情報処理学会論文誌 Vol.58 No.2 495–505 (Feb. 2017)

リスクアウェア複製システムにおいて構成変更による再配置データ量を抑制する複製先部分再選択方式

松本 慎也[1,2,a)] 中村 隆喜[2] 村岡 裕明[2]

受付日 2016年5月8日, 採録日 2016年11月1日

概要：本論文では，リスクアウェア複製を適用し運用中の分散ストレージシステムにおいて，システム構成を変更する際に高効率なデータの再配置方式を提案する．リスクアウェア複製は，地方自治体の支庁舎や企業の支店など地域に分散した拠点で構成する分散ストレージシステムにおいて，災害リスクに基づいて各拠点間のデータ複製先を決定することで，災害環境におけるシステム内のデータの可用性を最適化する方法である．容量追加などシステム構成が変更された際，リスクアウェア複製の再適用によってより安全な拠点に複製先を変更すれば，データの可用性がより向上するという期待がある．しかし，多くの複製先を変更した場合に生じる大量のデータ再複製処理は，長時間にわたってシステムに高い負荷を与えうる問題がある．そこで，本論文では，少ないデータ複製量で効率良く可用性を向上させることを目指し，システム全体でなく部分的にのみ複製先変更を許容する 2 つの方式を提案する．複製先変更拠点の選択とその複製先の決定とを数理計画法で同時に行う同時選択方式と，災害リスクの高さに基づいて複製先変更拠点を選択した後，その複製先を決定する事前選択方式である．この 2 方式を地震災害シミュレーションで比較評価した結果，事前選択方式が，リスクアウェア複製を再適用する場合に要するデータ複製量の 34%でそれと同等の可用性を得て，同時選択方式に比べて実用上の効率に優れることを示す．

キーワード：災害，数理計画法，ストレージシステム

Partial Sites Determination Method to Mitigate Data Re-transfer for Reconfiguring Risk-aware Replication System

Shinya Matsumoto[1,2,a)] Takaki Nakamura[2] Hiroaki Muraoka[2]

Received: May 8, 2016, Accepted: November 1, 2016

Abstract: This paper proposes a highly efficient data reallocation method for a system reconfiguration case such as storage expansion in an operated distributed storage system applied Risk-aware Data Replication (RDR) to. RDR is a data availability optimization method against a widespread disaster such as an earthquake, which makes site pairs for data replication based on the site-to-site disaster risks in a distributed storage system such as branch offices of governments or companies. RDR can improve the availability of the reconfigured system because the reconfiguration changes parameters for the optimization. However, it can cause huge data replication that burdens the system for a long time. Therefore, this paper proposes two methods for changing replication sites, which allows to reselect only a part of the sites, to improve the data availability with small amount of replication data. The earthquake simulation results show that the preselection method, which makes higher risk sites reselect new replication sites on the priority basis, gets the same availability with the RDR reapplication case with 34% of data replication amounts, and that is much better than the combined selection method, which gives each site an opportunity to reselect new replication sites based on the original objective function.

Keywords: disaster recovery, mathematical programming, storage system

1. はじめに

様々な情報サービスの常時提供が求められている．地震や津波などの深刻な災害直後の被災地においても，身元確認や被災者ケアのために，電子化された戸籍情報や医療情報の継続的な提供が重要である．

被災地での継続的な情報サービスの提供のために，ストレージシステムにおいて，災害に対して安全な複製先を見出し，データをあらかじめ複製するリスクアウェア複製という手法が提案されている [1]．この手法では，ストレージシステムは複数の離れた拠点に設置された複数のストレージ装置（以下，拠点ストレージと呼ぶ）からなる．災害によって各々の拠点ストレージどうしが同時に損壊するリスクの高さを数理計画法で評価し，全体としてリスクが低くなる複製先の組合せを見出す．クラウドなどの遠距離ではなく，近隣の拠点で構成されるストレージシステムにこれを適用することで，広域網が断絶するような深刻な災害直後であってもその被災地でデータを保護し，災害環境におけるデータ可用性を維持する．

従来のリスクアウェア複製の研究では，各拠点ストレージが複製先を初めて選択する場合を主な対象としていた [1], [2], [3]．すなわち，分散ストレージシステムにリスクアウェア複製を新規適用する際の複製先決定問題に重点を置いていた．

このようなシステムでは，平時の運用において，拠点ストレージへの容量追加などの構成変更が生じるため，いったん定めた複製先の組合せは見直されるべきである．たとえば，災害リスクとして津波を考慮するシステムにおいて，海側に比べて山側の拠点ストレージの容量が追加された場合，山側へデータを複製する拠点ストレージを増やすことで津波からデータをより良く保護できる．

しかし，構成変更のたびに従来のリスクアウェア複製を適用すると，新たな複製先へのデータ複製に時間がかかるという問題が生じうる．すでに運用されデータを蓄えた拠点ストレージの複製先を変更するには，データ差分だけではなくデータ全体を複製しなければならないためである．

したがって，本研究では，構成変更を行ったリスクアウェア複製システムにおいて，少ないデータ複製量で効率良く可用性を向上させるために，全拠点ではなく部分的な複製先の変更だけを許容する 2 つの方式を示す．複製先を変更する拠点を，その複製先拠点の決定と同時に数理計画法で選択する同時選択方式と，複製先拠点の決定前に災害リスクの高い順に複製先を再選択する拠点を選択する事前選択方式である．同時選択方式は，事前選択方式に比べて多くの組合せをテストするため，可用性向上効果がより高く，計算時間がより長くなることが予想される．計算時間が短いほど，構成変更に合わせたより安全な複製先拠点を早期に再選択できるため，不意の災害に対するデータの安全性をより早く高められる利点がある．そのため，これら 2 方式の可用性向上効果と計算時間を評価する．

[1] 株式会社日立製作所
Hitachi Ltd., Yokohama, Kanagawa 244–0817, Japan
[2] 東北大学
Tohoku University, Sendai, Miyagi 980–8577, Japan
[a)] shinya.matsumoto.gn@hitachi.com

2. リスクアウェア複製

2.1 リスクアウェア複製の概要

リスクアウェア複製は，分散ストレージシステムが，広域網への通信が断絶するような大災害直後に被災地で情報サービスを継続的に提供するための技術である [1]．

リスクアウェア複製は，複製先を選択するステップと，データを複製するステップの 2 つからなる．複製先を選択するステップでは，サービス拠点の近隣拠点のうち災害に対するリスクの小さい拠点を複製先拠点として選択する．分散ストレージシステムを構成する複数のサービス拠点が，それぞれ複製先拠点を選ぶ際に全体として最も安全な複製先の組合せを選ぶことが重要であり，その複製先の組合せを探し出す問題を複製先決定問題と呼ぶ．データを複製するステップでは，複製先決定問題を解いて得られた組合せに，各拠点が実際にデータを複製する．

4 拠点からなるシステムの複製先決定問題を**図 1** に示す．ここで，各拠点が持つストレージ装置は複製データを格納するための空き容量 F と複製データのデータ量 D を持つ．また，各拠点間は複製した場合のデータ喪失の危険度を示す災害リスク P を持つ．これらの文字に対する添え字はその拠点番号を示す．複製先決定問題は，4 拠点がそれぞれ所定の数のデータを複製する際に，各拠点がその空き容量を超えることなく，全体としてデータ喪失の危険度が最も低い複製先の組合せを探索する問題である．

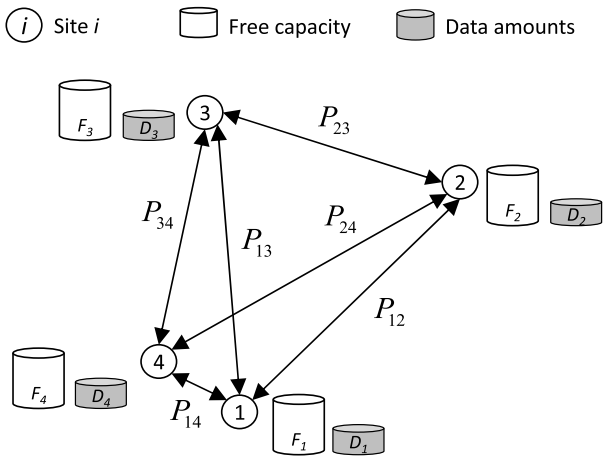

図 1 4 拠点からなる複製先決定問題
Fig. 1 Replication site decision problem in 4 sites case.

2.2 複製先決定問題の数理モデル

複製先決定問題は，整数計画問題として表現できる．その問題表現は，目的関数，冗長度制約，空き容量制約の3つからなる．以下，簡単のため，複製数1の場合の複製先決定問題についてその整数計画問題表現を示し，複製数2以上の場合の定式化方法を説明する．

最小化すべき目的関数は想定する災害に対して各拠点のデータ喪失見込み量の総和であり，式 (1) のように表現される．

$$\min f(x_{ij}) = \sum_{i \in S} \sum_{j \in S} D_i P_{ij} x_{ij} \tag{1}$$

ここで，S は全拠点の集合である．たとえば4拠点で構成されるシステムでは $S = \{1, 2, 3, 4\}$ であり，各々の数字が異なるストレージ装置を示す．x_{ij} は拠点 $i \in S$ から $j \in S$ への複製を作るか否かを示す変数であり，0 または 1 の値をとるバイナリ変数である．$x_{ij} = 0$ のとき拠点 i から拠点 j への複製を作らないことを意味し，$x_{ij} = 1$ のとき拠点 i から拠点 j への複製を作ることを意味する．なお，同じ拠点に複製を作らないために $x_{ii} = 0$ と定義する．D_i は拠点 i の格納するデータ量，P_{ij} は拠点 i から拠点 j にデータを複製した場合に予想されるデータ喪失確率であり，0 から 1 間の実数値をとる．たとえば，震源は対象とするフィールドのどの地点でも同じ確率で起きえて，地震の強度の減衰が地震動減衰式 [4] に基づくと仮定すると，地震に対するデータ喪失確率を式 (2) のように表すことができる．

$$P_{ij}(d_{ij}) = \begin{cases} \dfrac{1}{1 + e^{a(\log_{10} d_{ij} + b)}}, & i \neq j \\ 1 \qquad\qquad , & i = j \end{cases} \tag{2}$$

ここで，d_{ij} は拠点 i と拠点 j との間の地理的距離を示す．a および b は想定する災害の伝播を指定するパラメータであり，a は距離に対するリスクの伝播の「傾き」，b はリスクが全体の半分になる値を示す「中央値」である．このデータ喪失確率は震源を仮定しないため，それによるデータ喪失の影響が小さい解を得ることができる．また，他のデータ喪失確率の例として，地震ハザードステーション [5] が提供するようなハザードマップ情報を数値的に用いることができる．

次に，冗長度制約を式 (3) に示す．

$$\sum_{j \in S} x_{ij} = 1, \forall i \in S \tag{3}$$

ここで，右辺の値 1 は拠点 i の複製先拠点の数を示す．この制約は，各々の拠点に設定する複製先拠点の数を 1 つに制限する．

次に，空き容量制約を式 (4) に示す．

$$\sum_{j \in S} D_i x_{ij} \leq F_j, \forall j \in S \tag{4}$$

ここで，F_j は拠点 j が複製データを受け入れるために用意する空き容量である．この制約は，各々の拠点が受け入れるデータ複製量をユーザが指定した値に制限する．

式 (1)，(3)，(4) で表現される整数計画問題を，たとえば分枝限定法 [6] のような整数計画問題のアルゴリズムで解くことで，データ喪失の危険度が低い複製先の組合せを得られる．得られた組合せを，実際のシステムで各拠点がデータを実際に複製して実現することで，災害環境におけるシステムの可用性を向上させることができる．

複製数が2以上の場合の複製先決定問題も，複製先が1の場合と同様の考え方を用いて定式化できる．すなわち，目的関数としてデータ喪失見込み量，制約条件として，冗長度制約と空き容量制約をそれぞれ定義する．

複製数2の場合の目的関数を式 (5) に示す．

$$f\left(x_{ijk}\right) = \sum_{i \in S} \sum_{j \in S} \sum_{k \in S} D_i P_{ijk} x_{ijk} \tag{5}$$

ここで，$x_{ijk} \in \{0, 1\}$ であり，拠点 $i \in S$ から拠点 $j \in S$ と拠点 $k \in S$ の両方へ複製を作るか否かを示す変数である．$x_{ijk} = 1$ のとき拠点 i から拠点 j と拠点 k の両方へ複製を作ることを意味する．同じ拠点に複製を作らないために $x_{iii} = x_{iij} = x_{iji} = x_{ijj} = 0$，同じ拠点について複製先を入れ替えた場合を区別しないように $k > j$ のとき $x_{ijk} = 0$ と定義する．P_{ijk} は拠点 i から拠点 j と拠点 k の両方にデータを複製した場合に予想されるデータ喪失確率である．複製数1の場合と同様に，ハザードマップ情報から値を得られる．

複製数が2の場合の冗長度制約と空き容量制約は，上記で定義した x_{ijk} を用いてそれぞれ定義する．冗長度制約では，各々の拠点が複製先を2つ選ぶために，それぞれの拠点 i に対して，すべての x_{ijk} のうち1つだけが1であるよう制約する．空き容量制約では，各々の拠点が受け入れる複製データの数を制約するために，x_{ijk} にデータ量 D_i を掛け合わせたものの総和が，受け入れ拠点の空き容量以下であるよう制約する．

なお，複製数が3以上の場合も，複製数2の場合と同様に定式化できる．たとえば，複製数を3つ選ぶ場合のデータ喪失見込み量の総和を目的関数として定義する．

2.3 リスクアウェア複製における課題

容量拡張などの構成変更がなされたストレージシステムにリスクアウェア複製を再適用すれば，災害時の可用性を高められる．構成変更によって空き容量などのパラメータが変動した複製先決定問題を再度解くことで，より安全な複製先の組合せを得られるからである．

しかし，実際のシステムでは，複製先決定問題を解いて得られた組合せが，運用上の理由で実現できない可能性がある．リスクアウェア複製の再適用によって複製先を変更することになった拠点が変更先の拠点にデータセットを複

製する処理が，長期的に大きな負荷をシステムにかける恐れがあるためである．

複製先拠点を変更する際に拠点が送信すべきデータ量は一般に大きいことが予想される．一般的なバックアップシステムでは，日々の複製を差分バックアップで行うなど複製量を抑えているものの，複製先拠点を変更する場合はデータセット全体を複製しなければならない．さらに，複数世代のデータセットを複製先に格納するような一般的なバックアップ運用の場合には，その世代数に応じてデータ量が増加する．

以上のことから，本研究における課題は，構成変更を行ったリスクアウェア複製システムにおいて，少ないデータ複製量で効率良く可用性を向上させることである．

3. 提案方式

3.1 アプローチ

少ないデータ複製量で効率良く可用性を向上させるために，高い可用性向上が見込まれる拠点にだけ複製先を変更させるアプローチをとる．以下では，容量拡張による構成変更があった場合に，高い可用性向上が見込まれる拠点を見出す手順が異なる 2 つの方式を示す．

3.2 同時選択方式

同時選択方式は，複製先変更拠点の選択をその複製先の決定と同時に数理計画法で行う方式である．複製先の変更により生じるデータ複製量を制約する条件の範囲で，目的関数を最小化する拠点の組合せを求める．そのため，少ないデータ複製量で高い可用性を得る期待がある．一方で，複製先変更拠点のすべての組合せの中から最適解を探索する仕組みのため，計算時間が長くなる懸念がある．データ複製量の上限値は，運用状態を考慮してユーザが指定することを想定する．

以下，同時選択方式の数理表現を示す．簡単な表記のため，複製先の再選択前後で各拠点の格納するデータ量は一定とする．

まず，各拠点の複製数が 1 の場合，同時選択方式における数理計画問題は，次に示すデータ複製量の制約条件 (6) を，複製先決定問題 (1)，(3)，(4) に加えた問題である．

$$\sum_{i\in S}\sum_{j\in S}\left(1-x^*_{ij}\right)D_i x_{ij} \le T \tag{6}$$

左辺は，各拠点がデータを複製し直す量の総和を示す．D_i は拠点 i の格納するデータ量，S は全拠点の集合を示す．x^*_{ij} は前回の複製先選択時に拠点 i が j を複製先としたかどうかを示す定数であり，拠点 i が j を複製先にしていた場合に 1，していない場合に 0 をそれぞれとる．x_{ij} は今回の複製先選択において拠点 i が j を複製先にするかどうかを示す変数であり，拠点 i が j を複製先にする場合に 1，しない場合に 0 をそれぞれとる．したがって，左辺のとりうる具体的な値は**表 1** のとおりである．

表 1 拠点選択による各拠点のデータ複製量
Table 1 Replication data amounts of each site in selecting targets.

Case #	x^*_{ij}	x_{ij}	値
1	1	1	0
2	1	0	0
3	0	1	D_i
4	0	0	0

Case #1 は，前回の複製先選択で選択した複製先を選択する場合である．新しくデータを複製する必要はないため，データ複製量は 0 である．Case #2 は，前回の複製先選択で選択した複製先を選択しない場合であり，同様の理由でデータ複製量は 0 である．Case #3 は，前回の複製先選択で選択した複製先と異なる拠点を選択する場合である．全データを複製する必要があるため，データ複製量は複製元のデータ量 D_i である．Case #4 は，前回の複製先選択で選択した複製先とは異なる拠点を選択しない場合である．新しくデータを複製する必要がないため，データ複製量は 0 である．

式 (6) の右辺 T は，ユーザが指定するデータ複製量の上限値である．システムの性能やバックアップウィンドウ，システムの運用スケジュールを考慮し，許容可能なデータ複製量をユーザが入力する．

許容可能なデータ複製量の決定方法は複数考えられる．1 つの簡易な方法は，システム全体のデータ量の割合を固定的に用いる方法である．たとえば，管理者はシステム全体のデータ量の 10%の値を用いる．別の方法として，バックアップウィンドウを考慮して複製完了が見込める上限値を用いる方法がある．管理者やシステム構築者が，あらかじめ全拠点が同時にデータ複製を行ったときの各拠点のデータ複製性能を計測し，その総和 S(GB/s) を得ておく．そして，バックアップウィンドウ W(s) が与えられたときに，上限値 T(GB) としてその積 SW を用いる．たとえば，データ複製性能の総和 $S = 0.010$ のシステムを，バックアップウィンドウ $W = 28800$ で運用する場合，上限値 $T = 288$ を用いる．

次に，各拠点の複製数が 2 の場合も同様の考え方を使って問題を表現できる．同時選択方式における数理計画問題は，次に示すデータ複製量の制約条件 (7) を，複製数 2 の場合の複製先決定問題に加えた問題である．

$$\sum_{i\in S}\sum_{j\in S}\sum_{k\in S}\left(1-x^*_{ijk}\right)D_i x_{ijk} \le T \tag{7}$$

左辺は，複製数 2 の場合の各拠点がデータを複製し直す量の総和を示す．$x^*_{ijk} \in \{0, 1\}$ は，前回の複製先選択時に拠点 i が拠点 j と拠点 k の両方を複製先としたかどうかを示す定数である．拠点 i がその両方を複製先にしていた場合

に 1 をとる．

なお，複製数が 3 以上の場合も，複製数が 2 の場合と同様である．複製数が 3 以上の場合の複製先決定問題に，式 (7) のようなデータ複製量を制約する条件式を加える．

3.3 事前選択方式

事前選択方式は，複製先拠点の決定前に，災害リスクの高い順に複製先を再選択する拠点を選択する方式である．複製先を再選択する拠点を，複製先の変更により生じるデータ複製量の上限値を超えない範囲で災害リスクの高い順に選択した後，その複製先を再選択する．災害リスクの高い複製関係をいったん解消して，その複製元の新しい複製先を再選択する一方で，すでに災害リスクの低い複製関係を維持するため，少ないデータ複製量で高い可用性を得る期待がある．ただし，複製先を再選択する拠点に対してのみ複製先を探索するため，同時選択方式に比べて可用性向上効果が小さい一方で，計算時間が短いという期待がある．データ複製量の上限値は，運用状態を考慮してユーザが指定することを想定する．

以下，簡単な表記のため，複製先の再選択前後で各拠点の格納するデータ量は一定とする．

複製数が 1 の場合，事前選択方式による複製先の再選択は，複製元を選択する過程（Step 1）と，選択された複製元の新しい複製先を選択する過程（Step 2）で構成される．ここで，S_{SEL} は複製先を再選択する対象として選択された複製関係の複製元拠点の集合，x^*_{ij} は再選択前の複製関係，すなわち，前回の複製先選択時点で拠点 i が j を複製先として選択したかどうかを示す定数である．

Step 1：ユーザが与える複製データ上限値 T を超えない範囲で，災害リスクが高い複製関係から順番にその複製元拠点を選択し，複製先再選択対象拠点集合 S_{SEL} を得る．

Step 2：以下に示す目的関数 (8)，冗長度制約条件 (9)，空き容量制約条件 (10) で構成される複製先決定問題を解く．

目的関数を式 (8) に示す．

$$\min f(x_{ij}) = \sum_{i \in S_{SEL}} \sum_{j \in S} D_i P_{ij} x_{ij} + \sum_{i \in \neg S_{SEL}} \sum_{j \in S} D_i P_{ij} x^*_{ij} \tag{8}$$

式 (8) は，データ喪失量の総和を示す．第 1 項は複製先を再選択する拠点によるデータ喪失量の総和を意味する．第 2 項は複製先を再選択しない拠点によるデータ喪失量の総和を意味し，この項は変数でなく定数である．上記では第 2 項を記載したが，定数のため省いて記載しても等価である．

冗長度制約を式 (9) に示す．

$$\sum_{j \in S} x_{ij} = 1, \forall i \in S_{SEL} \tag{9}$$

式 (9) は，複製先を再選択する拠点に関する複製数制約条件である．再選択しない拠点は複製をしないため，すでに冗長度制約を満たしており，制約が不要である．

空き容量制約を式 (10) に示す．

$$\sum_{i \in S_{SEL}} D_i x_{ij} \leq F_j - \sum_{i \in \neg S_{SEL}} D_i x^*_{ij}, \forall j \in S \tag{10}$$

式 (10) は，複製先を再選択する拠点に関する空き容量制約条件である．左辺は複製先を再選択する拠点による複製データ量を意味する．右辺は，拠点の空き容量から，複製先を再選択しない拠点による複製済みのデータ量を差し引いた残りの空き容量を意味する．複製先を再選択しない拠点は複製をこれ以上増やさないため，右辺は定数である．

複製数が 2 以上の場合も同様の考え方で問題を定義できる．すなわち，上記 Step 2 における複製先決定問題のように，再選択する拠点がすべての拠点の中の一部であることと，再選択しない拠点がデータを複製済みであることを複製先決定問題に反映すればよい．

複製数が 2 の場合，事前選択方式による複製先の再選択は，上記 Step 2 における複製先決定問題において，以下の式 (11) に示す目的関数を用いる．

$$f(x_{ijk}) = \sum_{i \in S_{SEL}} \sum_{j \in S} \sum_{k \in S} D_i P_{ijk} x_{ijk} + \sum_{i \in \neg S_{SEL}} \sum_{j \in S} \sum_{k \in S} D_i P_{ijk} x^*_{ijk} \tag{11}$$

式 (8) と同様に，式 (11) はデータ喪失量の総和を示す．第 1 項は複製先を再選択する拠点によるデータ喪失量の総和を意味する．第 2 項は複製先を再選択しない拠点によるデータ喪失量の総和を意味し，この項は変数でなく定数である．上記では第 2 項を記載したが，定数のため省いて記載しても等価である．

複製数が 2 の場合の冗長度制約と空き容量制約も，同様の形式でそれぞれ定義する．冗長度制約では，複製先を再選択する拠点が複製先を 2 つ選ぶために，それぞれの拠点 i に対して，すべての x_{ijk} のうち 1 つだけが 1 であるよう制約する．空き容量制約では，各々の拠点が受け入れる複製データの数を制約するために，x_{ijk} にデータ量 D_i を掛け合わせたものの総和が，受け入れ拠点の，複製済みデータを差し引いた空き容量以下であるよう制約する．

なお，複製数が 3 以上の場合も，複製数 2 の場合と同様にして定義できる．

4. 評価

提案する 2 方式に関する評価を示す．

4.1 評価指標

評価指標は，データ残存割合 A と計算時間 C である．

データ残存割合 A は，構成変更後のシステムにおける可用性を示す指標である．データ残存割合は，対象システムにおける，災害前の全データ量に対する災害後の全残存

表 2 シミュレーションパラメータ
Table 2 Simulation parameters.

分類	パラメータ	値
対象システム	拠点数	100
	拠点配置方法	ランダム
	フィールドの広さ	40 km × 20 km
	各拠点のデータ量D_i	10
	各拠点の空き容量F_i	10
複製条件	複製数	1
	想定地震における震源距離とリスク値(d_{ij}, P_{ij})	(5 km, 0.2), (20 km, 0.1)
	複製関係を決定するアルゴリズム	分枝限定法
再選択条件	拠点への追加空き容量	10 (*)
	複製先を再選択する拠点数の割合	10-100 %
	再選択時に複製関係を決定するアルゴリズム	分枝限定法
地震条件	震源場所(X, Y, Z)	(15 km, 10 km, -50 km)
	マグニチュードM	7.215
	地震の減衰特性(傾きa, 中央値b)	(43.9, 1.75)

*) フィールドの端に近い 20 拠点について，その空き容量を 10 だけ追加する．

データ量の割合と定義する．ただし，複製による重複データはデータ量には含めないものとし，たとえば，オリジナルデータと複製データの両方が生存したとしても 1 つのデータとしてカウントする．定義より，データ残存割合の値は大きいほど可用性の高い優れた方式である．

データ残存割合の評価では，システムを構成する拠点のうち半数を損壊させる災害に対してデータ残存割合を測定する．データ複製量上限値をパラメータとしてデータ残存割合を複数測定し最大のデータ残存割合と同等の値となる最小のデータ複製量がいくらかを比較し，効率を評価する．

計算時間 C は，複製先を再選択する際のコストを測る指標である．計算時間は，複製先再選択のための複製先決定問題を解くのにかかる時間と定義する．定義より，計算時間の値が小さいほど高い頻度で複製関係を再選択できる優れた方式である．

計算時間の評価では，2 方式間の公平な評価のため，事前選択方式における複製元を選択する過程（3.3 節 Step 1）と，新しい複製先を選択する過程（3.3 節 Step 2）との含めた時間を計算時間とする．

4.2 評価方法

評価では，計算機上で生成した対象システムのモデルに対して複製関係を再選択し，その後災害を模擬するシミュレーションを行う．実行するシミュレーションは以下のとおりである．

まず，対象システムを定義する．計算機で，シミュレーションフィールドの大きさ，拠点数，各拠点の位置，複製すべきデータ量，複製データを格納する空き容量を設定する．次に，対象システムの複製先を決定する．定義した対象システムに対し地震に対して最もリスクの低い複製関係を生成し，その複製先へのデータ複製を行う．ここで，地震後に生存したデータ量を集計し，これをデータ複製量 0 に対するデータ残存割合 A の値とする．次に，地震前の状態に戻し，対象システムの複製先を再選択する．対象システムの中から想定する災害リスクが低い拠点に容量を追加し，2 つの提案方式を用いて，それぞれ複製先を再度選択させる．このとき，2 つの提案方式において再選択にともない要するデータ複製量の上限値を変化させ，再選択する拠点の数を変化させ，この再選択にかかる計算時間 C を得る．再選択後の複製関係にデータ複製を行い，そのときのデータ複製量の値を得る．最後に，複製先再選択前と同じ条件の地震を起こして生存したデータ量を集計し，データ複製量に対するデータ残存割合の値をそれぞれ得る．

このシミュレーションに用いるパラメータを**表 2** に示す．

対象システムは 40 km × 20 km のフィールドにランダムに配置された 100 拠点からなるストレージシステムである．各々の拠点のストレージ装置は 10 の複製すべきデータを持つ．また，各々のストレージ装置は複製先の選択前に 10 の複製データ格納用の空き容量を持つ．

対象システムが複製関係を生成する際の条件は次のとおりである．各ストレージ装置の複製数は 1 とする．リスクとして，式 (2) で示したデータ喪失確率を利用し，その設計パラメータ (a, b) を，震源から 5 km と 20 km 離れた地点のデータ喪失確率の値がそれぞれ 0.2，0.1 になるように定める．いずれの方式においても，複製先決定問題を解くために，整数計画問題の汎用アルゴリズムとして広く用いられる分枝限定法（Branch-and-Bound method）[6] を用いる．

対象システムが複製先を再選択する際の条件は次のとおりである．震源から遠い順に選ばれた 20 拠点のストレージ装置の空き容量を拡張し，別の拠点の複製データをさらに格納可能にする．その拡張量は 10，すなわち，システム全体の容量の 20%に相当する．容量拡張後に，対象システムを構成する拠点のうち再選択を行う対象は，全システムの 10%から 100%まで 10 刻みとする．これは，同時選択方式では再選択にともなうデータ複製量の上限値 T が 100 から 1000 まで 100 刻みであること，事前選択方式では拠点の選択数が 10 から 100 まで 10 刻みであることをそれぞれ意味する．また，再選択のための複製先決定問題で用いるアルゴリズムは分枝限定法である．

なお，実用的なケースの評価のため，各拠点は複製先を選択する際に 10 のデータのうち一部を複製できるものとする．この条件を実現するため，10 のデータを持つ 1 つの実際の拠点を，1 のデータを持つ 10 の仮想的な拠点としてシミュレーション上に表現する．すなわち，シミュレーション上では，仮想的な拠点数は 1000 であり，各仮想的な拠点はデータを 1 だけ持つ．1 拠点を表現する 10 の仮想的な拠点のうち 1 つだけが空き容量 10 を持ち，他の空き容量は 0 とする．これにより，シミュレーション上，冗長度制約は 1 のままであるが，実際の拠点の複製先は 2 つ以上にできるよう構成する．

対象システムに与える地震の条件は次のとおりである．地震はシステムに最も大きな被害をもたらすフィールド中央（15 km, 10 km）を震源とする直下型地震とし，その深さは 50 km である．また，マグニチュードおよび地震動の伝播を伝えるパラメータは拠点の半数が損壊するように調整する．地震のシミュレーションには，地震強度の減衰が地震動減衰式 [4] に従うとして設計・実装したシミュレータ [1] を用いる．このシミュレータでは，以下のステップに従って各拠点の損壊確率が計算され，地震により損壊するか否かがシミュレーションごとに確率的に決定される．

Step 1：地震動減衰式 [4] を示す式 (12) から，シミュレーションフィールド上の各点での地震強度 E を得る．

$$\begin{aligned}&\log_{10} E = 0.58M + 0.0038Z - 1.29\\&-\log_{10}(d + 0.0028 \cdot 10^{0.50-M}) - 0.002d\end{aligned} \tag{12}$$

ここで，M は地震のマグニチュード，Z は震源の高さ，d は震源から各点への距離である．なお，式 (12) における地震強度 E とは地表面が移動する速さを意味しており，一般に言及される震度はこの速さから算出される．

Step 2：得られた地震強度を用いて，式 (13) から各拠点の損壊確率 Q を得る．

$$Q = \frac{1}{1 + e^{a(-E+b)}} \tag{13}$$

ここで，a，b は地震の減衰係数であり，地震動が伝播する地面の地質などから定まる定数である．

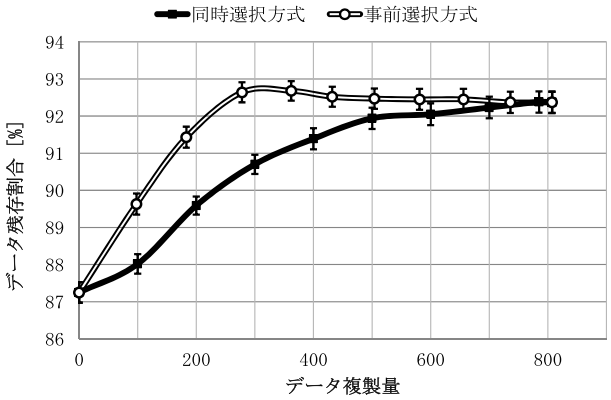

図 2 データ残存割合のシミュレーション結果
Fig. 2 Simulation results on data available ratio.

上記シミュレーションを，2 コア 1.87 GHz の Intel Xeon [*1] E5502 のプロセッサと 6 GB のメモリを搭載するサーバで実行し，結果を計測する．なお，統計誤差を打ち消すため，10 通りの拠点配置パターンを用意し，各パターンについて空き容量を追加して複製先の再選択を行った後，500 回の地震シミュレーションを実行する．評価指標は，この結果測定されたデータの平均をとる．すなわち，データ複製量に対するデータ残存割合 A は計 5000 通り，計算時間 C は計 10 通りのデータの平均である．

4.3 評価結果

複製先再選択時のデータ複製量 M に対するデータ残存割合 A の測定結果を**図 2** に示す．なお，各データ点のエラーバーは，標準誤差による信頼度 95%の信頼区間を示す．

図 2 から，2 方式とも，データ複製量が大きいほどデータ残存割合が大きくなる傾向があることを示しており，データ複製量の最大値 807 であることを示している．807 を超えるデータ複製量を指定してもデータ残存割合は向上しない．これは，構成変更後のシステムにリスクアウェア複製を再適用した場合，すなわち従来手法により得た複製関係に相当する．

なお，図 2 において，事前選択方式は，全拠点を選択する場合に比べて，一部拠点のみを選択する場合にデータ残存割合が高くなっているが，これは測定誤差であると考えられる．データ複製量 300 付近以上の各データ点の可用性に関する信頼区間は互いにほぼ一致しており，統計上，それらの間に有意な差は認められない．

図 2 は，事前選択方式が効率に優れることを意味する以下 2 つの事実を示す．

第 1 に，従来手法と比較して，事前選択方式は，少ないデータ複製量で同等のデータ残存割合を得る．従来手法のデータ複製量 807 と同等のデータ残存割合 92.4%を，データ複製量 278 以上であればつねに達成する．このデータ複

[*1] Intel および Xeon は，アメリカ合衆国および/またはその他の国における Intel Corporation の商標である．

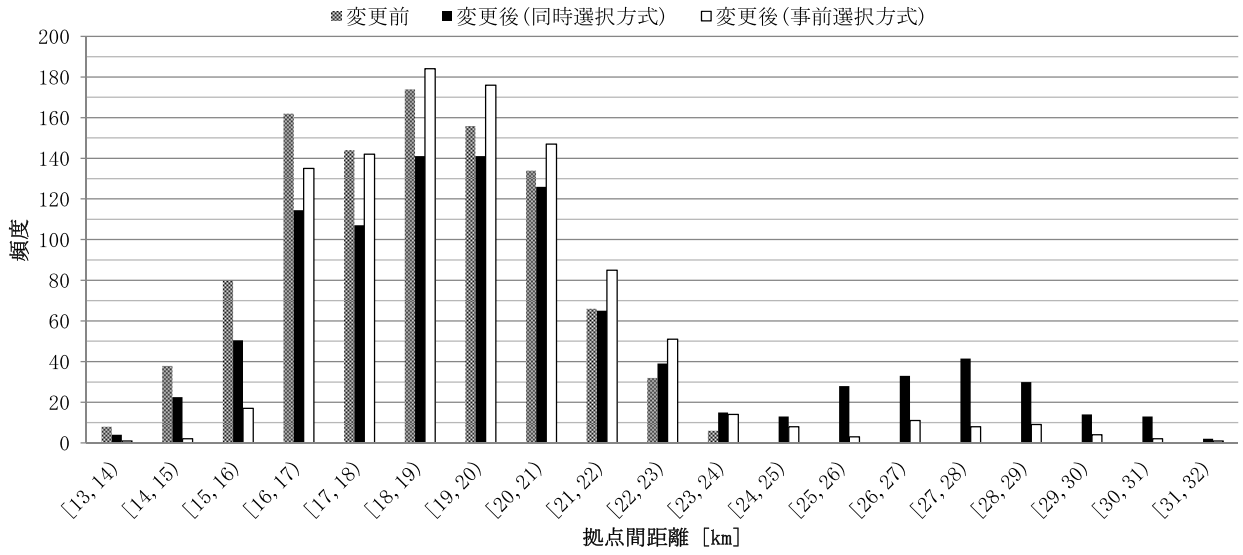

図 3 複製関係を持つ拠点間距離の分布
Fig. 3 Histogram of distance between two replication sites.

製量は，従来手法の 34%と少ない．一方で，同時選択方式は，データ残存割合 92.4%を得るためにデータ複製量 807，データ残存割合 92.0%を得るのにもデータ複製量 500 を要する．このデータ複製量は，従来手法の 62%に相当する．

第 2 に，同時選択方式に比べて，事前選択方式のデータ残存割合は，複製データ量が少ないときにより大きい．2 方式によるデータ残存割合の差は，データ複製量 200 から 300 のときに 2%以上である．これは，同時選択方式が事前選択方式と同等のデータ残存割合を得るために，200 から 400 ほど多くのデータを複製する必要があることを意味する．

数理計画問題を使って解く同時選択方式に比べて，事前選択方式のデータ残存割合が大きい理由として，災害リスクの高い拠点から順に複製先を再選択することが，本評価で起こした地震の強度や震源場所に対して効果的に働いたためと予想される．

次に，この予想を説明する測定結果として，複製関係を持つ拠点間距離の分布を**図 3** に示す．図 3 は，複製先を再選択する拠点数の割合が 20%の場合のものである．横軸は複製元拠点から複製先拠点までの距離を示しており，距離が大きいほどリスクが低い安全な複製関係である．縦軸はその拠点間距離を持つ拠点の頻度（拠点数）を示す．

図 3 は，同時選択方式はより強い災害に対して，事前選択方式はより弱い災害に対して，それぞれデータの安全性を強化する性質があることを示す．同時選択方式の場合，変更前に比べて，拠点間距離 15 km から 19 km の頻度が極端に減少し，25 km から 29 km の頻度が極端に増加する．これに対し，事前選択方式の場合，変更前に比べて，同時選択方式よりも近い拠点間距離 13 km から 16 km 周辺の頻度が極端に減少し，17 km から 23 km の頻度が極端に増加するが，同時選択方式のように遠い 24 km 以上の頻度は大きく増加しない．つまり，事前選択方式は，同時選択方式と異なり，過剰に高リスクな複製関係と過剰に低リスクな複製関係を作らないように動作する．実際，同時選択方式による拠点間距離は，平均値 20.3 km，分散は 14.8 km^2，事前選択方式は平均値 19.4 km，分散 6.4 km^2 であり，事前選択方式は全災害に対して平均的にリスクが高い一方で，分散が小さいために，弱い災害に対してはより良くデータを保護できることが分かる．

図 3 より，図 2 に示したデータ残存割合のシミュレーション結果において事前選択方式が同時選択方式よりもつねにデータ残存割合が大きくなった理由は，このような事前選択方式の性質が同時選択方式の性質よりも有効に動作したことである．さらに，図 2 の拠点間距離分布において，地震によるデータ喪失は拠点間距離が小さい順に起こりやすいことから，同時選択方式は，システムを構成する拠点の 80%程度が損壊する非常に強い地震の場合に，事前選択方式よりもデータ保護効果が高くなることが見込まれる．

次に，複製先再選択割合に対する計算時間 C を**図 4** に示す．なお，各データ点のエラーバーは，標準誤差による信頼度 95%の信頼区間を示す．

図 4 において，従来手法は，事前選択方式における複製先選択割合 100%のとき計算時間に相当する．これは，全拠点が複製先を選択しなおすケースだからである．なお，このケースで同時選択方式と事前選択方式の計算時間が一致しない理由は後述する．

図 4 は次の 2 つの事実を示す．

第 1 に，従来手法と比較して，事前選択方式の計算時間は小さい．複製先再選択割合が 10%のとき，従来手法の計算時間の 1.0%，50%のとき 14%である．この理由は，事前選択方式はあらかじめ選択された拠点だけを計算対象にするため，解くべき複製先決定問題の問題規模が従来手法

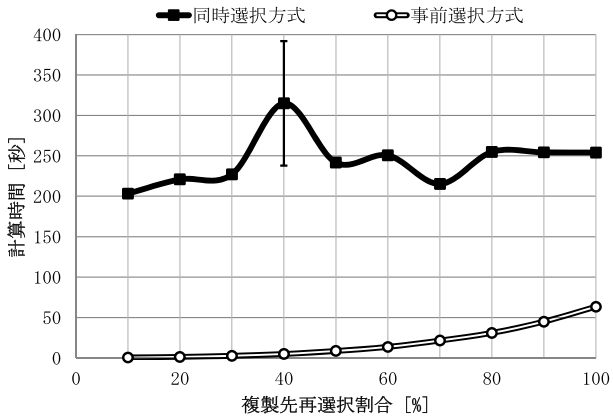

図 4 計算時間のシミュレーション結果
Fig. 4 Simulation results on computation time.

に比べてつねに小さくなることである．一方，同時選択方式の計算時間は，従来手法と比較して，どの複製先再選択割合のときも大きい．複製先再選択割合が 10%のとき従来手法の計算時間の 3.2 倍，50%のとき 3.8 倍である．この理由は，従来手法にない制約条件である式 (6) が加わったことである．分枝限定法が解を探索する過程で生成する部分問題の解（暫定解）が最適解になる可能性が低くなり，結果として限定操作が有効に働かず，多くの部分問題を解く必要が生じたと考えられる．なお，複製先再決定割合が 100%のとき，同時選択方式と事前選択方式の値が一致しないのは，この理由によるものである．

第 2 に，同時選択方式と異なり，事前選択方式の計算時間は，複製先再選択割合が大きくなるにつれて大きくなることである．この理由は，再選択する対象の範囲を大きくするにつれ，複製先決定問題の問題規模である複製先再選択割合，すなわち，複製先を再選択する拠点の数が大きくなり，その組合せの数が多くなることである．一方，同時選択方式は複製先再選択割合にかかわらず，対象システムを構成するすべての拠点の複製関係を基本的にすべて検査し直すために計算時間について増加傾向も減少傾向も見られない．ただし，複製先再選択割合が 40%のときに計算時間が他に比べてやや大きくなっているのは，実験で使用した最適化計算ライブラリで分枝限定法が効率的にはたらかない初期値が用いられてしまったためと考えられる．

複製数や拠点数が増加した場合，複製先決定問題の計算時間は指数関数的に増加することが従来研究 [2] より知られている．したがって，複製数や拠点数をより多くした場合では，両方式の計算時間の差も指数的に増加することになる．計算時間の差は解候補である全拠点の組合せの数の差に基づくためである．

以上から，システムを構成する拠点の半数が損壊する比較的強い災害条件であっても，従来手法と同等のデータ残存割合を，従来手法の 34%のデータ複製量で得ることができ，かつ，計算時間も従来手法より短い事前選択方式が，実用上優れることが分かった．

5. 関連研究

分散ストレージシステムにおいて，それを構成するストレージノードに格納されたデータを再配置する際にデータ複製量を小さく抑える仕組みが研究されている．

Dynamo [7] は，Consistent hashing [8] を用いて，複数のストレージノード間でデータを分散して格納する．Dynamo は，各ノードにリングハッシュ値空間の範囲を割り当て，空間内で後ろに続く値の範囲を担当するノードを，そのノードの持つデータの複製先とする．この仕組みにより，一部ノードの故障や追加・削除にともなって起こるデータの再配置は，そのノードの持つリングの周辺のノードのみに限られるため，システム全体に波及しないという利点がある．一方で，リスクを考慮して複製先を決定するシステムに用いると，データの安全性が極端に低下する恐れがある．データが再配置される際，複製先のリスクの大きさが考慮されないためである．

また，リスクを考慮してデータやサービスを配置するシステムに関する研究がある．

文献 [1], [2], [3] は，地理的に分散した複数の拠点ストレージからなるシステムにおいて，地震や津波などの自然災害リスクを定量評価して複製先を決定することで，災害に対するデータの安全性を高めるリスクアウェア複製に関するいくつかの手法を提案する．文献 [1] は，複製数 1 の場合の複製先決定問題が，定量化した災害リスクを用いることで整数計画問題として定式化できることを示す．さらに，汎用の整数計画アルゴリズムを用いて得た解の安全性を，災害シミュレーションを用いて評価し，リスクを考慮することによる有効性を示す．文献 [2] は，複製先が 2 以上の場合の複製先決定問題において，短い計算時間で高い安全性を持つ解を得られる反復計算アルゴリズムを提案する．複製数の増加により拠点ストレージの複製先の組合せが膨大になるという問題を，元の複製先決定問題を分割して得た複製数 1 の分割問題を反復的に解くことで解決できることを示す．さらに，得られた分割問題の解を，災害シミュレーションを用いて評価し，元の問題の解と同等の高い安全性を持つことを示す．文献 [3] は，複製数が 2 以上の場合の複製先決定問題の解を求める際に，データの安全度を指標として，拠点ストレージごとに異なる複製数を決定する手法を提案する．さらに，この手法により得られた解を，災害シミュレーションを用いて評価し，拠点ごとの安全性を平準化できることを示す．これらのうちのいずれの手法も，容量拡張等などの構成変更時に，それ以前に定めた複製先への複製済みデータを考慮して新しく複製先を決定する手法への言及はない．

文献 [9], [10] では，大規模破壊兵器（WMD）による攻撃などの災害のリスクを考慮して，プライマリなデータや

サービスを配置する手法が提案される．これらの研究では，リスクの低いデータの格納先やサービスの実行先を決定する点が共通する．しかし，容量拡張などの構成変更を理由として，それ以前に配置したデータやサービスの配置を変更する手法への言及はない．

6. おわりに

本研究では，リスクアウェア複製を適用する分散ストレージシステムの容量拡張時にシステムの災害時の可用性を，少ないデータ複製量で向上させる複製先部分再選択方式を示した．高い可用性向上が見込まれる拠点にだけ複製先を変更することを実現する方式として，システムの災害リスクを評価し複製先を決定する過程で複製先変更拠点を選択する同時選択方式と，災害リスクの高い拠点を複製先変更拠点としてあらかじめ選択した後，システムの災害リスクを評価し複製先を決定する事前選択方式を提案した．2 方式をシミュレーション評価した結果，半数の拠点が損壊するような大災害環境下で，事前選択方式が，全拠点の複製先を再選択する場合に要するデータ複製量の 34%で，全拠点の複製先を再選択する場合と同等の可用性を得られ，実用的な範囲で効率に優れることが分かった．

今後の課題は，可用性に関するさらなる評価である．本研究では，分散ストレージシステムを構成する装置の半数を壊す災害条件で行ったが，実際の災害では様々な損壊状況がありうる．実用上より起こりやすい災害条件であるより少数が損壊した場合に，2 つの提案方式がそれぞれどのような可用性を実現するか比較し，災害に対するロバスト性を評価すべきである．

謝辞 本研究は，文部科学省の委託研究「高機能高可用性情報ストレージ基盤技術の開発」の成果の一部である．本研究の推進にあたり，プログラム作成および評価作業に協力いただいた株式会社日立超 LSI システムズの新堀氏，一ノ宮氏に心より感謝する．

参考文献

[1] Matsumoto, S., Nakamura, T. and Muraoka, H.: Risk-aware Data Replication to Massively Multi-sites against Widespread Disasters, *Rangsit Journal of Information Technology*, Vol.1, No.2, pp.22–28, July-December (2013).
[2] Matsumoto, S., Nakamura, T. and Muraoka, H.: Redundancy-based Iterative Method to Select Multiple Safe Replication Sites for Risk-aware Data Replication, *IEEJ Trans. Electrical and Electronic Engineering*, Vol.11, No.1, pp.96–102 (2016).
[3] Matsumoto, S., Nakamura, T. and Muraoka, H.: Risk-based Method for Data Redundancy Determination to Improve Replica Capacity Efficiency, *Proc. 3rd Asian Conference on Information Systems* (*ACIS 2014*), pp.529–536 (2014).
[4] Si, H. and Midorikawa, S.: New attenuation relations for peak ground acceleration and velocity considering effects of fault type and site condition, *Proc. 12th World Conference on Earthquake Engineering* (*WCEE 2000*), CD-ROM, No.532 (2000).
[5] 防災科学技術研究所：J-SHIS 地震ハザードステーション，防災科学技術研究所 (オンライン)，入手先 ⟨http://www.j-shis.bosai.go.jp⟩ (参照 2016-04-17).
[6] Lowler, E.L. and Wood, D.E.: Branch-and-Bound Methods: A survey, *Operations Research*, July/August, Vol.14, No.4, pp.699–719 (1996).
[7] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., et al.: Dynamo: Amazon's highly available key-value store, *Proc. ACM SOSP*, pp.205–220 (2007).
[8] Karger, D., Lehman, E., Leighton, T., Panigrahy, R., et al.: Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web, *Proc. 29th Annual ACM Symposium on Theory of Computing* (*STOC '97*), pp.654–663 (May 1997).
[9] Ferdousi, S., Dikbiyik, F., Habib, M.F. and Mukherjee, B.: Disaster-Aware Data-Center and Content Placement in Cloud Networks, *Proc. IEEE International Conference on Advanced Networks and Telecommunications Systems* (*ANTS*), pp.1–3 (2013).
[10] Savas, S.S., Dikbiyik, F., Habib, M.F. and Mukherjee, B.: Disaster-Aware Service Provisioning by Exploiting Multipath Routing with Manycasting in Telecom Networks, *Proc. IEEE International Conference on Advanced Networks and Telecommunications Systems* (*ANTS*), pp.1–3 (2013).

松本 慎也 (正会員)

2006 年名古屋大学大学院工学研究科機械理工学専攻修士課程修了．同年 (株) 日立製作所入社．現在，同社研究開発グループ研究員．ストレージシステムの研究開発に従事．

中村 隆喜 (正会員)

1998 年大阪大学大学院工学研究科精密科学専攻博士前期課程修了．同年 (株) 日立製作所入社．2012 年東北大学電気通信研究所准教授．ストレージシステムの研究開発に従事．博士 (情報科学)．

村岡 裕明

1981 年東北大学大学院工学研究科博士課程修了．1991 年東北大学電気通信研究所．現在，東北大学電気通信研究所教授．高密度磁気記録および情報ストレージ工学の研究に従事．工学博士．