

修士論文

広域医療連携システムにおける
アクセス時間とレプリカ保有量
の決定手法

同志社大学大学院 生命医科学研究科
医工学・医情報学専攻 医情報学コース
博士前期課程 2012年度 1005番

藤井 亮助

指導教授 廣安 知之教授

2014年1月24日

Abstract

Wide-area medical cooperation system manages vast amounts of medical data in distributed sites and enables many users in each site to share their data. However, it may take a long time to retrieve a big data via a wide network. In such a distributed system, data replication is widely employed to solve this problem. There is a trade-off relationship between access time and volume of replicas, and their required values vary depending on the target system and the system administrator. Therefore, an optimal replica placement does not exist. In this study, we propose a method of determining each value in wide-area medical cooperation system. Our method calculates an optimal relationship between access time and volume of replicas and determines each value. In order to calculate the relationship, we formulate the problem of searching for a replica placement to multi-objective optimization problem and optimize replica placements. In experiment, for investigating the effectiveness of the proposal method, we applied the method to the virtual model of wide-area medical cooperation system. As the result, it was shown that the method be able to determine each value in a more wide range.

目次

1	序論	1
2	医療データの共有とデータレプリケーション	2
2.1	医療データの共有	2
2.2	データレプリケーション	4
3	最適化問題と進化計算アルゴリズム	6
3.1	最適化問題	6
3.2	GA : Genetic Algorithm	6
3.3	MOEA/D : A Multiobjective Evolutionary Algorithm based on Decomposition	7
4	提案手法	9
4.1	提案手法の概要	9
4.2	性能とコストの関係の推定	9
5	実験 -広域医療連携システムに対する提案手法の適用-	11
5.1	実験内容	11
5.2	実験結果	12
6	結論	14

1 序論

近年、患者の症状に応じた的確な診察・治療を継続的に受けられるように医療連携が求められている。医療連携とは、医療機関ごとの機能分担や専門化を進め、各医療機関が有する特性を補い合うことである。また、主治医以外の複数の医師から検査や治療に関する診断を受けるセカンドオピニオンの促進にも期待されている。円滑な医療連携を実現するにあたり、複数の医療機関における医療データの共有が必要となる。

そこで、医療データの共有を可能にする医療連携システムに関する取り組みが行われている。現在、地域的な範囲での医療連携システムとして K-MIX：かがわ遠隔医療ネットワークがある¹⁾。K-MIX は、地理的に離れたサイトにいる各医療機関から生成される医療データを外部のデータセンターで一元的に管理し、各医療機関はデータセンターを通して医療データを利用する。しかし、地域ごとに孤立したデータセンターでは地域を超えて移動する利用者には対応できない。より広域な範囲で連携を行うために医療機関をさらに拡大した場合、データの一元的な管理手法は限界があると考えられる。広域医療連携を実現する方法として、各医療機関ごとに医療データを管理し必要に応じて外部のデータへアクセスする分散管理による共有システムが提案されている²⁾。しかし、分散型の共有システムは広域範囲で多数のユーザが膨大なデータを共有できる一方、広域ネットワークを介したデータ取得時に大幅な遅延が伴う。

分散型共有システムでは、この問題を解決するために、データのレプリカを別のサイトに配置するデータレプリケーションが有効的な方法として使われている³⁾。データレプリケーションにおいて、アクセス時間が最良となるレプリカ配置は全データのレプリカを全サイトへ配置する事である。しかし、アクセス時間とトレードオフにあるレプリカ保有量も考慮しなければならないため、アクセス時間の許容とレプリカ保有量の削減が必要となる。各項目をどの程度の値にするかという決定問題において、アクセス時間とレプリカ保有量の対応関係が不明確であるため、各項目について要求する値を適切に決定できない。従来の手法は対応関係が不明確の中で、要求するアクセス時間の値へレプリカ保有量を調整するものであった⁴⁾⁵⁾。

そこで本研究では、広域医療連携システムにおけるアクセス時間とレプリカ保有量の決定手法について提案する。提案手法では、過去のアクセス履歴やデータの配置からデータレプリケーションにより得られるアクセス時間とレプリカ保有量の関係を事前に推定し、システムの管理者がその関係から要求する値を決定する。この関係の推定において、レプリカ配置の組み合わせ問題をいわゆる多目的最適化問題としてとらえ、アクセス時間とレプリカ保有量を目的関数として最適化を行う。

本論文では、第2章で医療データの共有とデータレプリケーション、第3章では提案手法で用いる最適化問題と進化計算アルゴリズム、第4章では提案する手法、第5章では広域医療連携システムを対象とした提案手法の適用実験について述べる。

2 医療データの共有とデータレプリケーション

2.1 医療データの共有

2.1.1 医療連携

現在、患者一人ひとりが自身の症状に合った医療機関で適切な医療を受けられる事を目的として、各医療機関は明確な役割と機能により分類されている。病床数が19以下の医療機関を診療所、また20以上の医療機関を病院と呼び、さらに病院を一般病院、地域医療支援病院、および特定機能病院に分類することができる。各医療機関の役割を以下に示す。

- 診療所
初期診察などの1次医療を行う医療機関である。
- 一般病院
一般的には入院・検査・手術までの2次医療を行う役割がある。
- 地域医療支援病院
一般病院と特性機能病院の中間に位置し、地域の診療所や一般病院からの紹介患者および救急患者の診察を主とする医療機関である。また、地域の医療従事者に向けた研究を実施するなど、地域医療の中核施設としての役割もある。
- 特定機能病院
高度な先進医療を提供する3次医療を行う医療機関である。大学病院、国立がんセンター、国立循環器病センターなどが特定機能病院に属する。

また、機能別の分類以外に疾患からも各医療機関の特徴を見る事ができる。現在、日本における医療費は、病気の種類と診療内容によって分類された診断群分類別包括制度DPC (Diagnosis Procedure Combination) と呼ばれる区分に基づいて、定額部分と出来高による部分を組み合わせて計算される。このDPCのデータは厚生労働省より公開されており、各主要病院における疾患別患者数などがわかる⁶⁾。DPCにおいて、疾患は神経系、眼科系、耳鼻科、呼吸器系、循環器系、筋骨系、皮膚系、乳房系、内分泌系、腎・尿路系、生殖器系、血液系、新生児系、小児系、外傷系、精神疾患系、およびその他の17種類に分類することができる。DPCデータから同じ地域医療支援病院や特定機能病院に分類される病院においても、どの疾患を専門として行っているかは大きく異なる事がわかる。以上より、患者は的確な診察・治療を継続的に受けられるように、自身の病気の疾患と症状の段階に応じた医療機関の選択をする事が重要であり、そのためには医療機関の連携が必要不可欠である。

2.1.2 地域医療連携システムに関する先行研究

現在、複数の医療機関による医療連携を目指し、医療データの共有を可能にする医療連携システムについて、研究および試験的な導入が進んでいる。その実用的なサービスの1つが香川県で稼働しているK-MIX：かがわ遠隔医療ネットワークである。このシステムは、各医療機関から生成される医療データを外部のデータセンターで一元的に管理し、各医療

機関はデータセンターからデータを取得する。

しかし、地域ごとに孤立したデータセンターでは地域を超えて移動する利用者には対応できない。より広域な範囲での医療連携を目指して医療機関をさらに増やした場合、データセンターのような場所に一元的に集約管理する手法は限界があると考えられる。これは、医療データの中でも特に医用画像は画像サイズが非常に大きいだけでなく、毎日膨大な枚数が追加されており、今後も医療機器やシステムの発展にともない、その情報量は爆発的に増加するためである。現在、アメリカにおいて2014年の一年間に10億以上の画像診断が行われ、それに伴い約100ペタバイトのデータが生成されると予測されている⁷⁾。

2.1.3 広域医療連携システムに関する先行研究

広域環境で複数の医療機関が医療データを共有するシステムとして、階層的なメタデータサーバを有する広域分散ファイルシステムを用いたPACS (Picture Archiving and Communication Systems) が提案されている。Fig. 1にシステムのイメージ図を示す。

このシステムは、各医療機関内のデータサーバ、およびメタデータサーバの2層のサーバから構成される。

- データサーバ
データサーバは各医療機関内に設置されている。各医療機関から生成された医療データは、その機関内にあるデータサーバで管理される。
- メタデータサーバ
メタデータサーバは、各医療機関と接続され、各医療機関内で管理されている医療データに関するデータの情報や所在であるメタデータのみを一元的に管理する。

医療機関のデータサーバに医療データが保存された時、医療データのメタデータが抽出され、メタデータサーバに保存される。またクライアントが医療データを取得する際、メタデータサーバにアクセスを行い目的ファイルの保存先情報を得る。その後、目的ファイルを保持している外部の医療機関のデータサーバとクライアントマシン間の通信により、ファイルを取得する。

同様に広域な医療連携を目指した実用的なサービスとしてドルフィンプロジェクトがある⁸⁾。各地域内ごとに複数の医療機関の医療データを管理するデータセンターを設置し、各医療機関とその地域のデータセンターを専用の回線経路にて結ぶ。さらに各地域に保有されているデータのメタデータをメタデータサーバにて一元的に管理を行い、メタデータサーバと地域のデータセンターをインターネットを介して結び、地域を超えたデータの利用を想定している。現在、京都府内の医療機関を連携するまいこネット⁹⁾や東京都医師会による医療連携事業HOTプロジェクト¹⁰⁾など、地域範囲での連携サービスの運用に留まっているが、今後は日本全国的な範囲での医療データの共有に向けて取り組まれている。

2.2 データレプリケーション

2.2.1 分散システム

広域な範囲で多数のユーザによる膨大なデータの共有を可能にするシステムとして、膨大なデータを広域に点在するサイトで分散して管理する分散システムが広く用いられている¹¹⁾。各サイトは地理的に離れた別のサイトとネットワークで接続され、各サイトにいるユーザは遠方のサイトで管理しているデータを広域なネットワークを通して取得することができる。分散システムの応用分野には、データインテンシブアプリケーションと呼ばれる大容量データの入力を必要とするアプリケーションの実行基盤であるデータグリッドや、不特定多数のユーザ間でのファイル共有を目的とした P2P ファイル共有システムなど、様々である。

分散システムでは、データが広域に分散するという特性を持つため、様々な課題が生じる。これらの課題は、対象の応用システムの特性によって大小は異なるが、一般的に以下の5つが挙げられる。

- 性能 (Performance)
- 拡張性 (Scalability)
- 一貫性 (Consistency)
- 信頼性 (Reliability)
- 安全性 (Security)

例えば、データグリッドでは、データインテンシブアプリケーションの実行時間を削減する事を目的としており、アプリケーションが要求するデータが遠方のサイトにある場合、実行時間に大きな影響を与える。このため、必要なデータを短時間で取得する事ができるように、性能の向上が最重要課題である。一方、P2P ファイル共有システムでは、不特定多数のノードが参加・離脱を頻繁に繰り返す特性を持つため、信頼性や拡張性の向上が重要な課題である。このように分散システムでは、応用システムの特性によって課題の優先順位が異なる。広域医療連携システムでは頻繁な機関の参加・離脱は考えられず、本研究ではファイルへのアクセス性能を解決するためにデータレプリケーションを使用する。

2.2.2 データレプリケーションの概要

データレプリケーションとは、データのレプリカを別のサイトに配置することであり、分散システムの課題を解決するための有効的な手法として広く用いられている¹²⁾¹³⁾¹⁴⁾¹⁵⁾。データレプリケーションの概要図を Fig. 2 に示す。

この例では、全てのサイトからファイルへのアクセス頻度は均一であるとする。レプリカを配置しない場合、サイト 0 を除く全てのサイトはネットワークを介してデータを取得しなければならない。このため、データへのアクセス時間に大幅な遅延が生じる。一方、レプリカを全てのサイトへ配置する場合、全てのサイトはローカルからファイルを取得する

事ができ、アクセス時間は最小となる。しかし、レプリカを保有するためのストレージ容量が最大になる。よって、実問題ではファイルの数は膨大であるため、できる限りレプリカの配置量を削減したレプリカの配置を行わなければならない。

レプリカの配置量を削減したレプリカの配置の例がパターン A とパターン B である。パターン A では、サイト 1 や近隣であるサイト 2 からは非常に短いアクセス時間でファイルを取得できる。一方、サイト 3, 4, 5 は、サイト 0 から遠方のサイトであるため、システム全体としての平均アクセス時間は大きくなり、決して良好なレプリカの配置とは言えない。よって、パターン B のような全てのサイトから短いアクセス時間でファイルを取得できるレプリカの配置を行わなければならない。

一般的にサイトによってファイルへのアクセス頻度が異なるため、アクセスの距離のみを考慮したレプリカの配置が常に良好とは限らない。そのため、アクセスの距離と頻度の 2 つの項目を考える必要があり、最適なレプリカ配置の組み合わせは非常に難しい問題となる。

3 最適化問題と進化計算アルゴリズム

3.1 最適化問題

最適化問題とは、与えられた制約条件の基で最良にする設計変数を求める問題であり、一般的に式 (3.1) および式 (3.2) のように表される。制約条件を満たす実行可能領域 F の中で目的関数 $f(x)$ を最小あるいは最大化する設計変数 x を求め、またそのような変数 x を最適解と呼ぶ。

$$\min(\max) f(x) \tag{3.1}$$

$$\text{subject to } x \in F \tag{3.2}$$

最適化問題の中でも目的関数が1つである問題を単一目的最適化問題、複数存在する問題を多目的最適化問題と呼ぶ。つまり、多目的最適化問題は、複数個の目的関数を与えられた制約条件の下で最小あるいは最大化する問題である。目的関数が複数存在し、かつそれらが互いに競合し合う時、全ての目的関数の値を最良とするような最適解を求めることはできない。そのため、多目的最適化ではパレート最適解 POS (Pareto-optimal solution) と呼ばれるある目的関数の値を改善するためには少なくとも他の1つの目的関数値を改悪せざるを得ない解の集合を求める事が目標となる。

本稿では、単一目的最適化問題の最適解を求める手法として GA (Genetic Algorithm)、また多目的最適化問題の POS を求める手法として MOEA/D (Multi Objective Evolutionary Algorithm based on Decomposition) について述べる。

3.2 GA : Genetic Algorithm

GA は生物が環境に適応して進化していく過程を工学的に模倣した確率的な最適化手法である¹⁶⁾。自然界における生物の進化過程においては、ある世代を形成している個体集団の中で環境に適応した個体がより高い確率で生き残る。そして、生き残った個体が次世代に子を残す。この生物進化のメカニズムをモデル化し、環境に対して最もよく適応した個体、すなわち目的関数に対して最適値を与えるような解を求めることが GA の概念である。

GA では1つの解候補を1個体として扱い、個体の集団を用いて解探索を行う。各個体は目的関数値に応じた適合度を有し、ある世代を形成している個体集団において、適合度の高い個体ほど高確率で生き残るように選択を行う。加えて、交叉および突然変異といった個体生成の遺伝的操作によって子個体を生成し、次世代の個体集団を形成する。この世代更新の繰返しによって適合度の高い個体が集団内に増加し、最適解に集団を収束させるのが GA のメカニズムである。GA の流れを以下に示す。

Step1:初期化 初期母集団を形成する複数の個体をランダムに生成し、各個体の適合度を評価する。

Step2:終了判定 あらかじめ定められた終了条件に基づいて、GA の処理を終了する。この

時の母集団で適合度の最も高い個体を最適解として採用する。一般的には、世代数による終了条件が使用される。

Step3:評価 各個体に環境に合わせた適合度、すなわち目的関数値を計算する。

Step4:選択 次世代の子を生成するための親個体候補を選択する。

Step5:交叉 親個体 A の遺伝子と親個体 B の遺伝子を入れ替えることにより新しい子個体を生成する。

Step6:突然変異 個体内のある遺伝子を突然変異率に従って他の対立遺伝子に置き換える。

Step7:生存選択 交叉、突然変異によって生成された子個体から、次世代に残る個体を選択する。本 Step が終了後、Step2 に戻る。

3.3 MOEA/D: A Multiobjective Evolutionary Algorithm based on Decomposition

MOEA/D は、2007 年に Qingfu よって提案された多目的進化型計算のフレームワークである¹⁷⁾。MOEA/D の特徴は、多目的最適化問題を複数の単一目的最適化問題として扱うことである。各個体にそれぞれ重みの異なるベクトルを与え、参照点と重みベクトルを計算に含んだスカラー化関数を用いることで、単一目的最適化問題として個体の評価値を計算していく。また、重みベクトルを用いて近傍を定義し、近傍内で交叉・突然変異といった遺伝的処理と近傍内での優れた設計変数の更新を行うことによって探索を進めていく。MOEA/D の流れを以下に示す。

Step1:初期化 個体数 N の母集団 P を生成

Step1.1 各個体 $P(i)$ に対して異なる重みベクトル λ_i を与え、与えられた重みベクトル間のユークリッド距離を用いて各個体の近傍 $B(i) = i_1 \dots i_T$ を設定する。近傍サイズ T は任意に設定する。

Step1.2 各個体 $P(i)$ の各設計変数 x_j に対してランダム値を設定する。

Step1.3 参照点 z に対して母集団内の各目的関数における最良値を与える。

Step1.4 各個体 $P(i)$ に対して各目的関数 $f(x)$ を算出し、参照点 z ・重みベクトル λ ・各目的関数 $f(x)$ からスカラー化関数 $g^{te}(x|\lambda^j, z)$ を計算する。

Step2:更新 For $i = 1, \dots, N$ do

Step2.1:再製 各個体に対して、近傍 $B(i)$ からランダムに 2 個体を選択し、遺伝的処理を用いて新たな解 y を生成する。

Step2.2:修正 新たに生成した解 y が制約条件から外れている場合、任意の方法で制約条件内の解 y' を生成する。

Step2.3:参照点の更新 参照点 z を新たに生成した解 y' から得られる $f(y')$ を用いて更新する.

Step2.4:近傍解の更新 近傍 $B(i)$ の各個体に対して新たに生成した y' から得られる $g^{te}(y' | \lambda^j, z^*)$ の更新を行う.

Step3:終了条件 事前に設定した条件 (世代数など) に該当する場合終了し, それ以外ならば Step.2 に戻る.

4 提案手法

4.1 提案手法の概要

本稿では、性能を表す指標は、ファイルを取得するためにかかる平均アクセス時間 AAT (Average Access Time) とする。またコストを表す指標は、システムが保有する全レプリカの容量 TRV (Total Replica Volume) とする。広域医療連携システムにおけるシステム管理者に要求される AAT と TRV の値を決定することによって、最適なデータレプリケーションが実行できる。そこで、広域医療連携システムにおいて要求される AAT と TRV の決定手法について提案する。提案手法の概要図を Fig. 3 に示し、また流れを以下に示す。

Step1:推定情報の収集 各サイトからメタデータセンターへ以下の情報を集める。

- システム全体のネットワーク構成と速度
- サイト i におけるストレージの空き容量 ASV_f (Available Storage Volume)
- ファイル f に関する配置場所とファイルサイズ V_f (Volume)
- 過去の期間に行われたアクセスの履歴

Step2:関係の推定 収集された情報を用いて、レプリカの配置により変動する AAT と TRV のトレードオフ関係を推定する。推定において、同値な TRV を持つレプリカの配置においても、その配置の組み合わせによって AAT は異なるという問題がある。AAT と TRV を正確に求めるためには、AAT と TRV を互いに最適にするレプリカの配置の組み合わせを探索する必要がある。本手法では、レプリカ配置の組み合わせ問題を多目的最適化問題へ定式化し、解くことによりこの問題を解決する。

Step3:要求する値の決定 AAT と TRV のトレードオフ関係をシステムの管理者に提示する。システムの管理者は、この関係図から対象のシステムにおいて要求する AAT と TRV の値を決定する。

Step4:システムへの適用 一意に決定した AAT と TRV の値を満たすレプリカの配置を対象のシステムに適用する。

4.2 性能とコストの関係の推定

AAT と TRV が互いに最適となる関係を推定するために、レプリカ配置の組み合わせ問題を多目的最適化問題へ定式化し、解く。

設計変数はレプリカの配置の組み合わせとする。サイトが $i = 1, 2, \dots, n$ まであり、ファイルが $f = 1, 2, \dots, m$ まである時、設計変数は $n \times m$ 分の式 (4.1) を持った 2 ビットの配列として表現できる。 $W_{f,i} = 1$ はファイル f のレプリカをサイト i へ配置する事を表し、また $W_{f,i} = 0$ はレプリカを配置しない事を表す。

$$W_{f,i} \in \{0, 1\} \quad (4.1)$$

目的関数は、AAT と TRV の2つとする。TRV はシステム内にある全てのレプリカの全容量を意味し、式 (4.2) に示す通りに表される。

$$TRV = \sum_{f=0}^m \sum_{i=0}^n V_f \times W_{f,i} \quad (4.2)$$

AAT は、過去の期間において設計変数のレプリカ配置表現に従ってレプリケーションを行った場合に、1 ファイルを取得するのにかかる平均アクセス時間を意味する。AAT は、式 (4.3) に示す通りに表される。AC (Access Count) は、過去の期間において各サイト i からファイル f へアクセスした回数を表す。式 (4.3) は、アクセス回数が多く、かつアクセスまでの距離が遠いサイトにしかないファイルを優先的にレプリケーションする事を意味する。

$$AAT = \frac{\sum_{f=0}^m \sum_{i=0}^n AC_{f,i} \times AT_{f,i,j}}{\sum_{f=0}^m \sum_{i=0}^n AC_{f,i}} \quad (4.3)$$

$AT_{f,i,j}$ (Access Time) は、ネットワーク速度 $BW_{i,j}$ (BandWidth) を持つサイト $i-j$ 間を介して、ファイル f を転送するのに要する時間を表し、式 (4.4) に示す通りに計算される。

$$AT = \min\{BW_{i,j} \times V_f\} \quad (4.4)$$

さらに各サイトで保有できるレプリカの容量に限界があるため、制約条件として式 (4.5) を加える。

$$\sum_{f=0}^n V_f \times W_{f,i} \leq ASV_i \quad (4.5)$$

各サイト i のストレージの空き容量 ASV_i (Available Storage Volume) が、そのサイトへ配置される全レプリカの容量を超える事を意味する。以上を踏まえ、レプリカ配置の組み合わせ問題は式 (4.6)、式 (4.7)、および式 (4.8) に示す通りに定式化できる。

$$\min TRV \quad (4.6)$$

$$\min AAT \quad (4.7)$$

$$\text{subject to } \sum_{f=0}^n V_f \times W_{f,i} \leq ASV_i \quad (4.8)$$

この定式を多目的最適化手法を用いて、TRV と AAT が共に最小になる POS を見つけ出す。

5 実験 -広域医療連携システムに対する提案手法の適用-

5.1 実験内容

本実験は、広域医療連携システムに対する提案手法の有効性の検証を目的とする。有効性を検証するために、広域医療システムのシナリオを想定したモデルに対して提案手法を適用し、アクセス時間とレプリカ保有量を適切に決定できる事を確認する。また、広域医療連携システムのモデルの構築として、想定されるシナリオから広域医療連携システムの要件を定義し、要件を満たすモデルを設計する。本実験では、多目的最適化問題を解く多目的最適化手法としてMOEA/D (A Multiobjective Evolutionary Algorithm based on Decomposition)¹⁷⁾、またシミュレーションソフトとしてOptorSim¹⁸⁾を用いた。

5.1.1 広域医療連携システムのシナリオ

広域医療連携システムは以下の流れで行われる事を想定している。

Step1:初期状態 広域医療連携システムに属する全ての医療機関は、互いにネットワークで接続されており、機関間のデータの通信を可能にする。

Step2:医療データの生成と保有 各医療機関から疾患種類の医療データが生成される。これらの医療データは、その機関内のデータサーバで管理される。

Step3:ファイルアクセス 各機関に在するユーザは、必要に応じて外部の機関で管理されている医療データへネットワークを介してアクセスする。

本稿で定義する要件の項目は、各医療機関が接続されるネットワークに関するトポロジ、および各機関から実行されるファイルアクセスに関するパターンの2つである。トポロジに関する要件として、関連・系列にある医療機関同士は近隣に接続されるものとした。この設定の理由は、各大学附属病院には関連・系列にある病院が存在し、その関連・系列にある病院間が専用のネットワークで接続される事が考えられるためである。

アクセスパターンに関する要件として、各医療機関から実行されるアクセス回数はその機関の患者数に比例し、またそのアクセス先のファイルの疾患種類はその機関の疾患比率に比例するものとした。例えば、循環器系を中心に診断・治療を行っている医療機関は、循環器系のファイルについて外部の機関へ多くアクセスする。この設定の理由は、患者の総数が多いほどその機関へ転移する患者数も多く、また特定の疾患比率が高い機関ほどその疾患に関する診断・治療を目的として転移する患者が多いと考えられるためである。

5.1.2 ネットワークトポロジの設定

要件を満たしたトポロジを設計するために、現実の各大学附属病院の関連・系列病院の情報を使用した。使用したのは京都府立医科大学附属病院、京都大学附属病院、および大阪大学附属病院とその関連・系列にある病院である。この関連・系列病院の情報を基に、要件を満たすトポロジを Fig. 4 の通りに設計した。大学附属病院とその系列・関連のあ

る病院は近隣のネットワークで接続されている。また、関連・系列でない病院間は広域ネットワークを介して接続される。

5.1.3 アクセスパターンの設定

Fig. 4 に示されたトポロジーに属する各機関から実行されるファイルアクセスのパターンを設計する。要件を満たすアクセスパターンを設計するため、各医療機関の疾患別患者比率の情報を用いる。なお、本実験では、循環器系と消化器系の2種類の疾患のみを扱う。この2種類のみを扱う理由は、その2種類の疾患が診断・治療のためにCTなどのデータサイズの大きな医用画像データを生成する機器を使用するためである。各医療機関の循環器系と消化器系の患者比率を Fig. 5 および Fig. 6 に示す。これらの情報は厚生労働省から公開されている DPC に関するデータを用いた⁶⁾。この疾患別患者比率の情報を基に、設計したアクセスパターンを Fig. 7 に示す。例えば、サイト 12 である大阪循環病研究センターでは、循環器系比率の患者数が非常に多いため、循環器系のファイルへのアクセス回数は多く、消化器系のファイルへのアクセス回数が非常に少ない。また、サイト 10 である京都南病院では、両疾患比率が他の機関に比べて非常に少ないため、アクセスの総数も非常に少ない設定となっている。

5.1.4 実験パラメータ

本実験における広域医療連携システムのモデルのパラメータを Table 1、また推定に使用した MOEA/D のパラメータを Table 2 に示す。各サイトの機関のデータサーバの容量は、それぞれ 25 GB とした。各サイトのデータサーバに 1 GB のサイズを持つ消化器系と循環器系のファイルがそれぞれ1つずつ配置される。その後、ファイルアクセスが各サイトから合計 10000 回実行され、指定回数実行された後に提案手法が適用される。

5.2 実験結果

MOEA/D により得られた AAT と TRV の関係を Fig. 8 に示す。パレート最適解は探索された解の中で AAT と TRV が互いに最小とする解である。また、劣解は探索された解の中で AAT と TRV の値がパレート最適解に比べ劣る解である。パレート最適解について、TRV が 0 GB から最大レプリカ保有量である 250 GB までほぼ均等に解の探索ができていた事を確認した。

TRV が 50 GB および 100 GB である時の各サイトのレプリカ保有量を Fig. 9 に示す。サイト ID の 0, 8, 12, 13, 14, および 15 に代表される多くのサイトにおいて、TRV の増加に合わせてサイトのレプリカ保有量も増加傾向にある事がわかった。一方、サイト 1, 4, および 11 のサイトは TRV の増加に合わせてレプリカ保有量が減少しているサイトも存在する事がわかった。これによって、TRV の値に合わせたレプリカの配置を探索できている事が示唆された。

次に、レプリカ配置による各サイトへアクセスする回数の変動を調査する。サイト 0 から 5 を Group0、サイト 6 から 10 を Group1、およびサイト 11 から 15 を Group2 とする。

TRV が 0 GB, 50 GB, および 100 GB の時のパレート最適解のレプリカ配置による各 Group へのアクセス回数の変動として, サイト 0 から見た変動を Fig. 10, サイト 6 から見た変動を Fig. 11, およびサイト 11 から見た変動を Fig. 12 に示す. Fig. 10 では, TRV の増加に合わせて Group1 と Group2 へのアクセス回数が減少した. 一方, サイト 0 が所属する Group0 へのアクセス回数は増加した. Fig. 11 および Fig. 12 においても同様に, サイト 6 およびサイト 11 が自身の所属する Group へのアクセス回数は増加傾向にあり, 異なる Group へのアクセス回数は減少傾向にあった. この理由として, 今回のトポロジーの場合では同じ Group 内へレプリカを配置する事によってアクセス時間を大きく削減できる. このため, 同じ Group 内へ自身が必要とするレプリカを多く配置し, アクセス回数が増加したと考えられる. これによって, 各 TRV において良好なレプリカの配置を探索できている事が示された. 以上より, AAT と TRV の最適な関係から要求する AAT と TRV の値を適切に決定できる事が示された.

6 結論

膨大な医療データを各医療機関に分散して共有する広域医療連携システムが提案されている。このシステムにおいて、トレードオフにあるアクセス時間とレプリカ保有量を各々どの程度にするかといった問題は、対応関係が明示されていないため要求する値を適切に決定する事はできない。

本研究では、広域医療連携システムにおけるアクセス時間とレプリカ保有量の決定手法の提案を目的とする。提案手法は、アクセス時間とレプリカ保有量の関係を推定し、推定した関係から要求する値を一意に決定する。最適な関係を推定するために、アクセス時間とレプリカ保有量を目的関数とした多目的最適化問題へ定式化し、複数の最適なレプリカ配置を探索する。

実験では、提案手法の有効性を検証するために、シミュレーション上に構築した広域医療連携システムのモデルに対して提案手法を適用した。モデルのトポロジーおよびアクセスパターンとして、各大学附属病院の関連病院、および各機関の患者数と疾患比率の情報をを用いて、実際のシナリオを想定した設定を行った。実験の結果、探索された複数のレプリカ配置は、どのレプリカ保有量においてもアクセス時間を最小にする良好なレプリカ配置を探索できている事を確認し、また各項目を最適な関係から適切に決定できる事を示した。提案手法により、広域医療連携システムにおける要求するアクセス時間とレプリカ保有量の値を適切に決定する事ができる。

謝辞

本修士論文は、筆者が同志社大学大学院 生命医科学研究科 医工学・医情報学専攻博士前期課程在学中に医療情報システム研究室において行った研究をまとめたものである。私が本研究室に配属されてからの3年間、同志社大学生命医学部の廣安知之教授には多大なる御指導、そしてご協力を頂き、心より御礼を申し上げます。また、本研究に関してご指導ご鞭撻を頂きました山本詩子助教、横内久猛先生に心より感謝致します。

そして、同じ医療システム班である西村祐二さん、井上楓彩さん、岡村達也さん、林沼勝利さん、滝謙一さん、埴賢哉さん、また私のコーチである布川将来人さんには日頃から研究について多く議論して頂きました。また本論文を作成するにあたり西村祐二さん、大西夏子さんに校正して頂きました。皆様にはこの場をお借りして感謝を申し上げます。

最後に、医療情報システム研究室の皆様には私の研究に関して数多くの議論や助言をして頂きました。皆様のおかげで、3年間すばらしい研究生活を送ることができました。この場を借りて厚く御礼申し上げます。

参考文献

- 1) 原量宏, 横位英人. 病院情報システムと遠隔医療かがわ遠隔医療ネットワークから日本版 EHR の実現へ. 医療機器システム白書 2008~2009, pp. 358–360, 2008.
- 2) H. Tomoyuki, M. Yoshiyuki, Y. Masato, and M. Mitsunori. Distributed pacs using network shared file system. *PDPTA*, Vol. 2, pp. 745–750, 2011.
- 3) T. Amjad, M. Sher, and A. Daud. A survey of dynamic replication strategies for improving data availability in data grids. *Future Generation Computer Systems*, Vol. 28, No. 2, pp. 337–349, 2012.
- 4) X. Sun, J. Zheng, Q. Liu, and Y. Liu. Dynamic data replication based on access cost in distributed systems. In *Computer Sciences and Convergence Information Technology, 2009. ICCIT '09. Fourth International Conference on*, pp. 829–834, 2009.
- 5) R. Alhajj R.M. Ken, K. Barker. Study of different replica placement and maintenance strategies in data grid. *Cluster Computing and the Grid*, pp. 171–178, 2007.
- 6) 第 5 回診療報酬調査専門組織・D P C 評価分科会.
<http://edg-wp2.web.cern.ch/edgwp2/optimization/optorsim.html>.
- 7) C. Walker A. Swan C. Davila D. Howard C.-C. Teng, J. Mitchell and T. Needham. A medical image archive solution in the cloud. *Software Engineering and Service Sciences (ICSESS), 2010 IEEE International Conference on*, 2010.
- 8) T. Akira et al. Dolphin project—cooperative regional clinical system centered on clinical information center. *Journal of Medical Systems*, Vol. 29, pp. 391–400, 2012.
- 9) まいこネット. <http://www.e-maiko.net/>.
- 10) 東京都医師会地域医療連携事業 HOT プロジェクト.
<http://www.ocean.shinagawa.tokyo.jp/hot/>.
- 11) T. Osamu, H. Kohei and S. Noriyuki. Gfarm grid file system. *New Generation Computing*, Vol. 28, pp. 257–275, 2010.
- 12) I. Foster K. Ranganathan. Design and evaluation of dynamic replication strategies for a high performance data grid. *International Conference on Computing in High Energy and Nuclear Physics*, Vol. 3, No. 4, 2001.
- 13) T. N. Shishavanc L. M. Khanli, A. Isazadeh. Phfs:a dynamic replication method to decrease access latency in multi-tier data grid. *Future Generation Computer Systems*, 2010.

- 14) A. S. Thanamani K. Sashi. Dynamic replication in a data grid using a modified bhr region based algorithm. *Future Generation Computer Systems*, Vol. 27, pp. 202–210, 2011.
- 15) Y. Chen M. Leea, F. Leub. Pfrf: An adaptive data replication algorithm based on star-topology data grids. *Future Generation Computer Systems*, Vol. 28, pp. 1045–1057, 2012.
- 16) David E. Goldberg. Genetic algorithms in search. *Optimization, and Machine Learning*. Addison-Wesley, 1989.
- 17) Q. Zhang and H. Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *Evolutionary Computation, IEEE Transactions on*, Vol. 11, No. 6, pp. 712–731, 2007.
- 18) OptorSim. <http://edg-wp2.web.cern.ch/edgwp2/optimization/optorsim.html>.

付 図

1	広域医療連携システムのイメージ図	1
2	データレプリケーションの概要図	1
3	提案手法の概要図	2
4	広域医療連携システムのネットワークポロジ	2
5	循環器系の患者比率	3
6	消化器系の患者比率	3
7	各サイトの疾患別アクセス履歴	4
8	広域医療連携システムにおけるレプリカ配置のパレート最適解	5
9	各サイトのレプリカ保有量	5
10	サイト 0 から各 Group へのアクセス回数の変動	6
11	サイト 6 から各 Group へのアクセス回数の変動	6
12	サイト 11 から各 Group へのアクセス回数の変動	7

付 表

1	広域医療連携システムのモデルのパラメータ	4
2	MOEA/D のパラメータ	4

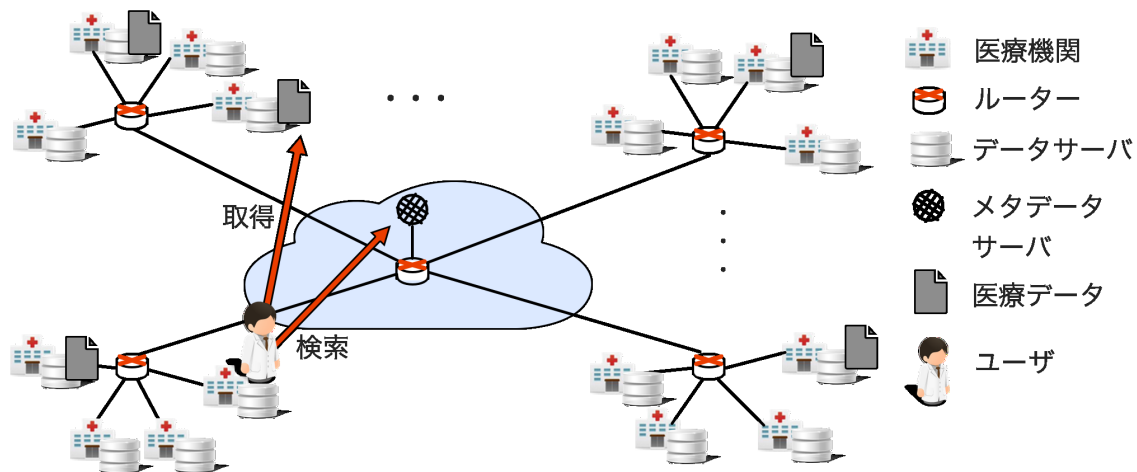


Fig. 1 広域医療連携システムのイメージ図

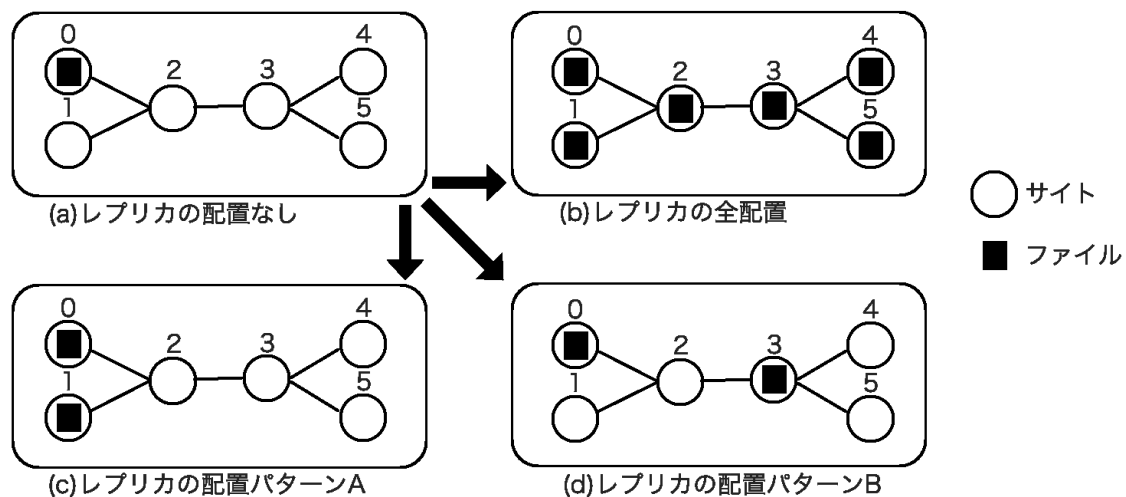


Fig. 2 データレプリケーションの概要図

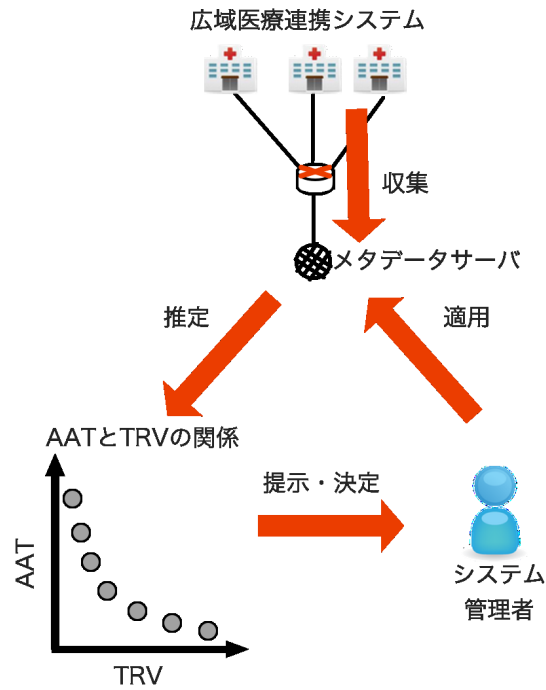


Fig. 3 提案手法の概要図

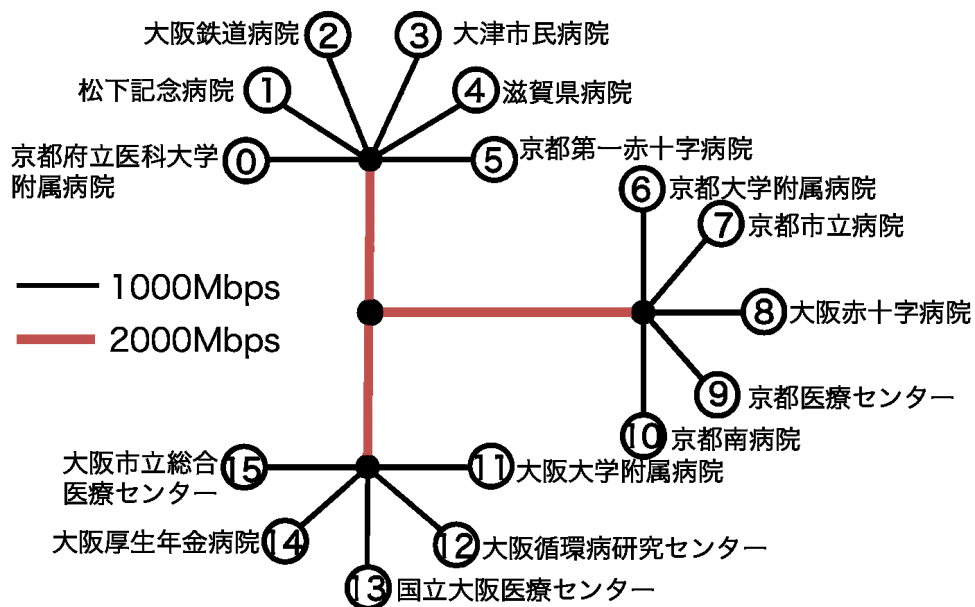


Fig. 4 広域医療連携システムのネットワークポロジ

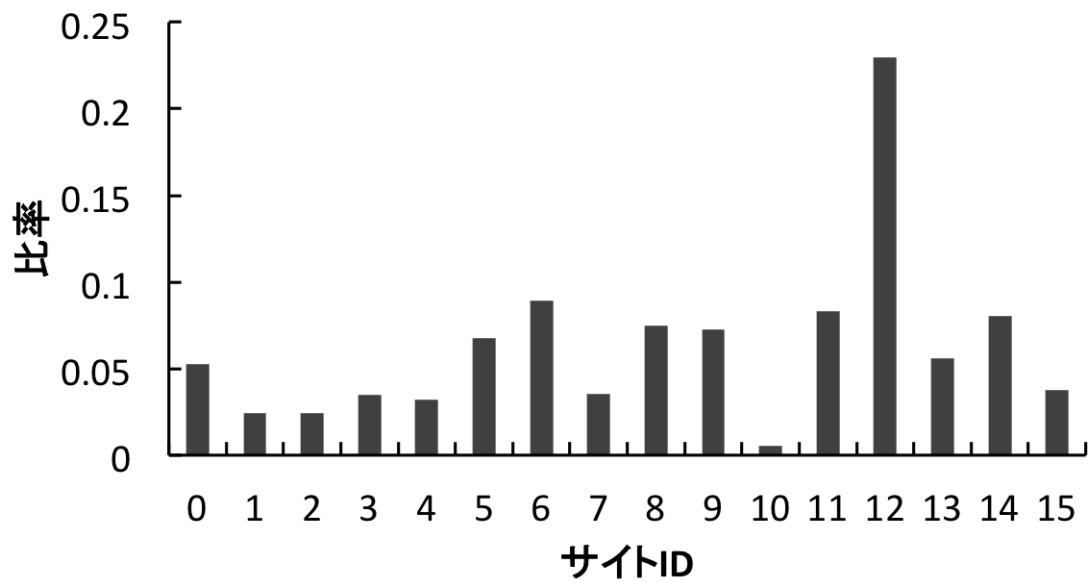


Fig. 5 循環器系の患者比率

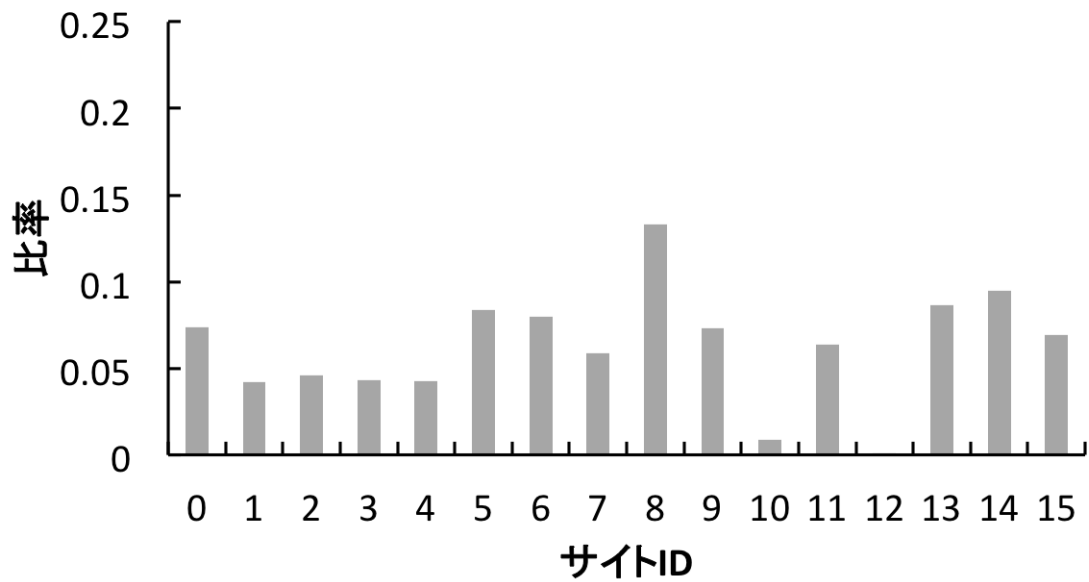


Fig. 6 消化器系の患者比率

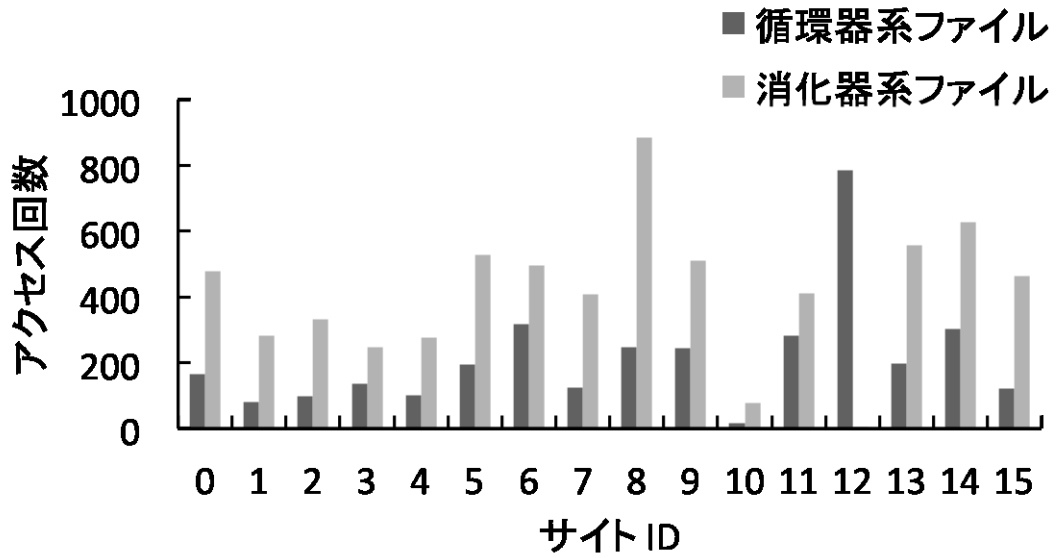


Fig. 7 各サイトの疾患別アクセス履歴

Table 1 広域医療連携システムのモデルのパラメータ

パラメータ	値
ファイル数	32
ファイルサイズ V	1 GB
各サイトのストレージ空き容量 ASV	25 GB
ファイルリクエスト回数	10000 回
サイト数	16

Table 2 MOEA/D のパラメータ

パラメータ	値
世代数	500
遺伝子数	512
個体数	1000
突然変異率	0.002
交叉率	1.0
交叉方法	2点交叉
選択方法	トーナメント選択
スカラー化関数	WeightedSum
近傍個体数	2

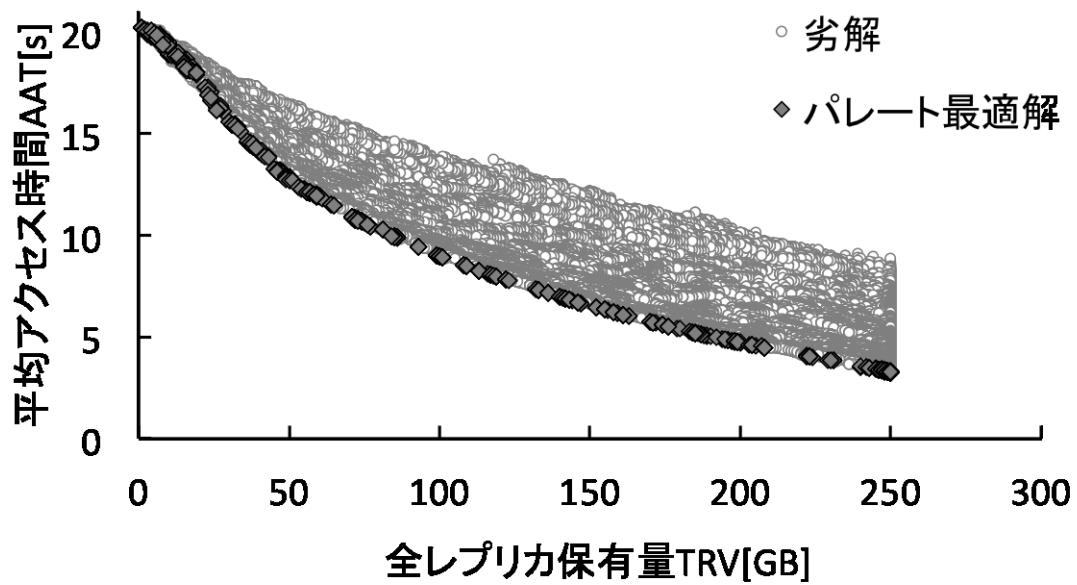


Fig. 8 広域医療連携システムにおけるレプリカ配置のパレート最適解

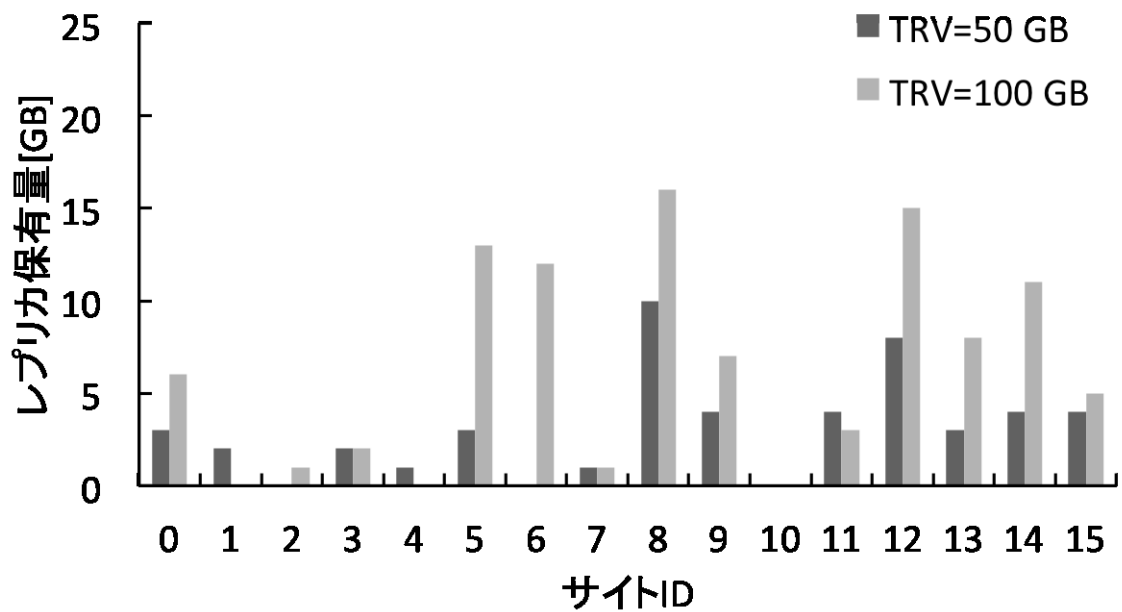


Fig. 9 各サイトのレプリカ保有量

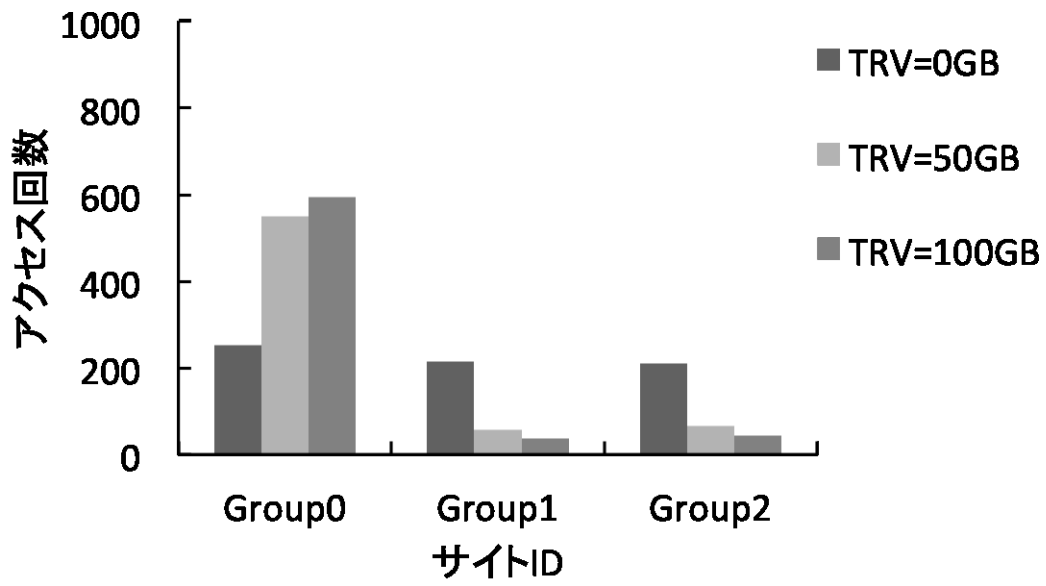


Fig. 10 サイト0から各 Group へのアクセス回数の変動

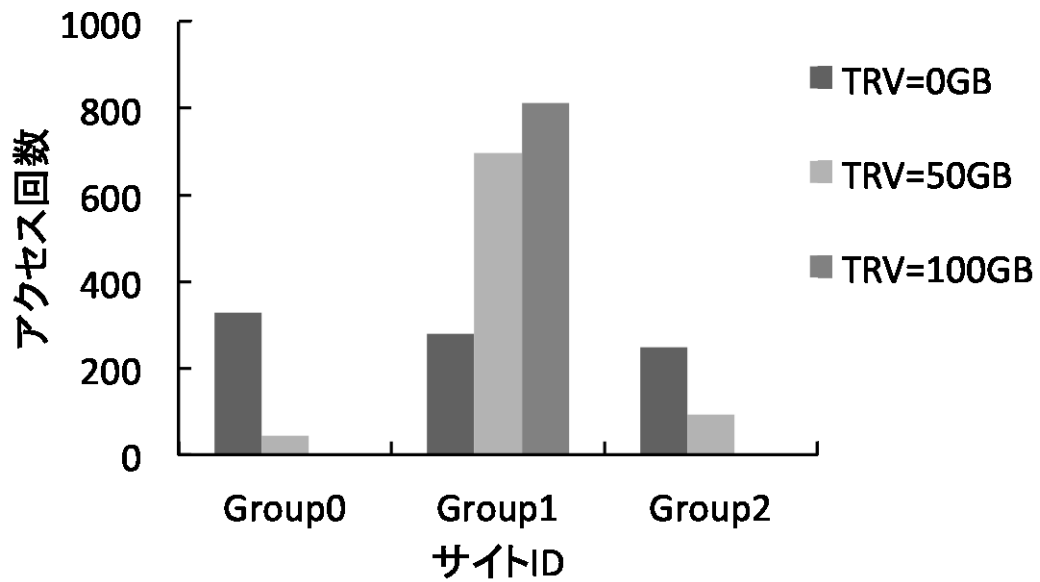


Fig. 11 サイト6から各 Group へのアクセス回数の変動

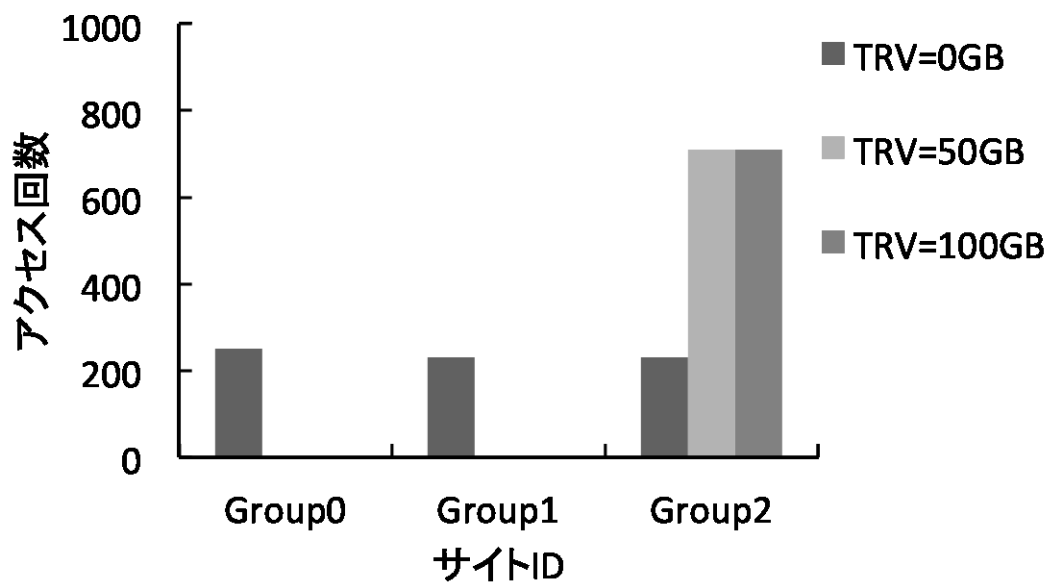


Fig. 12 サイト 11 から各 Group へのアクセス回数の変動