

# ストレージサービスを前提とした緊急時のバックアップ先スケジュール

田井 裕次郎<sup>†</sup>   Le Hieu Hanh<sup>†</sup>   横田 治夫<sup>†</sup>

<sup>†</sup> 東京工業大学 情報理工学院 〒 152-8550 東京都目黒区大岡山 2-12-1

E-mail: <sup>†</sup>{tai,hanh,hlh}@de.cs.titech.ac.jp, <sup>††</sup>yokota@cs.titech.ac.jp

あらまし    災害時のデータ損失を避けるため、オンプレミスのデータを、ストレージサービスを利用して遠隔バックアップする対策が広く行われるようになってきた。しかし、定期的に遠隔バックアップを取る場合、最新のデータを残すためには、災害が発生する直前の定時バックアップからの差分データを、緊急にストレージサービスのバックアップに適用させる必要がある。そして災害発生時に、そのような緊急バックアップが同一地域で多発した場合、ストレージサービスのアクセス集中により、バックアップが間に合わない可能性がある。その対策として、災害情報や転送が必要な差分データ量等の情報を基に、コストを考慮して緊急バックアップ先を推薦することでアクセス集中を避ける、バックアップ先スケジュールサービスについて検討する。

キーワード    災害、ストレージサービス、バックアップ、スケジューリング

## 1 はじめに

世界中で台風や地震などの災害による人々の生活への影響が多く見られる。災害で及ぼされる影響は人々の生活に直接的なものに留まらず、蓄えられたデータにも及ぶ。データが蓄えられたサーバーもまた被害を受け破壊される。実際に日本においても、2011年3月11日の東日本大震災を筆頭に多くの災害が被害をもたらしている [1]。そのような被害を受けて、データの損失を防ぐための対策というものも進められており、日常的なものとして、ストレージサービスを利用して遠隔地に定期的なバックアップを行う事でデータを保護する方法や、災害発生時の、限られた時間でのデータ確保を目的としたバックアップの研究などといった対策が進められている。

ストレージサービスを利用する場合の推奨的なバックアップの指針も [2] 等で公開されている。また、研究の例としては T.Nakamura らが迅速で確実なデータ復旧を行う技術の提案をしており [3]、この研究では早い時間でのデータ復旧を可能とし高いリカバリー率を得ることに成功している。

さらに、ストレージサービスは利用しないネットワークを使用した、バックアップによるデータ損失を防ぐための対策として高平らがデータ損失が発生する前に緊急的なバックアップを行う事でデータ損失を防ぐネットワーク制御手法を提案している [4]。

しかし、そのような対策によりデータ損失が阻止されてきてはいるものの未だ全てにおいて対策を取れているということではない。例えば、ストレージサービスを利用して12時間毎に遠隔地に定期バックアップを取っている場合において、そのバックアップの間に作成された新たなデータはオンプレミスにのみ存在するデータということになる。そのデータがストレージサービスにバックアップを取られる前に災害が起きた場合、サーバーの損壊により、データが失われる恐れがある。このようなオンプレミスに存在する最新のデータの損失を防がなければ、全てのデータを守ることが出来ない。

ば、全てのデータを守ることが出来ない。

現在、災害の多くは発生するタイミングで警報等が出されるようになっており [5]、災害を認知してから実際に被害が発生するまでに時間の猶予が存在する。その為、その猶予の間にバックアップを取ることで、オンプレミスに残る最新のデータのバックアップも取ることが出来ると考えられる。

このような対応をする場合想定される状況の例として、大地震が発生するという警報が発令されたとした時、発生地域の企業や個人がデータのバックアップを急遽取ろうとする。地震による直接の被害に限らず、津波の被害なども考えられることから多くの場所でバックアップを求めようとなると考えられる。その場合、バックアップデータの転送が特定箇所に集中してしまい通信に障害が発生し、バックアップが上手くいかない可能性がある。

そこで、そのような災害により時間制約が生じた状況における、特定箇所へのデータ転送の集中を避けるバックアップについて考える。データの損失を防ぐため、バックアップデータを分散させることも検討したバックアップ配置を考え、災害が発生した際にそのデータ配置を用いることで、データ損失を防ぐことが出来る方法を提案する。

データ損失を防ぐ手法として、時間制約とバンド幅、そしてバックアップに必要なデータ量から、時間内にバックアップを取るような制約条件を式として設定し、ソルバーを用いて、条件を満たしバックアップの際のコストを抑える最適解を求める。

コストとして実際のストレージサービスを参考に、ストレージサービスを利用する際にかかる費用を設定し、その費用について4つの手法で比較実験を行った。結果として損失率、コストの両面において大幅に抑えることが出来た。

本論文では、第2節において、関連研究を紹介する。第3節では、提案手法を提示し、第4節で、実験の手順、内容、結果とその考察をする。最後に、第5節でまとめと今後の課題を述べる。

## 2 関連研究

### 2.1 災害時のデータ復旧に関する研究

2011年に発生した東日本大震災では、多くの被害がもたらされた。その際には、ネットワークが接続できなくなり、情報提供システムも機能を止め、安否の確認や負傷者の管理もできなくなった。T.Nakamuraらは、Japan National Projectに参加し迅速で確実なデータ復旧が行える技術の提案をした[3]。

復旧の方法として、2つの状況を想定し、提案された。

- (1) ネットワークが使用できる状況でデータ復旧する方法
- (2) ネットワークを使わずに復旧する方法

後者は、東日本大震災において広範囲でネットワークの復旧までに1か月がかかったために想定された。前者では、近隣の地域のデータと併せて遠隔に置いたバックアップから復元する。

後者では、近くのサーバーにバックアップデータを置き、サーバーが失われた場所にサーバーごと運ぶことでデータの復旧を図る。そのため、バックアップデータが遠すぎず、プライマリーデータと同時に失われないような安全な位置に置くことを考える。ハザードマップやデータ量を参考に、各地域において被災確率と被災した際の損失の大きさから、その地点にバックアップを置く危険度を示すリスク関数を用意する。このリスク関数を用いることで、安全なデータ配置ができるようにしている。

この研究では、1つの災害に対して早い時間での復旧を可能にし、高いリカバリー率を得る。その一方で、ハザードマップが対応することができる状況のみを想定しており、ハザードマップに存在しない災害が起きた場合には対応できない。例として、千葉県千葉市や北海道札幌市などでは、地震の際の建物崩壊の危険性を纏めたハザードマップが存在することに対して、千葉県長生郡長生村や北海道登別市などでは、存在しない。

ハザードマップは自治体毎に作られている種類や数が異なり、この研究の手法を用いることが出来ない状況があるという欠点もある。また、同時に複数の災害が起きた場合を想定すると、リスク関数の値の設定が難しいという問題点も挙げられる。

### 2.2 災害発生直後におけるネットワーク制御手法

一方で、ネットワークを使用したバックアップ手法の研究として高平らは災害発生時において、データ損失が発生する前に緊急的なバックアップを行う事でデータ損失を防ぐネットワーク制御手法の提案がされている[4]。

この研究では、緊急時におけるデータ損失の対策として災害直近隣の安全なサーバーへ重要データを伝送する緊急バックアップが効果的であるとしており、災害速報などから決定したリンクの制限時間から通信経路の評価を行い、制限時間内に伝送が必要な残りのデータ量が多い拠点から優先的に帯域の大きい経路を割り当てることで大規模災害時に効率の良い緊急バックアップを目指している。

通信経路の制御は、緊急地震速報や津波被害推定システムなどの災害速報サービスから得られる拠点や通信経路が破損するまでの制限時間と、各拠点が伝送するバックアップデータ量に

基づいて行い、バックアップデータ量の多い拠点に大きい帯域の経路を割り当て、バックアップデータ量の少ない拠点に小さい帯域の経路を割り当てることで緊急バックアップにおける効率の良いデータ転送を実現する。

この研究では、拠点のデータ量と通信経路の破損を考慮して、この2つの内の片方を考慮していない手法2種と比較を行い評価を行い、結果として様々な環境・災害においても効率の良いデータ転送を実現できることを示している。

### 2.3 ストレージサービスを利用したバックアップ

#### 2.3.1 ストレージサービスを利用したバックアップの例

ストレージサービスの例としてAmazon S3やS3 Object Storage, Azure Storage [6], NEC Cloud IaaS [7]などがあり、これらを活用することにより遠隔地にバックアップを置きデータを保護する災害対策として利用することが出来る。

#### 2.3.2 ストレージサービスを利用したバックアップの一例

ストレージサービスを利用したバックアップの一例として[2]が推奨しているものがある。[2]が推奨するデータ運用では、NetBackup [8]を利用し、バックアップ体系として、

- (1) 業務システム側で日常的にバックアップを頻繁に行う。
  - (2) 日に1度などのペースで1でバックアップ(定期バックアップ)を取ったデータを用いて、遠隔バックアップを行う。
- の2つの段階に分けたバックアップ体系を推奨している。

(1)のバックアップにより、1つの業務用サーバーの故障などといった日常での不意なデータ損失を防ぐことに繋がり、(2)のバックアップにより、建物自体が被害に遭うような災害等の発生時のデータ損失を防ぐことに繋がると考えられる。

このようなストレージサービスを利用したバックアップ体系は確かに便利でデータ損失を防ぐことに貢献する。しかし、まだまだデメリットと言えるものは存在しており、

- クラウド側のシステム性能・負荷状況にデータ復旧時間が左右される場合があること。
  - 災害が発生した際には、遠隔バックアップのインターバル間に生じるオンプレミス側のデータの転送は出来ない。
- といったような問題がいまだ残されている。

### 2.4 課題点

関連研究における、災害発生により生じるバックアップを取るまでの時間制約やデータ量に考慮した複数の拠点を利用したバックアップ体制は、データ損失を防ぐことに有効であるが、複数のサーバーを用意しネットワーク環境を整える必要があり、その環境構築において初期費用が高いと考えられる。また、ストレージサービスを利用したバックアップでは、定期バックアップの間にオンプレミス側に生まれる差分データがデータ転送が行えず災害などでデータ損失する可能性があったり、データ転送が短時間に集中して行われてしまうと集中箇所の帯域が圧迫されデータ損失の可能性があるとといった課題点が残っている。

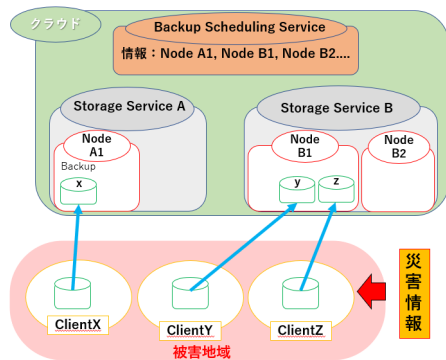


図 1 Backup Scheduling Service を利用したバックアップ (1/3)

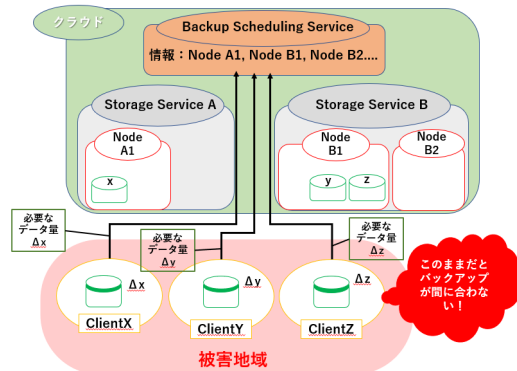


図 2 Backup Scheduling Service を利用したバックアップ (2/3)

### 3 提案手法

#### 3.1 概要

前節で挙げた課題点を解決する為に、コストが安価であるオンラインストレージサービスを利用した最新データのバックアップによりデータ損失を防ぐ手法を提案する。複数のストレージサービスを利用してデータ転送の集中していないストレージサービスを選択してデータを転送できる仕組みを考える。

上記のバックアップ配置の手法を実現するため、Backup Scheduling Service を利用したバックアップ配置を提案する。

Backup Scheduling Service で行うことは、簡単に以下の 4 つに分けられる。

(1) 災害情報 (時間制約に関する情報) や、通常時のバックアップ先の情報、各ストレージサービスにおけるバンド幅などの情報を取得

(2) バックアップを取る際に必要なバンド幅の計算

(3) 特定のストレージサービスへのバックアップの集中を避けた、時間制約内に行えるようなバックアップ配置の決定

(4) バックアップ先の推薦

これらの動作について例となる図 (図 1,2,3) を用いて説明する。

##### 3.1.1 通常時のバックアップ

Backup Scheduling Service は、初めに情報受取を行う (図 1)。図 1 で、下に並ぶクライアント (X, Y, Z) はサービスを受けることになる顧客 (企業, 個人, 自治体等) を指し、このクライアントが被災してそれぞれのオンプレミスに残されている最新のデータが失われる危機にあるという想定になっている。

水色の矢印は平常時にクライアントがバックアップを置いているストレージサービスを示しており、図 1 の例ではクライアント X にあるデータは NodeA1 に、クライアント Y, クライアント Z は NodeB1 にバックアップが取られている。

クライアント側で災害を検知した場合、それぞれのクライアントが「オンプレミスに残された最新のデータのバックアップを被害を受けるまでの限られた時間のうちにとりたい」となる。

##### 3.1.2 バックアップの指示の要請

それぞれのクライアントは、図 2 のように Backup Scheduling Service に対してバックアップしたいデータ量を報告する。

それぞれのクライアントが持つバックアップの必要なデータ (図 2 の例では  $\Delta x, y, z$ ) について、図 2 右側のクライアント Y, Z のデータのように同じ場所にデータを取っているクライアントが複数集まった場合はデータの転送が集中してしまい時間がかかってしまう恐れがある。Backup Scheduling Service がスケジューリングすることでそのような集中を避け、素早くデータ転送が行えるようにする。

##### 3.1.3 バックアップデータの配置

Backup Scheduling Service において、受け取った情報からデータの転送の集中を避けスムーズにバックアップを取るためバックアップ配置についてのスケジューリングを行う。スケジューリングの方法は 3.3, 3.4 において、説明する。

スケジューリングの後、図 3 のようにクライアント側では Backup Scheduling Service から指示を貰い、バックアップを取ることでデータ転送の集中を避けた時間制約内でのバックアップを可能にする。図 3 では青の矢印で示しているように、クライアント Z のデータを通常時のバックアップがある Node B1 ではなく別の場所である Node B2 に臨時バックアップデータとして置くことで、データ転送の集中を避けた時間制約内でのバックアップを可能になる。

また、バックアップを取る際には差分の適応には時間がかかると考えられることから、差分の適応をせずにデータを置くこととする。Delay と書いてある部分は後に行われる処理であり、差分データを元々取ってあるバックアップに合わせる動作を示している。

そして、本論文ではこの仕組みによるデータ転送の集中の回避をより確実にするために各バックアップデータを分割して複数のノードに分けておくことも検討する。

#### 3.2 データの分割

バックアップを取る際に、課題となる内の 1 つとしてデータ量の多いバックアップに対する対処がある。データ量の多いバックアップを取る際に、時間がかかり被災までの制限時間を超えてしまう恐れがあるからである。

そこで、本実験ではバックアップデータを分割して複数のストレージサービスに分けて置くことでこの問題の解決を図る。

バックアップデータの分割は、図 4 のように考える。初めに、

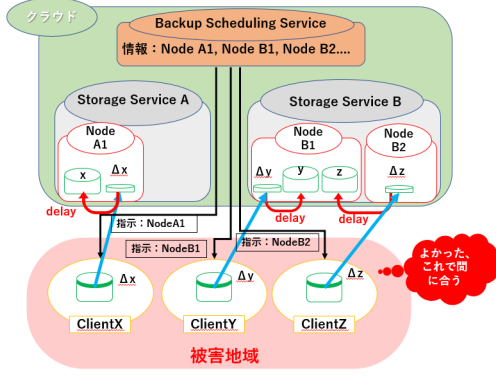


図3 Backup Scheduling Service を利用したバックアップ (3/3)

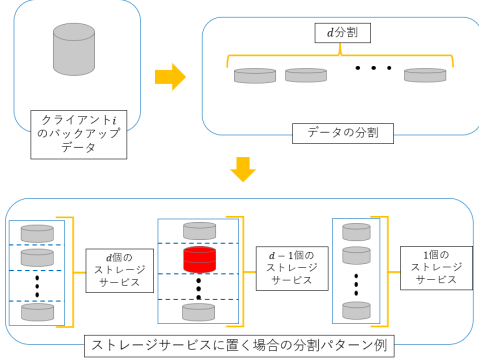


図4 バックアップデータ分割の図

それぞれのデータを  $d$  分割して、その細かくしたデータを配置するという考え方となっている。配置される先（ストレージサービス）が重複する際にはそのデータを統合して1つのデータとみなして考えるようにする。

$d$  を大きくするほど、バックアップを必要とするクライアントの中に大きなデータのバックアップを求めるクライアントがいる場合に対処がしやすくなる。一方で、分割をしすぎるデメリットとしてバックアップ配置を求める計算をする際に時間がかかるようになったり、データを多数に分割した場合データの復元にかかるコストも発生する。

### 3.3 バックアップデータの配置

実際にデータ損失を防ぐことの出来るバックアップデータの配置を考えていく。バックアップするデータ量とストレージサービスの許容するバンド幅、バックアップを取るまでの制限時間（災害により被害を受けるまでの時間）を考慮してバックアップデータを置くために次のように考える。

- クライアント数を  $m$  とする
- ストレージサービス数を  $n$  とする
- 1つのデータに対する最大分割数を  $d$  とする
- クライアント  $i$  のバックアップデータ  $j$  がストレージサービス  $k$  に存在するかを表す関数を  $x_{i,j,k}$  ( $= 0, 1$ ) とする
- クライアント  $i$  の持つバックアップデータ量を  $D_i$  とする
- クライアント  $i$  の持つバックアップデータ  $j$  をストレージサービス  $k$  に置く場合のデータ量を  $D_{i,j,k}$  とする（ただし、 $D_{i,j,k}$  は  $D_i$  の  $\frac{1}{d}$  のデータ量である）

- ストレージサービス  $k$  が許容するバンド幅を  $B_k$  とする
- 制限時間：被災までの時間を  $T_{disaster}$  とする
- ストレージサービス側に着目して制限時間内にバックアップを置かなくてはならないことから、
$$\sum_{i=1}^m \sum_{j=1}^{s_i} \frac{x_{i,j,k} D_{i,j,k}}{B_k} = T_{disaster} (k = 1, \dots, n)$$

- 同様にクライアント側においても制限時間内にバックアップを取らなくてはならないことから、
$$\sum_{k=1}^n \sum_{i=1}^m \frac{x_{i,j,k} D_{i,j,k}}{B_k} = T_{disaster} (i = 1, \dots, m)$$

- それぞれのデータについて幾つに分割してバックアップデータを取るかを定める（分割したデータの総和は常に  $D_i$ ）。
$$\sum_{j=1}^{s_i} \sum_{k=1}^n x_{i,j,k} D_{i,j,k} = D_i (i = 1, \dots, m)$$

- データ分割後のデータ数  $\sum_{i=1}^m \sum_{j=1}^{s_i} \sum_{k=1}^n x_{i,j,k}$  が最小となるバックアップの配置を考える

データ分割後のデータ数を最小にする目的関数としているのは、分割を増やした場合に複数のストレージサービス、ノードを利用することになりバックアップを用いたデータ復元の際の処理が必要になり、そのためのコストがかかるためである。

以上を纏めると、目的関数と制約条件は以下になる。

目的関数：

$$\text{Minimize Partition} = \sum_{i=1}^m \sum_{j=1}^{s_i} \sum_{k=1}^n x_{i,j,k}$$

制約条件：

$$x_{i,j,k} = \{0, 1\} (i = 1, \dots, m, j = 1, \dots, s_i, k = 1, \dots, n)$$

$$\sum_{i=1}^m \sum_{j=1}^{s_i} \frac{x_{i,j,k} D_{i,j,k}}{B_k} \leq T_{disaster} (k = 1, \dots, n)$$

$$\sum_{k=1}^n \sum_{i=1}^m \frac{x_{i,j,k} D_{i,j,k}}{B_k} \leq T_{disaster} (i = 1, \dots, m)$$

$$\sum_{j=1}^{s_i} \sum_{k=1}^n x_{i,j,k} D_{i,j,k} = D_i (i = 1, \dots, m)$$

### 3.4 利用料金を考慮したバックアップデータの配置

前節で提示した式においては、目的関数を分割数を最小にするように設定していたが目的関数の係数として、利用料金を設定し利用料金を抑えることを最優先とした式を立て、ソルバーを利用してバックアップ先を決める。

実際、各ストレージサービスではそれぞれ使用に費用がかかるため、本実験においても各ストレージサービスがそれぞれ使用に応じて費用がかかると考え利用料金を設定する。ストレージサービスの使用で発生する料金として2つのパターンを想定し、使用する際に月額などの定額料金が発生する場合と使用するストレージサービスの容量に応じて費用が増える従量課金制の使用料を考えている。

ストレージサービスでは量に応じて使用料が変化する仕様が一般的であるが、本実験では定時バックアップ後の差分を対象とするためデータ量が大きくないと想定し単純化して考える。

以上より、利用料金を考慮した式を立てる。

クライアント  $i$  の持つバックアップデータ  $j$  をストレージサービス  $k$  に置く場合にかかる利用料金を定める係数を  $c_{i,j,k}$  とし、 $c_{i,j,k} = (\text{ストレージサービス } k \text{ の契約料}) + (\text{ストレージサービス } k \text{ の使用料}) \times D_{i,j,k}$  とする。

制約条件は前節の式を利用し、目的関数に以下の式を立てることにより制限時間内で損失のないバックアップ配置の中で最も利用料金を少なくする配置を求める。

目的関数：

$$\text{Minimize Cost} = \sum_{i=1}^M \sum_{j=1}^{S_i} \sum_{k=1}^{n} c_{i,j,k} x_{i,j,k}$$

以上のように最適化問題の解法の中で比較的早く解の出せる線形計画ソルバーを利用した2つの手法を提案し、他2つの比較手法と併せて実験を行う。

## 4 評価実験

提案手法に加え、通常時のバックアップ先から変更しない手法やランダムにバックアップを配置する手法を考え、バックアップを確実に取ることが出来るか、コストは抑えることが出来るかを比較する。それぞれ、実際に被害に遭いデータが失われることを想定し、手法について比較し評価をした。

なお、3.2 で示したデータの分割数において本実験ではデータを定時バックアップ後の差分と考え、緊急時のバックアップのデータ量としては大きくないと想定し、1つのデータにつき最大3分割までとする。

### 4.1 評価目的

本論文では、損失率の比較と2つのパターンにおけるコストについて手法の比較を行い、その評価・考察を行う。

#### 4.1.1 比較対象

比較する手法として比較手法1、比較手法2、提案手法1、提案手法2の4つを用意した。

比較手法1では、全てのクライアントにおいて、通常時のバックアップ先にそのままバックアップを置く。

比較手法2では、空いてるストレージサービスの中からランダムなストレージサービスを選択し、バックアップを置く。また、この手法に限り数値の変化が大きいため全ての抜出数値は「30回行ったものの平均」を取ったものとする。

提案手法1では、3.3 で提示した式を用いて分割数を押さえることを最優先とし、ソルバーで求めた最適解に基づいてバックアップ先を決める。

提案手法2では、3.4 で提示した式を用いて、利用料金を抑えることを最優先とした式を立て、ソルバーで求めた最適解に基づいてバックアップ先を決める。

#### 4.1.2 使用するデータ

実験に使用するデータの作成は表1のようになっている。

バックアップデータは実際のトレースデータファイルの資料 [9] を基に1日で作成されるデータ量を推定して値を設定し、バンド幅は一般的なネット回線(広く普及している光ネットワーク)を利用した際の数値を参考に設定している [10]。

クライアント数、ストレージサービス数は今回の実験に当

表1 データの作成方法

|                                       |                               |
|---------------------------------------|-------------------------------|
| バックアップデータ量 (GB)                       | ランダムに 5~30                    |
| バンド幅 (GB/Sec)                         | ランダムに 0.3~1.0                 |
| クライアント数 (C*1)                         | 10, 15, 20                    |
| ストレージサービス数 (SS*1)                     | 5, 10                         |
| ストレージサービス使用料<br>(月額料金 (円/月)・使用料金 (円)) | 既存のサービスの料金を参考に<br>した料金プランを利用。 |

表2 料金プラン

|                | 料金プラン1                   | 料金プラン2                   | 料金プラン3                            | 料金プラン4                                    | 料金プラン5  | 料金プラン6       |
|----------------|--------------------------|--------------------------|-----------------------------------|---|---------|--------------|
| 月額料金<br>(円/月)  | 0                        | 0                        | 0                                 | 0   | 1,250   | 1,360        |
| 使用料金<br>(円/GB) | 6                        | 1.84                     | 1.25                              | 0.25                                      | 0       | 0            |
| 参考             | Azure<br>File<br>Storage | Azure<br>Blob<br>Storage | Amazon<br>S3(US)<br>標準低頻度<br>アクセス | Google<br>Cloud(US)<br>Archive<br>Storage | Dropbox | Google Drive |

たって表1のような数値に設定した。

ストレージサービス使用料は、実際のストレージサービスの料金を参考に各ストレージサービスが表2のいずれかの料金プランを持つようになっている [11-15]。

これらのデータを使用して、評価実験を行う。

#### 4.1.3 比較方法

##### ● 損失率の比較

制限時間内に、それぞれの手法についてどれだけバックアップが取れているかを測る。

損失率 =  $\frac{\text{バックアップの取れないデータ量}}{\text{バックアップが必要なデータ量}}$  として、手法毎の損失率の差について比較する。

##### ● ストレージサービス数が同一のデータにおける料金コストの比較

各手法についてストレージサービス数が同一のデータに対してバックアップを取った際にかかる料金コストを比較する。

また、提案手法を用いて損失率が0%になる最小の数値を時間 ( $T_{disaster}$ ) に代入して実験を行う。

##### ● 総バンド幅が同一のデータにおける料金コストの比較

各手法について総バンド幅が同一のデータに対してバックアップを取った際にかかる料金コストを比較する。

これにより、1つ1つのバンド幅が狭いストレージサービスが多くある場合と1つ1つのバンド幅が広いストレージサービスが少しある場合についての違いについて確認する。

また、こちらも提案手法を用いて損失率が0%になる最小の数値を時間 ( $T_{disaster}$ ) に代入して実験を行う。

#### 4.1.4 実験環境

今回の実験では、ソルバーを用いた解法を用いるため lp\_solve [16] を用いた。lp\_solve は、混合整数線形計画法のソルバーで、C 言語で記述されている。実験に使用したシミュレーション環境は表3の通りである。

## 4.2 結果

### 4.2.1 損失率の比較

損失率について比較したものが図5,6,7,8である。

1: 実験に使用するデータを、クライアント数 (C)、ストレージサービス数 (SS)、予想されるバックアップの最短時間 ( $T = \lceil \frac{\text{データ量総和}}{\text{バンド幅}} \rceil + (1or2)$ ) の数値を用いて  $C \cdot SS \cdot T$  と命名している。

表 3 シミュレーション環境

|         |  |
|---------|--|
| OS      | Windows 10 Home                                  |
| CPU     | Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90GHz |
| 使用言語    | Java   |
| 使用ライブラリ | lp_solve [16]                                    |

また、それぞれの手法についてプログラムの実行時間は、表 6 のようになっている。

ソルバーに与える制約条件の式の部分は、提案手法 1 と 2 で同じであるため、損失率としては同じ値となる。

また、提案手法により損失率が 0% になった値は図 5, 6, 7, 8 の順に 41, 28, 76, 49 となっている。これらの数値を以後の料金コストの実験にて  $T_{disaster}$  に代入して利用する。

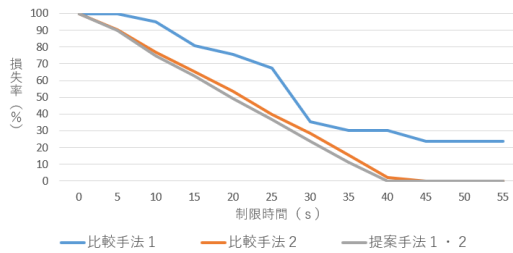


図 5 C10-SS5-T41 を用いた時の損失率の比較

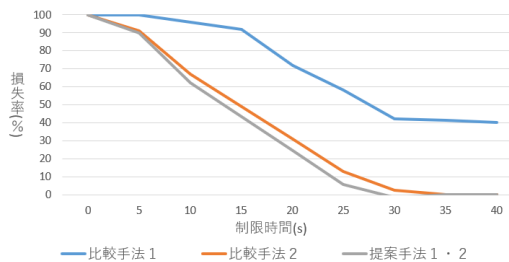


図 6 C10-SS10-T28 を用いた時の損失率の比較

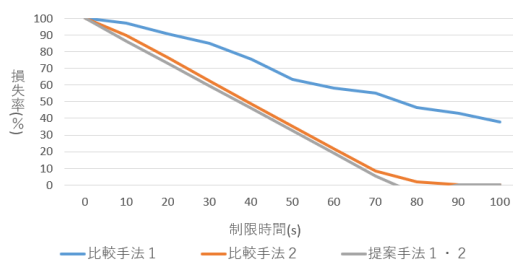


図 7 C20-SS5-T76 を用いた時の損失率の比較

#### 4.2.2 ストレージサービス数が同一のデータにおける料金コストの比較

ストレージサービス数が同一のデータ毎に料金コストの比較したものが図 9, 10 のようになっている。

また、それぞれの手法についてプログラムの実行時間は前節と同じであり、表 6 のようになっている。

また、比較手法 2 の結果として 30 回の平均を採用しグラフに反映させているが、その 30 回における最小値・最大値・95 %

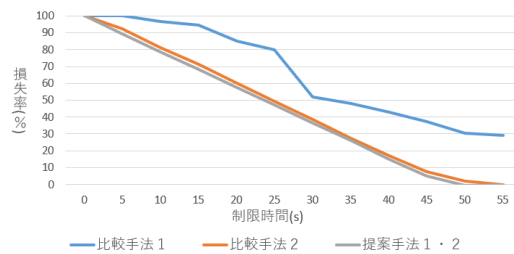


図 8 C20-SS10-T49 を用いた時の損失率の比較

表 4 図 9, 図 10 における比較手法 2 の料金コストのデータ (円)

|              | C10-SS5-T41 | C10-SS10-T28 | C20-SS5-T76 | C20-SS10-T49 |
|--------------|-------------|--------------|-------------|--------------|
| 最小値          | 1,695       | 4,401        | 3,312       | 2,518        |
| 最大値          | 4,217       | 8,394        | 10,232      | 6,663        |
| 95%信頼区間 (下限) | 2,418.976   | 5,260.876    | 4,805.564   | 3,967.638    |
| 95%信頼区間 (上限) | 3,245.024   | 6,130.124    | 5,987.836   | 4,758.029    |
| 平均値          | 2,832       | 5,695.5      | 5,396.7     | 4,362.833    |

信頼区間は表 4 のようになっている。

ただし、損失率のグラフでわかるとおり、比較手法 1 においてはバックアップを取れていないデータが多い。比較手法 1 に限ってはここで出されている費用は、「仮に全てのデータのバックアップが取れていた場合」にかかる費用となっている。他の 3 つの手法については全てのデータのバックアップが取れており、かつその状態の費用である。

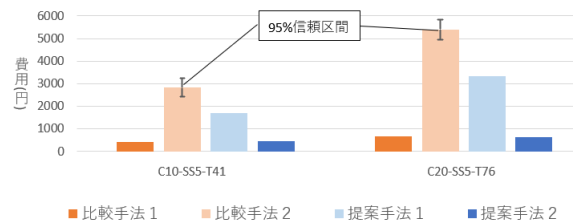


図 9 ストレージサービス数が同一のデータにおける料金コストの比較

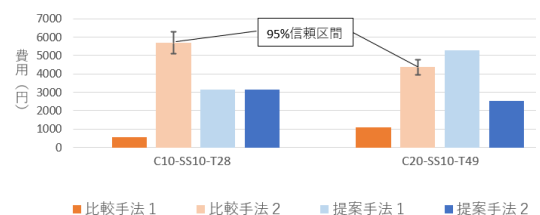


図 10 ストレージサービス数が同一のデータにおける料金コストの比較

#### 4.2.3 総バンド幅が同一のデータにおける料金コストの比較

総バンド幅が同一となっているデータにおいて、料金コストの比較したものが図 11, 12 のようになっている。

また、それぞれの手法についてプログラムの実行時間は、表 6 のようになっている。

また、比較手法 2 の結果として 30 回の平均を採用しグラフに反映させているが、その 30 回における最小値・最大値・95 % 信頼区間は表 5 のようになっている。

こちらも同様に比較手法 1 に限ってはここで出されている費



表 5 図 11, 図 12 における比較手法 2 の料金コストのデータ (円)

|              | C10-SS5-T35 | C10-SS10-T35 | C20-SS5-T73 | C20-SS10-T73 |
|--------------|-------------|--------------|-------------|--------------|
| 最小値          | 5,392       | 5,389        | 2,205       | 8,516        |
| 最大値          | 9,534       | 13,412       | 7,212       | 15,026       |
| 95%信頼区間 (下限) | 6,781.7     | 9,267.958    | 4,352.792   | 11,895.95    |
| 95%信頼区間 (上限) | 7,616.366   | 10,615.44    | 5,059.008   | 13,232.11    |
| 平均値          | 7,199.033   | 9,941.7      | 4,705.9     | 12,564.03    |

表 6 実験の実行時間 (s)

|              | 比較手法 1 | 比較手法 2 | 提案手法 1 | 提案手法 2 |
|--------------|--------|--------|--------|--------|
| C10-SS5-T41  | 0.001  | 0.004  | 1.223  | 4.182  |
| C10-SS10-T28 | 0.001  | 0.004  | 3.301  | 8.983  |
| C20-SS5-T76  | 0.001  | 0.005  | 0.324  | 5.503  |
| C20-SS10-T49 | 0.001  | 0.005  | 1.435  | 6.887  |
| C10-SS5-T35  | 0.001  | 0.004  | 1.376  | 4.174  |
| C10-SS10-T35 | 0.001  | 0.004  | 4.103  | 8.002  |
| C20-SS5-T73  | 0.001  | 0.004  | 0.288  | 5.722  |
| C20-SS10-T73 | 0.001  | 0.005  | 1.589  | 4.223  |

用は、「仮に全てのデータのバックアップが取れていた場合」にかかる費用であり、他の 3 つの手法については全てのデータのバックアップが取れており、かつその状態の費用である。

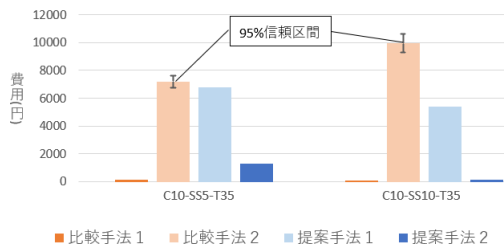


図 11 総バンド幅が同一のデータにおける料金コストの比較

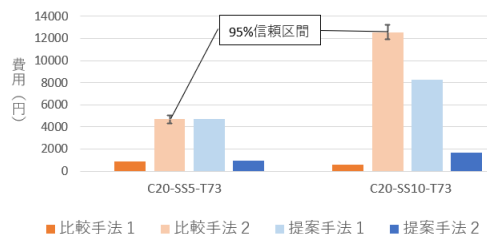


図 12 総バンド幅が同一のデータにおける料金コストの比較

### 4.3 考 察

図 5,6,7,8 のように全データにおいて、提案手法によりデータ損失を抑えられていることがわかる。また、損失率の観点のみならば比較手法 2 のランダムな配置でも成果が確認でき、バックアップの集中を避けることが有効であることがわかる。

図 5 の提案手法でグラフが急激に変化しているのは、データを分割したブロック単位でバックアップ可・不可の判定をしている為、ストレージサービスに空きがあってもバックアップを取れない状況があり時間が延びたタイミングで一気にバックアップを取ることに出来るデータが増えた為だと考えられる。

図 9,10 においては、バックアップが取れている比較手法 2、提案手法 1、提案手法 2 で比べると常に提案手法 2 が最も費用を抑えることが出来ている。これによりソルバーを用いて費用を抑える手法の効果が出ている事が確認できる。

また、C20-SS10-T49 のグラフに限り提案手法よりも比較手法 2 の方が費用を抑えられているが、その理由として、提案手法 1 では比較手法 2 と比べて、分割数を抑えるために高額な料金プラン 6 を持つストレージサービスに複数のクライアントから置く必要があったからだと考えられる。

図 11,12 においても、図 9,10 と同様にソルバーを用いて費用を抑える手法の効果が出ている事が確認できる。

また、図 11,12 より総バンド幅を統一した際にストレージサービスのバンド幅の大きさ (ストレージサービス数) を変化させても変わらず成果が出せていると言える。

さらに、費用は制限時間の変更で変化が考えられるが、図 5,6,7,8 の通り、制限時間が短くなると提案手法 2 種においてもデータの損失が生じる。仮にバックアップに成功したデータにだけ費用がかかると考えるとした際には、4 種の手法についてバックアップ配置を変更する要因が考えられないため、純粋に失われた分だけ費用がかからなくなると考えられる。

一方で、制限時間が長くなった場合、提案手法 2 では明らかに費用の安い料金プランを持つストレージサービスにデータが集まること、契約料を必要とする料金プランを持つストレージサービスの新規利用を避けることから比較手法 1 に数値が近づいていくことになる。また、比較実験 1 で求められている費用は最安値ではない (料金プラン 1 のストレージサービスの使用など) から、制限時間がバックアップを取るのに十分な数値を持つとき比較実験 1 の数値を下回ることになる。

そして比較手法 1・比較手法 2・提案手法 1 においてはストレージサービスの費用を考慮しない手法であり費用に大きな変化はないと考えられるが、比較手法 2・提案手法 1 においてはストレージサービス側の許容量が増え余裕が出来ることからデータを分割することが少なくなることがわかる。それに応じて、契約料の存在するストレージサービスを新規利用する確率が僅かに下がると予想でき、費用が安くなると考えられる。

また、それぞれのデータの実行時間が表 6 のようになっている。データ転送にかかる時間にこの結果を加えた結果が実際に対応するのにかかる時間とすると、提案手法 2 を有効的に利用するには、損失率を抑えるために提案手法 1 と比べて時間に余裕を持たせる必要があると考えられる。また、比較手法 2 は損失率が提案手法と近い値を出している中計算時間も短いためデータ数が増えた際にデータの損失を防ぐという目的に絞れば優れた成果が出ると考えられる。

## 5 おわりに

### 5.1 ま と め

本論文では、バックアップのデータ配置を Backup Scheduling Service によりスケジューリングし、災害が発生した際にそのデータ配置を用いることで、データ転送の集中を避け、データ損失を防ぐことが出来る方法について検討した。

各クライアントが取りたいバックアップのデータ量と各ストレージサービスが許容するバンド幅、バックアップを必要とするクライアント数とバックアップの配置先になるストレージ

サービス数を用意した。そして、災害が発生した状況を想定しデータ損失を防ぐために不可欠な制限時間の値を用意した。

Backup Storage Service へそれらの情報を送信し、制限時間内で、分割数や利用料金を抑えるようなバックアップの配置を求めた。そのデータ配置を、各クライアントにバックアップ先に推薦し、データのバックアップをストレージサービスに送ることでデータ損失も防ぐことが出来た。

バックアップ配置を求める際にはソルバーを利用した計算式を用いた分割数を最小にする手法と、利用料金を減らす手法、それに加えて比較用の手法を2つ利用した。これらの手法について、損失率と料金コストについて評価実験を行った。

結果として、損失率を最大52.3%削減でき、提案した手法はデータ損失に対する対策として有効であることがわかった。ランダムにバックアップを配置する手法(比較手法2)においても何もしない場合(比較手法1)と比べて大きく損失率を減らしていることから、データ転送の集中を避ける行為自体がデータ損失の対策として有効であることがわかった。

さらに、料金コストの比較において、分割数を減らす手法は、料金プランの選択によってコスト低下につながらないこともあるため、分割数だけでは不十分であることがわかった。一方で、利用料金を考慮した手法においては料金コストを最大で98.6%削減出来たことから料金コストを抑える上で有効であることを確認できた。

総バンド幅を固定させて行った実験からも、手法の有効性を確かめることができ、データの種類、バンド幅の大きさに問わず損失率を抑えつつ、費用も抑えることができる手法であることがわかった。

また、費用は制限時間を変更することで多少の変化は考えられるが、制限時間が短くなった場合は、失われたデータ分だけ費用がかからなくなると考えられるため変わらず提案手法で費用が抑えられた。制限時間が長くなった場合においても、提案手法が他の手法よりも費用が抑えられることがわかり、どのような状況においても提案手法が有効であることがわかった。

## 5.2 今後の課題

ソルバーを利用し分割数を最小にする手法を使っているが、その有効性について本実験の指標では十分に測ることが出来ない。分割することにより生じるメリットやデメリットを数値化することで有効性を十分に測ることが出来ると考えられる。

ストレージサービスの使用料金は従量課金制がほとんどであるため、厳密な設定をするためには既にストレージサービスに置かれているデータ量の想定も必要であると考ええる。さらに、費用としてバックアップによるデータ復元の際の通信料なども現実には必要となる為考慮できると考える。

また本実験では、バックアップ配置の導出にかかった計算時間等を考慮していない。例えば、本実験で使用したlp\_solveによる計算時間を見ると、提案手法2は提案手法1と比べてそれなりに時間がかかる。一方で比較手法2は計算時間がほとんどないとみなせる。このようなことを考慮してバックアップ配置を導くことで、より実際の状況での使用についての研究に繋が

ると考えられる。

さらに本実験では、時間制約と実際の災害の到達時間との比較ができていないが、最も制約が厳しいと考えられる緊急地震速報から被災までで10-20秒(直下型地震の場合はさらに短くなる)と言われている。これを研究を進めていくうえでバックアップをする時間の参考の一つとしていきたいと考えている。

現実性の向上の為、データセットとして地図を用意しクライアントを地図内に設置することで災害情報の受信から制限時間を求める部分、またクライアント数の設定の部分において実験使用するデータを現実に近い設定が出来ると考えている。

そして、Backup Scheduling Systemの実用的な活用に向けて複数のストレージサービスを利用することによるAPIの差異に対応するシステムの実装も今後必要となる。

また、今後ストレージサービス間の環境の差異による影響の調査や差異を考慮した実験を行っていきたい。

## 文 献

- [1] 災害時における情報通信の在り方に関する調査結果, “[https://www.soumu.go.jp/menu\\_news/s-news/01tsushin02\\_02000036.html](https://www.soumu.go.jp/menu_news/s-news/01tsushin02_02000036.html)”, 2020年12月21日閲覧。
- [2] 災害対策のための遠隔地バックアップ, “<https://jpn.nec.com/backup/netbackup/netbackup-sl-remotebackup.html>”, 2020年12月21日閲覧。
- [3] Takaki Nakamura, Shinya Matsumoto, Hiroaki Muraoka “Discreet Method to Match Safe Site-Pairs in Short Computation Time for Risk-Aware Data Replication”, IEICE TRANSACTIONS on Information and Systems Vol.E98-D No.8 pp.1493-1502
- [4] 高平 寛之, 畑 美純, ギリエルイス, 和泉 諭, 阿部 亭, 菅沼 拓夫 “災害発生直後における被災の影響とデータ転送量を考慮したネットワーク制御手法の提案とその評価”, 情報処理学会論文誌 Vol.60 No.3 738-749, Mar. 2019.
- [5] 警報・避難 - 内閣府防災情報, “[http://www.bousai.go.jp/kaigir/ep/kentokai/bousai\\_specialist2/02/pdf/shiryo2.pdf](http://www.bousai.go.jp/kaigir/ep/kentokai/bousai_specialist2/02/pdf/shiryo2.pdf)”, 2020年12月21日閲覧。
- [6] クラウドストレージを活用したデータの災害対策, “<https://jpn.nec.com/backup/netbackup/netbackup-sl-cc.html>”, 2020年12月21日閲覧。
- [7] NEC Cloud IaaS “[https://jpn.nec.com/cloud/service/platform\\_service/iaas.html](https://jpn.nec.com/cloud/service/platform_service/iaas.html)”, 2020年12月21日閲覧。
- [8] NetBackup “[https://www.networkworld.co.jp/product/veritas/pro\\_info/netbackup](https://www.networkworld.co.jp/product/veritas/pro_info/netbackup)”, 2020年12月21日閲覧。
- [9] SNIA IOTTA Repository, “[http://iotta.snia.org/historical\\_session?tracetype\\_id=3](http://iotta.snia.org/historical_session?tracetype_id=3)”, 2021年1月3日閲覧。
- [10] 2020年最速の光回線は? 主要サービス10社の通信速度を徹底比較, “<https://naruhodo-wifi.com/optical-line-speed-ranking/#i-2>”, 2021年1月21日閲覧。
- [11] 料金計算ツール Microsoft Azure, “<https://azure.microsoft.com/ja-jp/pricing/calculator/?service=storage>”, 2021年1月17日閲覧。
- [12] 料金 Amazon S3, “<https://aws.amazon.com/jp/s3/pricing/?nc=sn&loc=4>”, 2021年1月17日閲覧。
- [13] Cloud Storageの料金, “<https://cloud.google.com/storage/pricing?hl=ja#network-egress>”, 2021年1月17日閲覧。
- [14] Dropbox 全プランを比較, “<https://www.dropbox.com/plans>”, 2021年1月17日閲覧。
- [15] google Workspace お支払いプラン, “<https://workspace.google.com/intl/ja/pricing.html>”, 2021年1月28日閲覧。
- [16] Mixed Integer Linear Programming (MILP) solver, “<https://sourceforge.net/projects/lpsolve>”, 2021年01月10日閲覧。