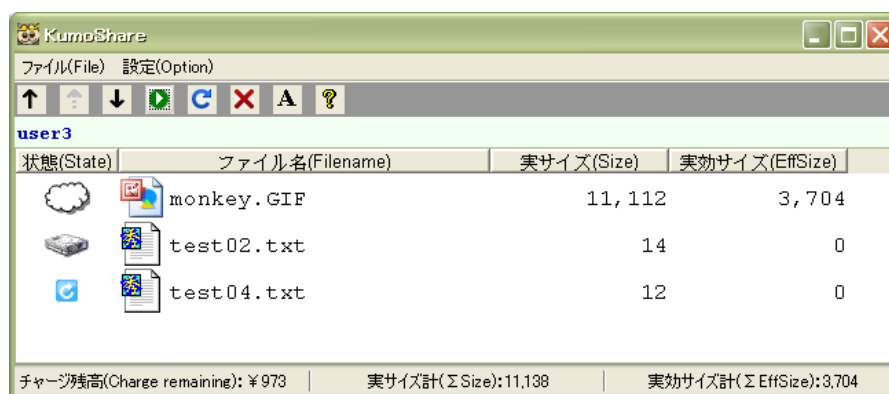


図 4 選択的同期機能

3-5. ファイル断片単位重複排除機能

ファイル単位のみで重複判定を行うと、ファイルの一部分のみが異なる場合でもファイル全体が別物として扱われてしまう。それに対してファイルを断片化して断片単位で重複判定を行えば、重複する断片については重複を排除できるので、より効率的に格納することができる。本プロジェクト期間中に、MPEG4 AVC フォーマットの動画ファイルに限定したファイル断片単位の重複排除機能を実現できた。

4. 従来の技術（または機能）との相違

ファイル共有サイトでも、皆が持っているファイルを効率的に保管することができる。しかしながら、ファイル自体を共有するため、ファイルの管理者が必要となる。その管理者が、ユーザの知らないうちにファイルを削除してしまうかも知れない。逆にユーザがファイルを削除してもらいたい場合には、管理者に削除依頼を出さなければならず面倒である。また、有料のストレージ上で共有するならば、誰がその費用を負担するかが問題となる。それに対して、本システムにおいては、自分のファイルは完全に自分の管理下に置かれるため、自分が削除しなければ削除されることはないし、削除したければ自分が削除すればよい。また、ファイル格納費用の分担も自動的に公平になされる。

また、Amazon S3をはじめとするクラウド上のストレージや、その他のオンライン・ストレージ・サービスにおいても、既に重複排除技術が利用されているかもしれない。しかしながら、それらのシステムでは、料金体系の中に重複状況が考慮されていないため、重複度の大きいファイルほど安く格納することができるという特徴を持たない。本システムでは、重複度の大きいファイルほど安く格納することができるので、重複度の大きいファイルを多く格納しようとしているユーザの利用を促すことができる。

5. 期待される効果

汎用のオンライン・ストレージとしての利用だけでなく、様々な Web アプリと併用することが可能である。例えば、ソーシャル・ブックマークと併用するならば、ブックマークを行うときにファイルの内容もすべて本システムに保存しておき、多くのユーザが保存したページは長期間保管するようにする。そうすれば、膨大なページを保管したとしても多くのユーザで格納コストを分担することができるので、一人当たりの負担額は抑制される。また、Web サービスとして API を公開することにより、他サイトとのマッシュアップも見込める。

6. 普及（または活用）の見通し

今後は実用面での検討を進め、1 年以内を目途に、本プロジェクトで実装したファイルサイズ分配機能（格納コスト分配機能）に基づく料金体系にてコンシューマ向けの有料オンライン・ストレージ・サービスの提供を目指す。

現在のところは、コンシューマ向けのオンライン・ストレージの市場規模は大きいとは言えないが、今後、クラウドコンピューティングが普及し、PC 内のハードディスクではなくクラウド上のストレージがメインストレージとなるにつれて、市場は急激に膨らむはずである。そのような時代に備えるため、数千万人～数億人が利用してもパンクしないようなスケーラビリティをもったシステムの実現を念頭に置いておきたい。

7. クリエータ名

後藤 誠

（参考）関連 URL

<http://kumoshare.com/>