

## ディザスタリカバリに向けた非同期リモートコピー構成資源算出方式

田口 雄一<sup>†</sup> 市川 直子<sup>†</sup> 山本 政行<sup>††</sup>

企業情報システムにおいては、遠隔地に設置された拠点にデータを複製することで、広域災害にあっても業務を再開可能とするディザスタリカバリの重要性が高まっている。重要度の高いデータにはディザスタリカバリ性能目標である RPO (Recovery Point Objective) をできるだけ短く設定することが求められる。その達成にあたってはリアルタイムでデータを遠隔地に複製する必要があり、ストレージシステムの備える非同期リモートコピー機能を活用したディザスタリカバリシステムが有用である。そこで本研究では、ディザスタリカバリシステム稼働開始前の設計段階で、資源量を過不足無くかつ高精度で見積もることが可能な非同期リモートコピー構成資源算出方式を提案する。提案方式は、システム構成とその処理手順をモデル化し、ジャーナル領域におけるデータ蓄積量を予測する。さらにその時系列推移から各時刻におけるリカバリポイントを計算し、RPO を達成するかどうか評価することで、同構成の妥当性を判定する。シミュレーションによる同方式の有効性検証の結果、RPO を満たしながらも最大 70% のコストを削減できることを確認した。

## Asynchronous Remote Copy System Resource Sizing for Disaster Recovery

YUICHI TAGUCHI,<sup>†</sup> NAOKO ICHIKAWA<sup>†</sup> and MASAYUKI YAMAMOTO<sup>††</sup>

The enterprise companies implement disaster recovery system in order to improve their IT system availability. A disaster recovery requirement is usually defined by RPO, Recovery Point Objective. In order to achieve a short term RPO, a remote copy function provided by storage system realizes continuous data protection. This research proposes an asynchronous remote copy resource design method that calculates adaptive size of system resource without the overs and shorts in advance to launch a disaster recovery. In this method, a configuration and processes of the asynchronous remote copy system is defined by an evaluation model. This model makes it possible to estimate data amount stored on journal volumes and recovery points. The validity of system configuration can be judged by a reference of estimated recovery points. A simulation of this paper verified that this method reduces 70% of system cost while achieving RPO.

### 1. はじめに

#### 1.1 ディザスタリカバリの重要性

今日、情報システムは企業活動にとって不可欠な基盤である。企業情報システムは日々大量のデータを生成し、その蓄積量は増加の一途をたどっている<sup>1)</sup>。昨今では企業が保有する膨大なデータから新たな知識や情報を発見しようとする試みが多くなされているように、データそのものが価値ある資産と認識されるようになってきている<sup>2),3)</sup>。また企業では、ひとたびデータ消失が起これば事業機会を損失するだけでなく、顧客からの信頼や社会的信用を失うといった重大なリスクが認知されている<sup>4)</sup>。このようにデータ保護は企業経営

にとって重要課題のひとつであり、様々な対策が講じられている<sup>5)~8)</sup>。

データ保護には重要度に応じていくつかのレベルがある。重要なデータについては、一般的なストレージの冗長化<sup>9)</sup> やバックアップにとどまらず、多発するテロや火災のような拠点規模の損害、さらには自然災害による広域被災への対策が求められる。こうした大規模災害においてもデータを保護し、業務継続を可能とするために、ディザスタリカバリシステムが有用である。ディザスタリカバリシステムは、距離を隔てた二つ以上の拠点にデータを複製し、冗長化しておくことで、ある拠点で障害発生した状況にあっても、代替システムで継続稼働しようとするものである。

#### 1.2 ディザスタリカバリ目標指標

ディザスタリカバリシステムの構築にあたっては、業務システムの可用性やデータの重要度に応じて適切に目標値を定める必要がある<sup>7)</sup>。この目標として一般

<sup>†</sup> 株式会社日立製作所横浜研究所  
Yokohama Research Laboratory, Hitachi Ltd.  
<sup>††</sup> 株式会社日立 LG データストレージ  
Hitachi-LG Data Storage, Inc.

に用いられる指標が RPO (Recovery Point Objective) と RTO (Recovery Time Objective) である。RPO は障害や被災が発生した時刻と、同時刻から遡ってデータを復旧可能とする時刻との差を目標値とする指標である。また RTO は被災発生から復旧までの時間を目標値とする指標である。

RPO はデータ消失のリスクを一定時間内に抑えるために定義される。一例として、表 1 の #2 に挙げるように、RPO が 60 秒であれば、被災時刻から遡って 60 秒前までに記録されたすべてのデータを復旧可能とすることが求められる。言い換えれば被災時刻から遡って 60 秒以内に記録されたデータの消失を許容する目標設定でもある。

従って、RPO は業務システムやデータの重要度が高ければ高いほど短く設定される。例えば金融業で扱われるデータはその完全性や正確性が何よりも重視されるため、RPO を短く設定することが求められる。

### 1.3 研究の目的

RPO を最小化するために、ほぼリアルタイムにデータを遠隔地に複製する技術として、ストレージシステムによるリモートコピー機能がある<sup>10)</sup>。

ストレージシステムの備えるリモートコピー機能は、ストレージに定義された記憶領域であるボリュームを単位として、書き込まれたデータを順次、対となるボリュームに複製する。リモートコピー機能には、同期方式と非同期方式があるが、同期方式は、遠隔地に配置されたストレージの書き込み完了報告を待たなければならないため、サーバへの応答性能がサイト間の距離によって生ずる転送遅延により低下する可能性がある。そのため、一般に遠隔地との間のディザスタリカバリには非同期方式が用いられる。

一方、非同期方式の場合、被災時にデータの一部を消失する可能性がある。データ消失の可能性を低減するためには、十分な量のシステム資源を用意すれば良いが、コスト増の要因になる。したがって、ストレージシステムにおける非同期リモートコピー方式の実現においては、データ消失リスクの低減と資源最適化の両立が課題となる。この課題を解決するため、本論文

表 1 RPO (Recovery Point Objective) 設定の例  
Table 1 Example of Recovery Point Objective

#	RPO	意味
1	0 sec	障害発生時刻までに記録されたすべてのデータを復旧可能とする目標設定
2	60 sec	障害発生時刻から遡って 60 秒前までに記録されたデータを復旧可能とする目標設定
3	1 week	障害発生時刻から遡って 1 週間前までに記録されたデータを復旧可能とする目標設定

では、ディザスタリカバリシステムの稼働開始前の設計段階で、資源量を過不足無くかつ高精度で見積もることが可能な非同期リモートコピー構成資源算出方式を提案し、その有効性を検証する。

## 2. リモートコピーシステムの概要と課題

### 2.1 リモートコピーシステムの概略構成

図 1 にリモートコピーシステムの概略構成を示す。リモートコピーシステムは、業務システムが稼働する正サイトと代替システムが稼働する副サイトで構成される。正サイト、副サイトは広域災害にあっても少なくともいずれか一方が存続するように、一定以上の距離を隔てたロケーションに設置される。

正サイトに設置されたデータセンタで業務システムが稼働する。業務システムはデータを生成、参照するアプリケーションを実行するサーバ群と、それらのデータを格納、保管するストレージで構成する。業務システムで利用されるストレージにはいくつかの種類があるが、ここではストレージエリアネットワーク (SAN) を介して複数のサーバから共有されるブロックストレージを例として取り上げる。すなわち、複数サーバで共有されたストレージに、多数のアプリケーションで生成されたデータが保管される構成となる。

### 2.2 非同期リモートコピー方式

非同期リモートコピー方式では、リモートコピーシステムの正サイトでサーバからストレージへの書き込みが発生すると、同ストレージでの書き込み処理完了後、即座にサーバへ完了通知を返し、書き込まれたデータを一時記憶領域であるジャーナル領域に格納する。ジャーナル領域に記録されたデータは FIFO 処理に従い、書き込みの古い順に副サイトストレージへ転

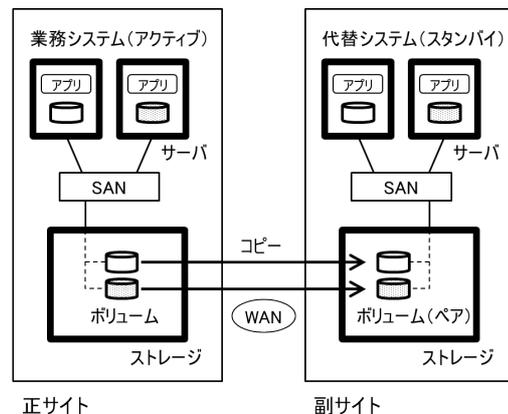


図 1 リモートコピーシステム構成  
Fig. 1 Remote Copy System Architecture

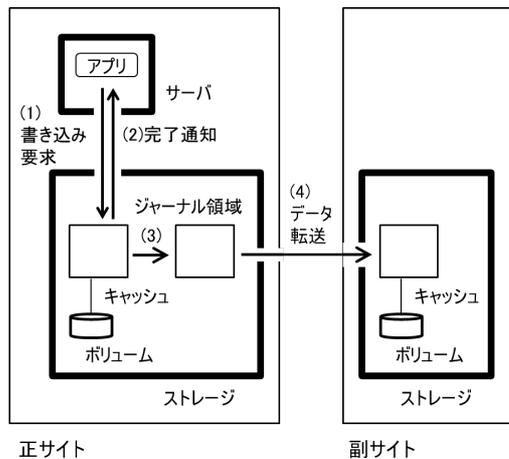


図 2 非同期リモートコピー方式  
Fig. 2 Asynchronous Remote Copy Mode

送される。このためサーバ側で記録完了を確認したステータスであっても、副サイトへの転送を終わっていないデータがジャーナル領域に残存する可能性がある。この状態で正サイトが被災した場合、未転送データは消失する。したがって、非同期方式では、目標や要件に応じてそのリスクを適切に制御する必要がある。次節ではリスク制御に向けた課題を述べる。

### 2.3 非同期リモートコピー構成設計の課題

非同期リモートコピーシステムの構成設計にあたっては、RPO を達成するために、正サイトストレージへの書き込みデータを RPO で定義された目標時間内に副サイトへ転送するだけの性能を担保する構成を設計することが求められる。性能を担保するには、通信回線やバッファ容量などの IT 資源を十分に準備する必要があるが、一般的にコスト増の要因にもなり得る。

そこで、リモートコピーシステムを必要最小限の IT 資源で実現する必要があったが、従来は、設計後に運用中の状況に応じて適宜構成を見直す手法を採用していたため<sup>11)~13)</sup>、初期構成時に、十分な IT 資源を準備するためのコストが必須であり、ディザスタリカバリシステムのコストを最適化することが困難であった。したがって、運用開始前に、RPO を満たす最適な資源量を算出することが課題となる。

### 2.4 解決方針

上述の課題を解決するために、以下の 3 つの方針をとる。

(1) シミュレーションのための評価モデルの策定  
運用開始前の算出を可能にするため、既設システムの書き込み量を入力としたシミュレーションに基づき、リモートコピー導入後のリカバリポイントを予測する

アプローチを採る。そのために実際の挙動を模擬するリモートコピーシステム評価モデルを開発する。

### (2) リカバリポイントの予測

RPO を満たすかどうかを判定するため、リカバリポイントを予測する。リカバリポイントとは副サイトにおける、現在時刻とその復旧可能時刻との差に相当する。

その実現にあたり、ある想定のリモートコピーシステム構成を対象に、各時刻における未転送データ量を計算する。未転送データ量はジャーナル領域のデータ蓄積量に一致し、書き込みデータ量と転送性能のギャップにより計算できる。このジャーナル蓄積データ量からリカバリポイントを予測する。

この処理では、システムへの書き込みデータ量を入力としてリカバリポイントを算出する点に特徴がある。一般に書き込みデータ量は監視しやすいパラメータであるだけでなく、すべての転送データに書き込み時刻のタイムスタンプを記録するといった独自実装が不要であるため、より汎用的なシステム構成や機器に適用可能である。

### (3) 最適資源量の算出

リカバリポイントの予測を用い、同構成が RPO を達成するかどうか判定する。次に RPO を達成する様々な構成を対象に、システム資源の単価設定を用いたコスト計算を行い、同コストが最小となる構成を導出する。

上述の解決方針に基づき、運用開始前にリモートコピーシステムの最適な資源量を算出する。次節では、非同期リモートコピー構成資源算出方式を説明する。

## 3. 非同期リモートコピー構成資源算出方式

### 3.1 非同期リモートコピー構成算出モデル

本研究では、回線帯域やジャーナル領域容量といった IT 資源の設計が適正であるかどうか評価するための非同期リモートコピー構成算出モデルを開発した。同モデルは正副ストレージ間の非同期リモートコピー構成とそのコピー処理過程を抽象化して表現したものであり、ある書き込み負荷が生じた場合の挙動を計算によって予測するために用いる。特に正サイトストレージに対してある書き込みを発生させた場合のリカバリポイントを算出し、同構成が RPO を達成するかどうかを評価する。この評価を様々な設計された構成で再帰的に繰り返すことで、RPO を満たし、かつ資源量が最小となるリモートコピーシステム構成を見つけることが本研究の目論見である。

ストレージならびにリモートコピーシステムの構成

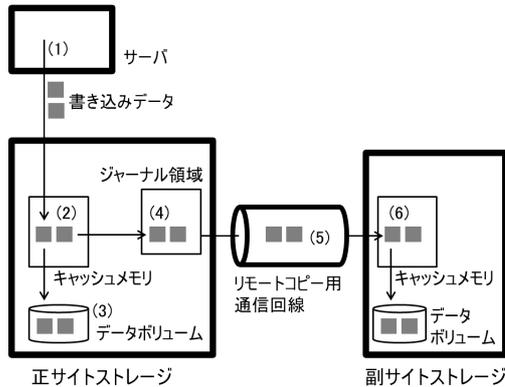


図 3 リモートコピー構成算出モデル

Fig. 3 A configuration sizing model for remote copy system

要素は多々あるが、本研究ではリカバリポイント算出に必要な部位に絞り、抽象化したモデルによってシステム構成を表現する。図 3 に示すように、まず正サイトに設置された正サイトストレージと副サイトに設置された副サイトストレージ、正副サイト間を接続するリモートコピー用通信回線を設ける。さらに各ストレージは書き込まれたデータを一時的に格納するキャッシュメモリと、リモートコピー専用のバッファであるジャーナル領域を有する。前者は揮発性メモリ、後者は揮発性メモリもしくはハードディスクなどの記録媒体による実装を想定する。データボリュームはデータの実体を保存する記録媒体である。

次に同モデルにおける非同期リモートコピーの処理手順を定義する。非同期リモートコピー運用において、正サイトストレージは書き込まれたデータ（処理 (1)）をキャッシュメモリに一時格納し、その時点でサーバに完了通知を返す（処理 (2)）。その後、キャッシュメモリに格納した内容をデータボリュームに記録する（処理 (3)）。さらにこれらの処理とは非同期で副サイトへの転送処理を実行する。転送データはジャーナル領域に時系列で格納された後（処理 (4)）、書き込み時刻の古い順にリモートコピー用通信回線を介して副サイトストレージに送られる（処理 (5)）。副サイトストレージでは受領したデータを、キャッシュメモリを介してデータボリュームに記録する（処理 (6)）。

すなわち同モデルにおいて、同データが正サイトストレージに書き込まれた時刻（処理 (2)）と、転送されたデータが副サイトに記録された時刻（処理 (6)）との差がリカバリポイントに相当する。図 4 に示すように、例えば正サイト内のサーバで生成されたあるデータが 12 時 00 分にストレージに書き込まれ、

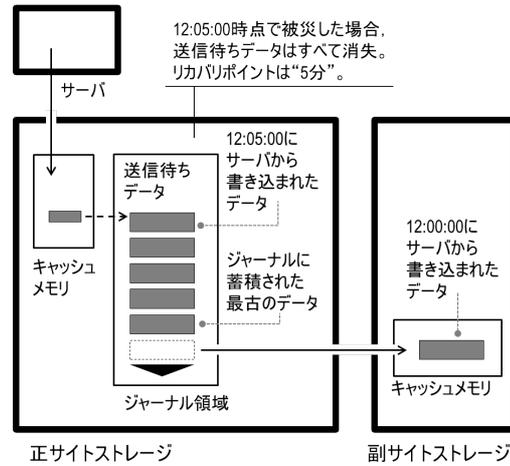


図 4 ジャーナル処理の例

Fig. 4 An example of journal processing

ジャーナル領域での滞留や長距離転送遅延を経て 12 時 05 分に副サイトストレージの記録を完了したケースでは、同 12 時 05 分時点のリカバリポイントは『5 分』である。このとき、仮に同 12 時 05 分に正サイトが被災した場合、12 時 00 分より後に記録されたすべてのデータは未転送であるため消失する。その際、事前に設定された RPO が 5 分より長ければ想定内であるためデータ消失も許容されるが、5 分未満であれば当初の RPO を達成できなかったことになる。

このように評価を行うことで、RPO を達成する構成であるかどうか判定することが可能となる。次節では同モデルを用いたリカバリポイント計算方法を定式化する。

### 3.2 非同期リモートコピー構成資源算出方法

#### 3.2.1 リカバリポイント算出方法

各時刻におけるリカバリポイントは、ジャーナル領域に蓄積されるデータ量から計算することが可能である。まず、ジャーナル蓄積データ量を算出する様子を図 5 に示す。

時刻  $T$  における正サイトストレージへの書き込みデータ量、すなわち流入量を  $In_T$ 、正サイトストレージから副サイトストレージへの転送データ量すなわち正サイトストレージからの流出量を  $Out_T$  と表記すると、ジャーナル蓄積データ量  $C_T$  は以下の式で算出することができる。

$$C_T = C_{T-1} + In_T - Out_T \quad (1)$$

同式はすなわち、前述の想定における書き込みデータ量  $In_T$  に対して、リモートコピー転送性能に依存する流出量  $Out_T$  が不足する場合に蓄積量が増加することを意味している。ここで  $C_{T-1}$  は時刻  $T$  よりひと

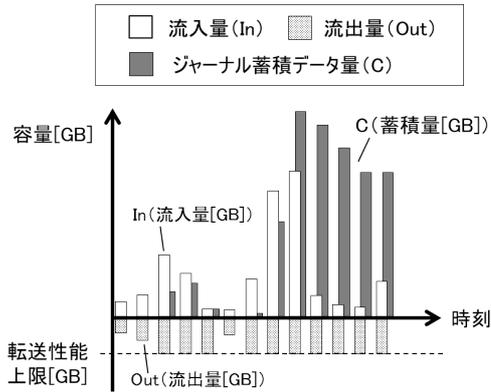


図5 ジャーナル蓄積データ量の時系列推移例  
Fig. 5 An example of journal data amount transition

つ前の時刻におけるジャーナル蓄積データ量を表す。このようにジャーナル蓄積データ量  $C$  もまた時系列で算出される。

流出量  $Out_T$  は書き込みデータ量  $In_T$  に対して転送性能が不足する場合、そのボトルネック箇所が上限となる。前述の算出モデルではハードディスクや揮発性メモリで構成されるジャーナル領域の入出力性能  $P_{JNL}$  と、リモートコピー用通信回線性能  $P_{LINE}$  の二つのパラメータがボトルネック箇所となる可能性がある。すなわち流出量  $Out_T$  は以下のように定式化される。通信回線性能  $P_{LINE}$  は回線帯域で表現すれば良い。

$$Out_T = \min\{In_T + C_{T-1}, P_{JNL}, P_{LINE}\} \quad (2)$$

次に、時刻  $T$  においてジャーナル領域に滞留する未転送データのうち、最も古いデータの書き込み時刻が、副サイトにおける復旧可能時刻に近似することに着目する。ジャーナル蓄積データのうち、最も古いデータ自体は未転送であるため復旧不可能だが、その直前の時刻までに書き込まれたデータは副サイトに転送済みであり、復旧可能であると想定できる。図4の例では12:00:00に書き込まれたデータがジャーナルに蓄積された最も古いデータの直前に書き込まれた、復旧可能データに該当する。このようにリカバリポイントを計算するためには、ジャーナル蓄積データのうち最古のデータを発見し、その書き込み時刻の直前をリカバリポイントと見なせば良い。

ジャーナル領域中、最も古いデータの発見には、ジャーナル蓄積データ量  $C$  と流入量  $In$  を参照する。時刻  $T$  におけるジャーナル蓄積データ量が  $C_T$  であれば、同時点に蓄積されたジャーナルのうち最古のデータが書き込まれた時刻を特定しようとする。これは同時刻から遡って流入量  $In$  を累加した値が  $C_T$  に達し

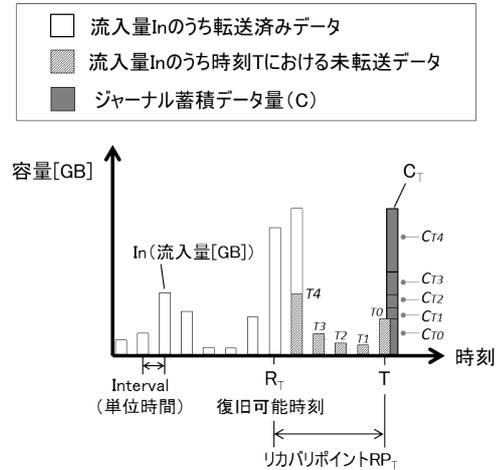


図6 リカバリポイント算出の考え方  
Fig. 6 Recovery point calculation approach

た時刻である。またその直前が復旧可能時刻  $R_T$  となる。すなわち、時系列の書き込みデータ量をヒストグラムで表現した図6における斜線部の累加値が  $C_T$  に達する時刻（図中時刻  $T_4$ ）を特定する。さらにその直前の時刻にあたる  $R_T$  までが副サイトにおける復旧可能時刻であり、そのリカバリポイント  $RP_T$  は時刻  $T$  と  $R_T$  の差に一致する。これらの手続きは以下のように定式化できる。

$$RP_T = T - R_T = \begin{cases} 0 & C_T = 0 \\ (n + 1) \times Interval & C_T > 0 \end{cases} \quad (3)$$

ただし  $n$  は以下を満たす最小の整数である。

$$C_T \leq \sum_{i=0}^n In_{T-i} \quad (4)$$

$RP_T$  は時刻  $T$  におけるリカバリポイントを表す。 $Interval$  は時系列推移の単位時間であり、前述の書き込みデータ量  $In$  のサンプリング間隔に相当する。また  $n$  は  $In$  を時刻  $T$  から遡って累加した値が同時刻におけるジャーナル蓄積データ量  $C_T$  に達した時点までの累加回数であり、そのひとつ前の時刻が復旧可能時刻  $R_T$  に一致する。

以上の手続きに従って算出したリカバリポイント  $RP_T$  の時系列推移を検証することで、各時刻においてRPOを達成しているか、また回線帯域やジャーナル領域の容量といったIT資源量に過不足がないかを評価し、最適なりモートコピー構成を導出できるようになる。

### 3.3 資源量算出方法

本節ではリカバリポイントの計算結果を用いたシステム構成評価と、その評価結果に基づく資源量算出方法を提案する。まずシステム構成評価の対象とする、既設システムの書き込みデータ量が大きい時間帯を抽出し、評価期間とする。この評価期間中すべての時刻において RPO を達成し、かつジャーナル領域容量と回線帯域の資源量のコストが最小となる構成を導出することを旨論む。図 7 にシステム構成評価の例を示す。

図 7 では回線帯域を変数とした 3 種類の構成を対象としたリカバリポイント  $R_T$  とジャーナル蓄積データ量  $C_T$  の計算結果をそれぞれ例示した。

構成 1 は回線帯域不足に起因して RPO を達成できず、データ消失リスク設計が不適切であるケースに該当する。

構成 2 はリカバリポイントの最大値が RPO より十分に低く抑えられていることから、同 RPO を達成しているが、回線帯域が過剰であり、無駄なコストが生じるケースと考えられる。一方で、その分ジャーナルに蓄積されるデータを少量に抑えられることから、ジャーナル領域に揮発性メモリなど高価な機器を利用する構成では、後述の構成 3 よりも総コストを抑えら

れる可能性がある。

構成 3 は評価期間を通じてリカバリポイントが RPO を達成できる事例を表す。通常稼働時は回線の転送性能が十分にあり、ピーク時には要求ジャーナル領域容量を満たすジャーナル領域を設けることで、バッファが有効に作用する。

以上の例に挙げたように、回線帯域を変数とした多様な構成に対して本評価方式を適用し、適正構成を導出することで、リモートコピーシステムを十分な性能かつ低コストで構築することが可能となる。すなわち、以下の式で表される総コストが最小になるシステム構成を採用すれば良い。

$$Cost = \min(Cost_{JNL} + Cost_{LINE}) \quad (5)$$

$Cost_{JNL}$  と  $Cost_{LINE}$  はそれぞれジャーナル領域と通信回線の所有コストおよび運用コストを表す。図 7 で述べた手順で導出した RPO を達成する構成において、それぞれのコストを見積もり、その和が最小となる構成を最適と見なす。

## 4. 評価

### 4.1 実データを用いた実用性検証

ストレージへの書き込みを模擬するデータに Cello99<sup>14)</sup> を用いたリカバリポイント計算結果を図 8 に示す。Cello99 は特定の企業内データセンタにおいて実際に発生したリード・ライトデータ量の一般公開情報であり、本方式の実用性を検証するには十分なデータである。本検証ではこのうち、あるデータボリュームに対する書き込みが局所的に増加した事象に着目し、その前後を含めた時間帯を評価期間とした。図 8 はそのデータボリュームへの書き込みを 15 秒刻みの時系

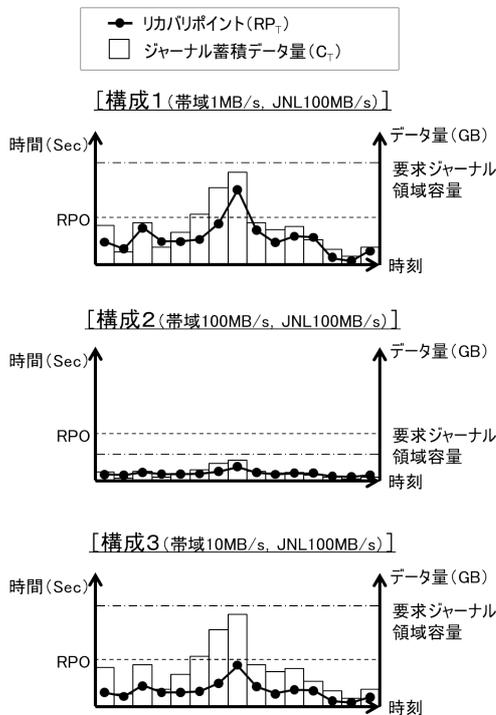


図 7 システム構成評価の例  
Fig. 7 An example of configuration evaluation

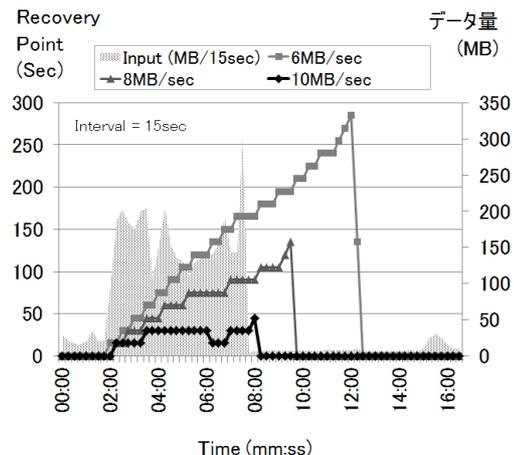


図 8 リカバリポイント時系列推移計算結果  
Fig. 8 A Recovery Point Simulation Result

列で集計し、リモートコピー用通信回線帯域を3通りに変動させたそれぞれの条件におけるリカバリポイント（主軸）を表す。本研究の計算手法を適用し、回線帯域を6MB/sec, 8MB/sec, 10MB/secと仮定した場合のリカバリポイントの最大値はそれぞれ285sec, 135sec, 45secと算出された。この評価により、例えば同ボリュームのRPOが45secであれば、少なくとも10MB/secの回線帯域を、RPOが300secであれば少なくとも6MB/secの回線帯域を必要とする結果が得られた。

さらに図7に示した形で、帯域10MB/sec, 6MB/secそれぞれの構成におけるリカバリポイントとジャーナル蓄積データ量の時系列推移を図9, 図10に示す。図9に示すとおり10MB/secの帯域設計時には少

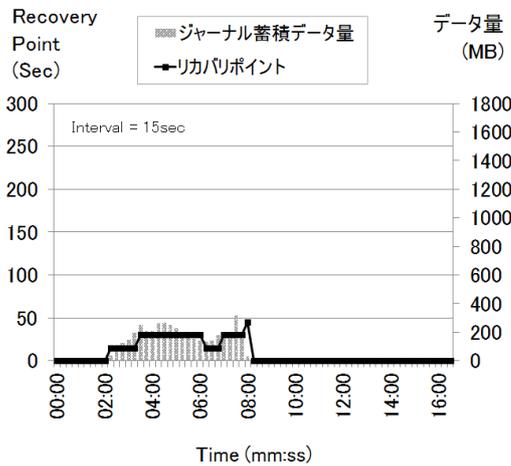


図9 システム構成評価結果 (10MB/sec の場合)  
Fig. 9 A Configuration Evaluation (case of 10MB/sec)

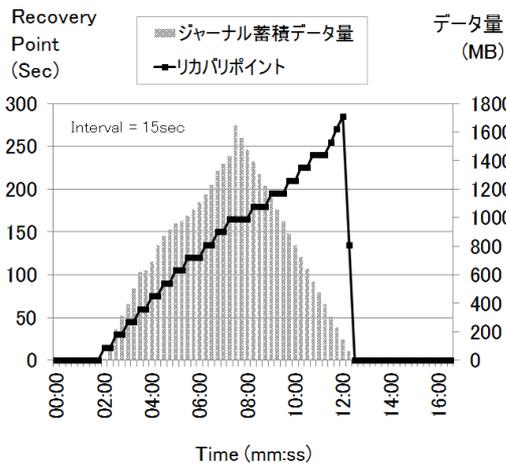


図10 システム構成評価結果 (6MB/sec の場合)  
Fig. 10 A Configuration Evaluation (case of 6MB/sec)

なくとも319MBのジャーナル領域容量を設ける必要がある。同様に図10に示すように6MB/secの帯域設計時には1,645MBのジャーナル領域を設ける必要がある。

回線帯域を変化させたその他の構成において同様の評価を実施した結果を図11に示す。前述の回線帯域10MB/sec, 6MB/secに加え、その他構成における回線帯域の増減を横軸に、それらの構成において本論文の資源量算出方式を適用した結果得られた要求ジャーナル領域容量を縦軸に表現する。さらにRPOを300secとした場合に、同目標値を達成する構成は■で、達成しない構成は×で図中にプロットする。

図11に示した構成のうち、RPOを達成し、資源量のコストが最小となる構成を選択することにより、本研究の目的が達成される。しかし実運用におけるコスト計算は所有コストと運用コストがある他、ジャーナル領域にハードディスクかフラッシュメモリ、あるいは揮発性メモリを用いるかで大きく単価が異なるため一概には定式化できない。そこで、回線帯域1MB/secあたりの単価に対する、ジャーナル領域容量1MBあたりの単価を1/10, 1/100, 1/1000と仮定した場合の、コスト評価結果を図12に示す。

各構成における回線帯域コスト  $Cost_{LINE}$  にジャーナル領域コスト  $Cost_{JNL}$  を加算した値が縦軸のコスト評価値である。ジャーナル領域コスト  $Cost_{JNL}$  は各構成における要求ジャーナル領域容量に前述の単価をかけた値となるため、その単価が高い1/10の場合が最も高額となる。

ジャーナル領域にハードディスクなど安価な記憶装置を用い、その容量あたり単価が回線帯域1MB/secと比べて1/1000となるケースでは、回線帯域が最小となる6MB/secの構成が最も低コストとなることが見込める。逆に揮発性メモリなど高価な装置を用い、容量あたり単価が回線帯域と比べて1/10となるケー

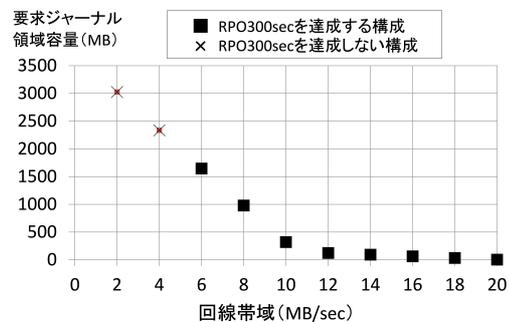


図11 リモートコピー構成資源量算出結果  
Fig. 11 Remote Copy System Resource Sizing Simulation

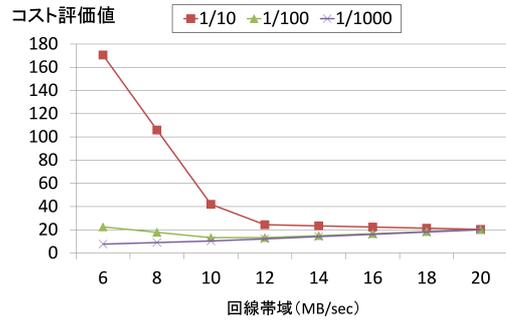


図 12 コスト評価結果  
Fig. 12 A cost simulation result

スでは、回線帯域が最大となる 20MB/sec の構成が最も低コストとなる。これらに対し、容量あたり単価が回線帯域 1MB/sec と比べて 1/100 となるケースでは、図 9 に挙げた帯域 10MB/sec の構成が最小コストと算出された。このように単価設定と要求ジャーナル領域容量に応じて、RPO を達成するために最小コストの構成の資源量を算出することが可能となった。

なお本評価では単一ボリュームを対象とした模擬データによるシミュレーションを行ったが、ディザスタリカバリ対象すべてのボリュームへの書き込みデータ量を合算した上で、システム構成を設計すればよいことは明らかである。

#### 4.2 本研究の有効性

従来は書き込み量のピークにあわせてシステム性能を設計することが一般的であった。例えば図 8 に示す評価期間における書き込み量のピークは 300MB/15sec、すなわち 20MB/sec である。本研究以前には RPO に関わらず、このピークにあわせて 20MB/sec の回線帯域を設けることが通例であった。

提案方式により、RPO が 300sec であれば 6MB/sec の帯域で足りることを、構築前に算出することができる。その結果、RPO を満たしつつ 70% の回線帯域を削減できる。

### 5. クラウドへの活用

本章では、提案方式の汎用性について、近年導入が進むクラウドサービスへの活用方法を通して考察する。

#### (1) 既存システムークラウド間ディザスタリカバリ

この活用方法では、企業 IT システムによって生成されたデータをクラウドに転送する。転送先にクラウドを用いる場合、図 1 に示したような正サイトと副サイトで対となるストレージを設けることができるとは限らない。

本提案方式は前述の算出モデルを改編することで、

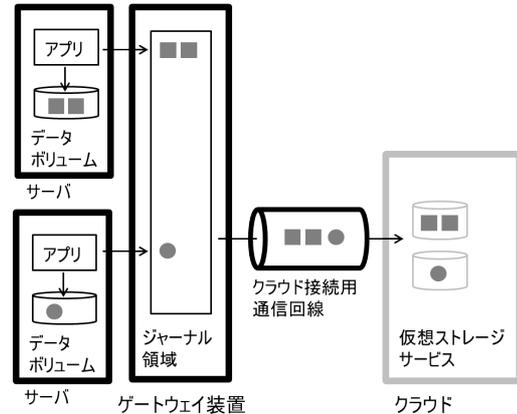


図 13 クラウド活用構成評価モデル  
Fig. 13 A configuration evaluation model of cloud service

ストレージのリモートコピー機能を用いないディザスタリカバリシステムにも適用可能である。

本モデルでは、例えばデータベースなどのアプリケーションがデータを生成・更新し、ゲートウェイ装置を通じてクラウドへ発送する構成を想定する(図 13)。ゲートウェイ装置には、ストレージサービスに接続するための専用アプライアンス、または一般的なキャッシングプロキシを利用する。

本モデルを用いた構成評価では、ゲートウェイ装置のジャーナル領域が図 3 の正サイトストレージにおけるジャーナル領域に該当し、3.2 節に挙げたリカバリポイント計算方法をそのまま流用できる。このとき、流入量 ( $In$ ) には各サーバが更新する転送対象データ量の時系列推移を、流出量 ( $Out$ ) の上限は同システムとクラウド間接続用の通信回線性能に相当する。

#### (2) クラウドークラウド間ディザスタリカバリ

本提案方式は、OpenStack<sup>☆</sup>などのクラウド管理ソフトウェアを利用したクラウドシステム間のディザスタリカバリにおける資源算出にも活用できる。

これらのクラウドシステムでは、生成された仮想サーバ、仮想インスタンスの存続中のみデータを保持するローカルストレージ領域の他に、永続的なデータ保管を目的としたブロックストレージサービスを利用することが可能である。同サービスを提供するストレージには、OpenStack における Cinder などの仮想ストレージシステムによる実装が知られており、それらは 1.3 節に述べた外付けストレージシステムの適用が可能であるため、このような構成のディザスタリカ

<sup>☆</sup> OpenStack は、米国における OpenStack,LLC の登録商標です。

バリシステムであれば、本提案方式を活用してコストを適正化できる。

## 6. 関連研究

従来はディザスタリカバリシステム構築前にリカバリポイントを予測するのではなく、構築後にシステム稼働状況を検証し、資源量の過不足に応じて構成を見直す方法がとられてきた。一方、提案方式は、構築前に最適資源量を見積もることができるため、従来方法よりもコストを低減できる可能性がある。

リカバリポイントを算出する方式について、従来稼働中システムのリカバリポイントを監視するために、転送対象データに付与されるシーケンス番号を用いてバッファ滞留時間を計測する方式が提案されている<sup>13)</sup>。また、リカバリポイントを容易に特定できるように、正サイトストレージ側で付与される転送データのタイムスタンプを記録する方法も提案されている<sup>15)</sup>。従来方式に比べ、本方式は、シーケンス番号やタイムスタンプを用いることなく、ストレージへのデータ流入量のみでリカバリポイントを予測できる。したがって、提案方式は従来よりも少ない情報量で、高精度な予測が可能になる。

## 7. おわりに

本論文では、ディザスタリカバリシステムの構成設計にあたり、通信回線帯域とジャーナル記憶領域の資源量を過不足無く適切に計算する手法を提案した。同計算にあたってはシステム構成とその処理手順をモデル化することで、ジャーナル領域におけるデータ蓄積量を算出可能とした。さらにその時系列推移から各時刻におけるリカバリポイントを計算し、ディザスタリカバリの性能目標であるRPOを達成するかどうか評価することで、同構成の妥当性を判定する手法を考案した。さらにシミュレーションを用いた評価実験により、その有効性を示した。

提案方式により、システム導入前の構成設計時にデータ消失リスクとシステム所有コスト、運用コストを適切に制御することが可能となり、要件に応じたディザスタリカバリシステムを構築できる。

## 参考文献

- 1) Gantz, J., Reinsel, D.: The Digital Universe Decade - Are You Ready?, IDC - IVIEW (2010)
- 2) LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big Data, Analytics and the Path From Insights to Value, MIT Sloan

- Management Review (2011)
- 3) Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers A.H.: Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute (2011)
- 4) Patterson, D.A.: A Simple Way to Estimate the Cost of Downtime, Proc. LISA '02: USENIX 16th System Administrators Conference (LISA '02), pp.185-188 (2002)
- 5) Toigo, J.W.: Disaster Recovery Planning, Principle Hall (2003)
- 6) 谷井成吉: コンピュータシステム災害復旧の対策, ダイヤモンド社 (2006)
- 7) Keeton, K., Santos, C., Beyer, D., Chase J. and Wilkes, J.: Designing for Disasters, Proc. 3rd USENIX Conference on File and Storage Technologies, pp.59-62 (2004)
- 8) Rudolph, C.G.: Business continuation planning/disaster recovery, IEEE Communications Magazine, Vol.28, Issue 6, pp.25-28 (1990)
- 9) Patterson, D.A., Gibson, G. and Katz, R.H.: A case for redundant arrays of inexpensive disks (RAID), SIGMOD '88: Proc. 1988 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, ACM Press, pp.109-116 (1988)
- 10) 大和純一, 管 真樹, 菊池 芳秀: 広域災害に対するストレージによるデータ保護, 電子情報通信学会, Vol.89, No.9 (2006901), pp.801-805 (2006)
- 11) 加倉井 宏一, 荻田 光一郎: 災害対策システムのリニューアルにおける現実的災害対策レベルの評価, 情報処理学会研究報告, Vol.2004, No.106, pp.1-6 (2004)
- 12) Gopisetty, S.: Automated planners for storage provisioning and disaster recovery, IBM Journal of Research and Development, Vol.52, No. 4/5, pp.353-366 (2008)
- 13) 江丸 裕教, 高井 昌彰, 原 純一: ディザスタリカバリにおける非同期リモートコピーのリカバリポイント監視方式, 情報処理学会研究報告, Vol.2010-EVA-31, No.1 (2010)
- 14) Mengzhi W, Kinman A, Anastassia A, Anthony B, Christos F, Gregory G: Storage Device Performance Prediction with CART Models, IEEE/ACM International Symposium, 2004
- 15) Shulman, R.R.: Disaster Recovery Issues and Solutions, HDS White Paper (2004)
- 16) 丸山 直子, 田口 雄一, 山本 政行: ディザスタリカバリシステムにおけるストレージリモートコピー構成評価モデルの提案, 情報処理学会第70回全国大会講演論文集 “4-539” - “4-540” (2008)