Vol.2011-OS-116 No.2
2011/1/24

狭帯域バックアップ回線を用いたネットワーク分断時の可用性向上

宮澤和徳[†1] 石井嘉明[†2] 廣岡誠之[†2]
品川高廣[†1] 加藤和彦[†1]

近年、インターネットサービスの可用性に対する要求が高まっている。我々は、ネットワーク分断時にも可用性を維持する分散オブジェクトライブラリの研究を行っているが、完全なネットワーク分断時には一貫性維持のためサーバで可能な操作が著しく制限され、通常時に比べると依然として可用性が大きく低下してしまう。本研究では、ネットワーク分断時にも利用可能な狭帯域のバックアップ回線の存在を前提とし、一貫性を維持しつつ少ない通信量で可用性を向上させるための手法について述べる。

Improving Availability with Spare Narrowband in Network-Partiotioning

Kazunori Miyazawa ,[†1] Yoshiaki Ishii ,[†2]
Nobuyuki Hirooka ,[†2] Takahiro Shinagawa [†1]
and Kazuhiko Kato[†1]

Recently, the demand for the availability of internet services is increasing. We had researched distributed object library to keep the application available and consistent when network is partitioned. But it is difficult to keeping consistency with no limit to availability in that case.In this papar, we propose a technique of improving availability and keeping consistency with using spare narrowband when network is partitioned.

†1 筑波大学大学院 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba
†2 富士ソフト株式会社
FUJISOFT Incorporated

1. はじめに

近年，インターネットサービスのインターネット経由で提供されるサービスの高機能化が著しい．例えば，クラウドコンピューティングで提供されるサービスの中には，電子メールやグループウェアなど，企業活動にとってなくてはならないサービスも提供されている．このようなサービスでは，サービスの可用性を維持することは最も重視すべき点の一つである．しかしそれらのサービスはインターネット経由で提供されるため，専用線の利用時と比べてネットワークの信頼性が低く，ネットワークの分断などでサービスの可用性が低下する問題がある．サービスの可用性を維持するアプローチの一つとしてサービスの分散提供が挙げられるが，分散型のサービスにおいては，そのサービスの可用性と一貫性の両方を、集中型のサービスと同程度のレベルに保つことが困難であるとされている．これは CAP 定理[1)] と呼ばれている．しかし，可用性と一貫性の両方を満たそうとせずどちらかを緩める、すなわち可用性もしくは一貫性のどちらかのレベルを集中型のサービスのそれよりも下げることで，ある程度の両立が可能になることが知られている．

本研究室では，可用性を緩めて一貫性を維持する手法で，ネットワーク分断にも対応可能な一貫性制御を行うライブラリ[6)] を研究している．このライブラリでは，ネットワークが分断されていない通常状態においては同期通信を行うことで各サーバ間の一貫性を維持し，ネットワークが分断されている状態においては，サーバごとに操作可能なデータの範囲に制限をもうけ，分断回復後に行われるデータのマージ処理を定義することで，可用性と一貫性を両立している．このライブラリにおいては，想定するネットワークの状態が通常状態とネットワーク分断状態の 2 種類であるが，本研究においては，衛星通信などを用いた回線が提供されており，バックアップ回線として利用可能な状態であることも想定する．ネットワーク分断が発生した場合でも，バックアップ回線を用いてサーバ間の通信を行うことで，分断時と比較して可用性をあげることが可能となる．しかし，衛星通信などは利用コストが一般的なネットワークと比較して高いため，なるべくコストを抑えるためにも利用する帯域は少ない方が良いと考える．

そこで本研究では，ネットワーク分断時にも利用可能な狭帯域のバックアップ回線の存在を前提とし，一貫性を維持しつつ少ない通信量で可用性を向上させるための手法を提案する．先行研究である分散オブジェクトライブラリを拡張し，狭帯域回線を利用して分断時に各サーバにかかる操作範囲の制限を緩和するための通信を行うことで，分断時と比較して一貫性を維持しつつ可用性を向上させることを目標とする．

本論文では，RUBiS というオークションサービスをベースにして，実際に提案した手法の実装を行った．この実験を行った結果，分断時と比較して狭帯域時には可用性が向上することを確認した．また通常時と比較して，狭帯域時にはサーバ間の通信量を削減できていることを確認した．

本論文の構成は以下の通りである．まず，第 2 章で関連研究と本研究との違いを述べる．続いて第 3 章で本提案において想定するネットワークの状態とその変化への対応について述べる．次に第 4 章で本提案の概要と先行研究である分散オブジェクトライブラリの概要を示し，本提案における拡張部分を示す．その後第 5 章で RUBiS に本提案手法を実装して実験を行った結果と考察を示し，最後に本論文のまとめを第 6 章に示す．

2. 関 連 研 究

CAP 定理で示された問題への対処手法として，(1) ネットワーク分断から回復した後のマージ処理を行うことで一貫性を回復する手法と，(2) 分散オブジェクトにより操作に一定の制約を設けることで一貫性を維持する手法が挙げられる．また，狭帯域な回線を用いた通信に関する手法もいくつか提案されており，それらの手法と本研究との違いについて述べる．

2.1 マージ処理による一貫性の回復

Dynamo[2] では，分断回復後のマージ処理を vector clock を用いて行っている．この手法では，更新操作が衝突しない限り最新の操作がどれかを判定でき，その操作を自動的に適用することができる．しかし同一のデータに対する更新操作が衝突した場合のマージ処理は，アプリケーションの開発者が記述する必要がある．

Harmony[3] は一貫性制御用のフレームワークであり，全てのデータを木構造に変換しマージすることで，統一された一貫性制御を行う．しかしこの手法においては，異なるユーザによる同一データへの操作は衝突とは見なされず，操作対象のデータを複製して各ユーザ固有のデータとして扱うことで衝突を回避しつつマージを行うため，適用するアプリケーションが限定されてしまう．

2.2 分散オブジェクトによる一貫性の維持

GlobeDB[4] では，一貫性制御を組み込んだ汎用的なデータベースのドライバーにより自動的に一貫性制御を行うことができるが，primary-backup プロトコルを用いているため，ネットワーク分断中には更新操作を完了することができない．

Gao らの研究[5] では，インターネットサービスの可用性を高めるために緩い一貫性制御を組み込んだ分散オブジェクトを提案している．しかし，書籍販売サイトに特化した設計になっているため，汎用的なサービスへの適用可能性は必ずしも明らかになっていない．

小長谷らの研究[6] では，緩い一貫性制御を組み込んだ汎用的に利用可能なデータ構造群をライブラリとして提供している．アプリケーション開発者はアプリケーションのセマンティクスに応じたデータ構造を選択することで，容易に分散型のインターネットサービスが構築可能である．

2.3 狭帯域回線を用いた通信

RDC[7] は，Windows Server2008 に提供されている，WAN 経由でのファイルの複製機能である．限られた帯域を用いて複製を行うため，ファイルの差分複製をサポートしている．送信するデータを限定するというアプローチは分散オブジェクトライブラリでもとられているが，複数箇所での更新はサポートされておらず，更新の衝突に対応するためには別にマージ処理を定義する必要がある．

Yui-Wah らの研究[8] では，データ自体をやりとりするのではなく，操作命令などのオペレーションを送信することで，帯域を節約する．オペレーションを送信することで帯域を節約するアプローチは本研究と似ているが，この研究における目的は更新の伝搬であり，本研究で提案するような可用性の向上のために用いる通信とは目的が異なる．

3. 分散オブジェクトライブラリ

本章では，先行研究である分散オブジェクトライブラリの概要と，内部で定義されたデータ構造について述べる．また，各データ構造内で利用されている一貫性制御手法，想定するネットワークの状態及び対応手法についても述べる．

3.1 分散オブジェクトライブラリの概要

分散オブジェクトライブラリは，開発者が分散型のインターネットサービスを構築する際に，一貫性制御を記述する負担を軽減するためのライブラリである．書き込みが想定される分散型のインターネット・サービスにおいてはサーバ間の同期は必須であるが，このライブラリでは各サーバの操作履歴を同期通信に利用しており，各サーバは他のサーバから送信された操作履歴を用いてマージ処理を行い，同期をとる．またネットワーク分断時の一貫性制御処理が記述された汎用的なデータ構造を複数提供しており，それらのデータ構造を利用することで，自動的に一貫性制御が行われる．各データ構造内のデータに対しては緩い一貫性制御を用いられているため，ネットワーク分断における可用性と一貫性の両立がなされている．

3.2 データ構造

分散オブジェクトライブラリで定義されているデータ構造はクラス階層構造となっている．上位クラスは，抽象的なデータ構造に対して想定される操作や制約条件を定義可能な抽象クラスとなっており，下位クラスは，上位クラスを継承して実際に分散型のサービス構築に有用な具象部分となっている．上位クラスは，主にセット構造と数値構造に分類され，下位も上位に従ってその 2 種類の構造に分類される．以下では、セット構造とその内部で定義される操作について述べる．

セット構造 セット構造は，複数の要素の集合を扱うデータ構造である．厳密には数学的な集合とは異なり，同じ内容の要素の重複も許容可能である．セット構造に対しては，要素の「追加」「削除」等が考えられる．またこれを継承する下位クラスとしては，key/value 形式でデータの集合を扱う Map クラスや，内包する要素の順番を気にしないで扱う Set クラスが考えられる．

3.3 一貫性制御手法

分散オブジェクトライブラリでは，各データ構造に緩い一貫性制御を組み込むことで，可用性と一貫性をある程度両立している．ここでは、内部で定義されたいくつかの一貫性制御手法のうち，操作範囲の制限について述べる．なお，内部で定義された各手法は独立の物ではなく，それぞれの手法を組み合わせることで一貫性制御が行われている．

3.3.1 操作範囲の制限

操作の範囲について，データ構造のセマンティクスに応じて制限する必要がある．例えば扱うデータ構造が下限が 0 の数値である場合，無制限に減算を認めるわけにはいかず，減算可能な範囲に制限がかかる．すなわち，全ての減算を適用しても 0 を下回らない範囲であれば，減算が可能であるとする．このように，データ構造のセマンティクスに応じて，操作 (演算) の範囲も制限する必要がある．なお，この制限は分断時にのみデータ構造にかかるものであり，通常状態ではこの制限はかからない．

3.4 想定するネットワークの状態と対応

分散オブジェクトライブラリでは，分散型インターネットサービスにおけるネットワークが取り得る状態は，帯域が広い回線を利用可能な通常状態と，いずれかのサーバ間でネットワーク分断が発生している分断状態の 2 状態だと想定している．通常状態では，各サーバはデータに対する操作ログを保持しておき，サーバ間でその操作ログを送信し合うことで同期を取る．それにより，サーバ間で一貫性を維持している．また分断状態では，各サーバに予め設定された一貫性制御のための制限に従ってサービスに制限を加え，分断回復後にマージ処理を行うことで，サービスとしての一貫性を維持する．

4. 提　　案

本節では，ネットワーク分断時にも利用可能な狭帯域のバックアップ回線の存在を前提とし，一貫性を維持するためにかかる制限を各サーバが管理し，必要なサーバに対して権限を委譲することで，可用性を向上させる．以下で，提案概要，各サーバにおける権限の管理及び，想定するネットワークの状態とその対応について述べる．

4.1 提 案 概 要

まず前提として，ネットワーク分断時にかかる操作範囲の，制限を狭帯域時にもかけることとする．また，操作可能なデータ範囲は各サーバに均等に分割されており，あるサーバが操作可能でないデータの範囲は他のサーバでは操作が可能なようになっている．そして本研究では，その制限を緩和するための通信を行う．すなわち，あるサーバにかけられている制限を緩和するために，他のサーバから，自サーバが操作可能でないデータの範囲を操作する権限の委譲を要求することで，各サーバの操作可能な範囲を広げる．それにより，分断時の制限がかかった状態と比較して制限が非常に緩くなるため，可用性が向上する．また，操作可能な任意の範囲を別のサーバに委譲したサーバにおいては，委譲した範囲が操作可能でなくなるため，データ全体で見ると，各サーバにおける操作可能な範囲は重複しない．よってサービス全体としての一貫性が，分散オブジェクトライブラリの想定する分断状態と比較してほぼ同程度に維持可能である．また必要なデータの権限をその権限を所持するサーバから委譲する際，サーバ間での通信を権限の委譲要求及びそれに対する返答のみに抑えることで，各サーバで行われた操作のログを送信する通常時の同期通信と比較して，通信量を削減することが可能になる．そのため，通信量を極力減らすことが求められる狭帯域回線の利用時においては，有用であると考えられる．

4.2 権限の管理

前提として，各サーバにどのような初期制限がかかっているかは，全てのサーバが知っている．また，各サーバにかかる初期制限から，必要な権限を所持するサーバを割り出すことが可能である．そのため，狭帯域時に必要な権限を持っているサーバに対して権限の委譲要求を出すことが可能である．また，委譲した権限についても権限管理テーブルに委譲先を記録しておく．それにより，もし委譲した権限が利用されていない場合は，その権限の委譲要求を出してきたサーバに対して，現在その権限を所持しているサーバの情報を渡すことで，他のサーバがその権限を所持するサーバに対して権限の委譲要求を出すことができるため，

利用可能な権限の有効利用が可能である．また権限の移譲先を記憶しておくことで，任意の権限を所持するサーバが必ずどのサーバからも分かるようになっている．また，各サーバに割り当てられた範囲の権限及び委譲されてきた権限の全てが，各サーバが持つ権限管理テーブルによって管理されている．しかし管理する権限の数の増加に伴い管理コストが非常に大きくなってしまう．その場合は権限管理テーブルに，Chord などの分散ハッシュテーブルなどを用いて，管理を他のサーバに任せることも可能である．しかし本研究では，提案を組み込むアプリケーションにおいて扱う権限の量を考慮して，各サーバが割り当てられた権限及び委譲されてきた権限を全て各サーバの管理テーブル上で扱う．

4.3 想定するネットワーク状態とその対応

分散オブジェクトライブラリがネットワークの状態を通常状態と分断状態の 2 状態と想定していたのに対し，本研究においては，帯域が狭いながらも利用可能な回線がある状態も加え，3 状態を想定する．狭帯域回線を利用するサーバ間では，各サーバに分割された操作可能なデータの範囲を緩和する．すなわちあるデータの範囲に対する操作権限を他のサーバに委譲するための通信を行う．権限の委譲については，個別委譲と範囲委譲の 2 種類の方式が考えられる．以下では，その 2 種類の方式について概要と利点，欠点について説明する．

個別委譲方式 各サーバ間で，すぐに利用する権限を個別に委譲する方式である．この方式では必要な権限のみの委譲を行うため，利用する権限が操作範囲の中にランダムに分布するようなワークロードの場合に，通信量が少なくてすむという点が効率的である．例えば，ランダムでつけられるユーザ識別用の ID などに関しては，偏りが無いと考えられるため，権限を予め委譲しておくことが難しい．よって，必要な権限を必要なだけ委譲することが望ましいと考える．しかし，利用する権限が操作範囲の中で偏るワークロードの場合には，範囲委譲方式と比較して委譲要求を出す回数が多くなるため，通信量が多くなるという欠点もある．

範囲委譲方式 各サーバ間で，すぐに利用する権限に加えてその権限近辺の任意の範囲の権限を委譲する方式である．この方式では 1 回の委譲要求で複数の権限を委譲できるため，利用する権限が操作範囲の中で複数あるいは 1 カ所に集中するワークロードの場合に，通信量が少なくてすむという利点がある．例えば，ユーザ名など何らかの形で偏りが現れるようなデータの権限委譲については，こちらの方式の方が良いと考えられる．特定の地域に多い名前などは，この形式で権限委譲を行うことでより通信量が少なくなり，またレスポンスの向上も期待できる．しかし，利用する権限が操作範囲の中にランダムに分布するようなワークロードの場合には，個別委譲方式と比較して委譲要求を出す回数が多くなるため，通信量が多くなるという欠点もある．

ここでは，セット構造に対してかかる制限を緩和する手法の説明するために，何らかのサービスにおけるユーザ登録を例に挙げる．セット構造においては，分断時は各サーバに対して追加可能な要素の範囲に対して制限がかかる．追加可能な要素の名前空間を各サーバに均等に分割するため，分断時には割り当てられた空間内の要素のみが追加可能となる．よってサービス提供サーバが 3 台あった場合は，分断状態における可用性は，単純に考えると通常状態における可用性の 1/3 となる．ユーザ登録においては，ユーザ名で利用可能な文字が大文字と小文字を区別したアルファベットのみであるとし，頭文字に利用可能なアルファベットに制限がかかるとすると，分断時には各サーバは，制限がかかっていない空間にあるアルファベットを頭文字に持つユーザのみが作成可能である．例えばあるサーバに対して，r〜Z の文字に制限がかかっているとする．その場合，分断時にはそのサーバに接続しているユーザは，制限がかかっていない a〜q までのアルファベットを頭文字に持つユーザ名のみが登録可能である．それに対して狭帯域回線が利用可能である場合には，狭帯域回線を通して通信可能なサーバと，それぞれが持つユーザの作成権限を委譲し合うことが可能となる．分断時に頭文字が a〜q までのユーザしか作成権限を持たなかったサーバに接続したユーザが，頭文字 D のユーザ名を持つユーザを登録する場合には，D の文字のユーザ作成権限を持つサーバに対して委譲要求を出すことで，そのユーザの作成権限を得て，ユーザを作成することができる．よって，ユーザ名の作成という面から見れば，分断時と比較して可用性は上がり，通常状態と同レベルの可用性になる．しかし，そのユーザを作成したことは他のサーバからは分かるが，作成したユーザの情報については作成したサーバに接続しないと参照ができない．よってサービス全体の可用性としては，通常状態よりは下がるが，分断状態よりは上げることが可能である．

5. 評　　価

本章では，提案する手法の有効性を確認するために，提案手法を組み込んだ分散オブジェクトライブラリを利用して，実際のインターネットサービスに組み込んだ上で行った実験の結果について述べる．本論文では，オープンソースのインターネットオークション・サービスである RUBiS[9] を用いた．RUBiS は，eBay を参考にしたオークションサイトの簡略化モデルとして設計されており，性能評価アプリケーションも用意されているため，使い勝手がよい．さらに簡略化されてはいるものの，一般的なオークションサービスとしての機能は一通りそろっている．そのため先行研究においてもこの RUBiS が実験で利用されており，

表 1 RUBiS で扱う情報
Table 1 Information used by RUBiS

内部情報	説明
商品情報	商品名，在庫数，即落札価格，出品者，入札日時等の情報
ユーザ情報	ユーザ名，パスワード，居住地域，評価等の情報
即落札情報	即落札者，即落札個数，即落札日時などの情報
入札情報	入札者，入札数，入札価格，入札日時等の情報
コメント情報	ユーザに対する評価やコメント情報
地域情報	ユーザの居住地域情報
カテゴリー情報	出品商品のカテゴリ情報

比較のため本研究でも RUBiS を用いて実験を行う．本研究における提案手法の有効性の確認のため，本論文ではネットワークの各状態におけるのユーザ登録の成功率をそれぞれ測定及び比較した．また通常時と比較して，狭帯域時に通信量が削減できていることを示すために，両状態におけるサーバ間の通信量を測定，比較した．

5.1 RUBiS への適用

RUBiS では，内部で扱う情報を**表 1** で示す 7 つに分類している．

また，表 1 の 7 つの情報をそれぞれ分散オブジェクトを用いて管理するために，7 つの情報に対して適用可能なオブジェクトとの対応関係を**表 2** に示す．RUBiS の機能上，商品情報やユーザ情報の編集は行うことができない．ここで，ユーザ情報について説明する．ユーザ情報は，RUBiS において分断発生時に操作範囲が制限される情報であり，狭帯域時にはその部分の権限をサーバ間で委譲しあうことで，可用性を向上させる．以降の実験についても，ユーザ情報に関して述べる．

ユーザ情報 RUBiS においては，分断時に作成可能なユーザを，ユーザの名前によって制限する．ユーザ情報の管理はセット構造を用いて行われる．分断時にはユーザの名前空間を各サーバに均等に分割するため，各サーバにおいては割り当てられた名前空間に名前が有るユーザのみが作成可能となる．それに対して狭帯域時には，各サーバが保持する利用可能な名前空間をやり取りすることが可能となるため，既に作成されているユーザ以外は作成可能となる．ただし狭帯域時には各サーバ間のデータの同期は行っていないため，あるサーバ上で動作している RUBiS において作成したユーザ名を利用して，他のサーバ上で動作している RUBiS にログインすることはできない．

5.2 実　　験

実験においては，3 台のサーバマシンと 3 台のクライアントマシンを用い，各サーバ上で

表 2 RUBiS において分断時にかかる制限
Table 2 Restriction in Network-partitioning

情報	使用データ構造 (区分)	操作自体の制限	操作範囲の制限
ユーザ情報	セット構造 (全体)	追加操作のみ	ユーザ名の頭文字による制限
	数値構造 (評価)	加減算操作のみ	なし
商品情報	セット構造 (全体)	追加操作のみ	なし
	数値構造 (入札数)	加算操作のみ	なし
	数値構造 (現在価格)	代入操作のみ	初期価格以上
	数値構造 (在庫数)	減算操作のみ	0 以上
入札情報	セット構造 (全体)	追加操作のみ	なし
即落札情報	セット構造 (全体)	追加操作のみ	なし
コメント情報	セット構造 (全体)	追加操作のみ	なし
カテゴリー情報	セット構造 (全体)	操作禁止	なし
地域情報	セット構造 (全体)	操作禁止	なし

表 3 実験環境
Table 3 Machine Environment

CPU	Intel(R) Xeon(R) CPU E5502 @ 1.87GHz
Memory	4GB
OS	Linux 2.6.18(CentOS 5.5)

RUBiS を動作させた．実験に用いた各サーバ及びクライアントのスペックは**表 3** のとおりである．以下で，分断時と比較して狭帯域時に可用性が向上するかどうか確認するための実験について述べる．また，通常時と比較してサーバ間の通信量が削減できているかどうか確認するための実験についても述べる．

5.2.1 サーバ間の通信量の測定

この実験においては，各クライアントは 3 台のサーバに対して 20 分間接続し，各クライアント上であらかじめ RUBiS に付属しているクライアントエミュレータを動作させ，それぞれ 240 人分のユーザの動作をエミュレーションした．また，ネットワーク分断が発生した後，少し時間が経過してからバックアップ回線に切り替わることを想定したため，始めの 6 分を通常状態，次の 1 分を分断状態，その次の 8 分を狭帯域状態にして実験を行った．狭帯域状態においては，20 分のうち 8 分は 1 台のサーバと他の 2 台のサーバ間の回線を狭帯域にし，それらのサーバ間の通信量を調べた．また、権限の委譲方式を範囲指定と個別指定に切り替えて、上記と同様の環境で通信量を測定した．

ネットワーク通信量を測定した実験結果について，**図 1** 及び**図 2**、**図 3** で示す．図 1 は，ネットワークが通常状態の際のあるサーバの同期通信の受信量を示しており，図 2 及び図 3

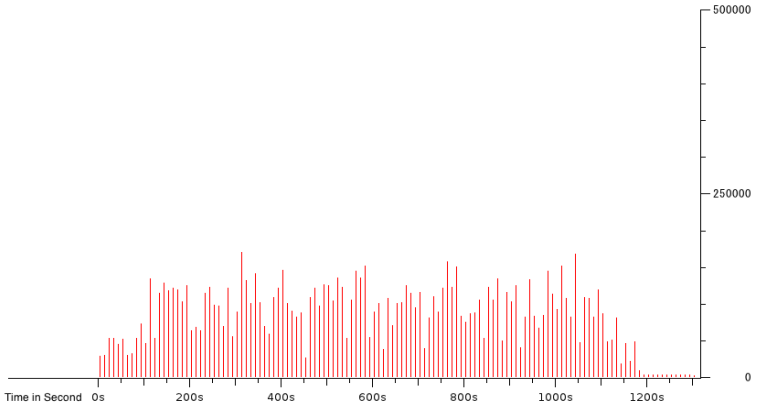

図 1 通常時のサーバの通信量
Fig. 1 Received byte in common network situation

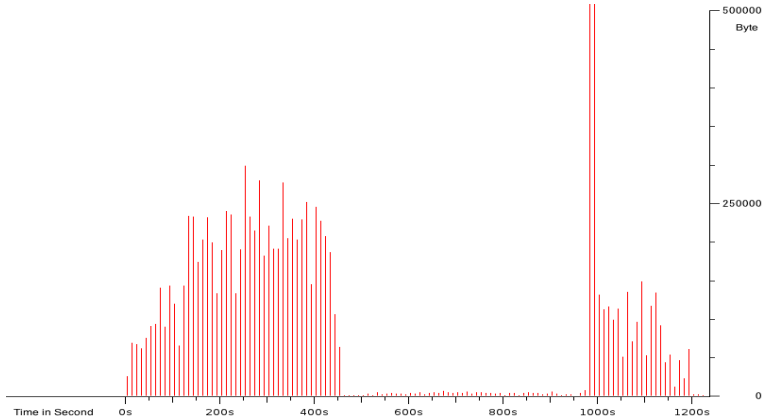

図 2 個別委譲方式を利用した場合の狭帯域時のサーバの通信量
Fig. 2 Received byte in Narrowband network with indeividual

は，図 1 で示したサーバと他のサーバ間のネットワークが狭帯域状態になった場合の，サーバの受信量を示している．また、図 3 は、権限の委譲方式を範囲指定にして測定した結果であり、図 2 は個別指定にして測定した結果である．

図 3 及び図 2 においては，図 1 と比較して狭帯域時の通信量を削減することに成功している．また，図 3 及び図 2 においては 500 秒あたりから 1000 秒あたりまでが狭帯域状態である．どちらも 1000 秒近くに通信量が急増しているが，これは分断及び狭帯域中に行われた全ての操作ログを他の各サーバから受信しているためである．また一回の委譲要求にかかる通信量は、返答も含めると約 600Byte であり、750 回程度の委譲要求が行われた．よって、委譲要求の秒間通信量は、RUBiS のワークロードにおいては約 930Byte/s であった．なお先行研究である分散オブジェクトライブラリを組み込んだ RUBiS において、本実験と同一ワークロードでエミュレーションした場合が図 1 であるが、通常時の同期通信における秒間通信量は約 5KByte/s であり、それと比較しても通信量が削減できていると言える．今回の実装においては、要求と返答に同一のパケット形式を利用しており、また権限の委譲方式が異なる場合でも同一のパケット形式を利用しているため、それらの部分の実装を工夫することで、さらに権限委譲要求の通信量を削減することが可能となると考えられる．また今回の実験においては、権限の個別指定と範囲指定による通信量の有意差は見られなかった．これは、上述した共通パケットの利用とともに、RUBiS のワークロードの特性が関係していると考える．例えばユーザ名について考えると、特定の地域で作られやすいユーザ名や、あるユーザ名のあとに作られやすいユーザ名といった形で、何らかの偏りが見られる場合には、予めそれらの空間を範囲指定方式による委譲しておくことで、権限の委譲要求を減らすことが可能となる．

5.2.2 各サーバにおける可用性の測定

この実験においては，作成するユーザのユーザ名を予め 1000 人分決めておき，そのユーザ名を 3 分割して各クライアントに割り当て，各クライアントは通常状態・分断状態・狭帯域状態の各状態にある各サーバに対して自分に割り当てられたユーザの登録を行い，その成功率をサーバごとに確認する．各サーバをサーバ 1，サーバ 2，サーバ 3 とした場合のユーザ作成の成功率を**表 4** に示す．なお，基本的に 5.2.1 で示した実験環境においてユーザ作成を行分断状態において各サーバの成功率に多少ずれが見られるが，原因としてはランダムなユーザを作成していること及び，実装におけるハッシュ関数の利用が考えられる．また実装においては名前空間の分割にハッシュ関数を用いており，そのハッシュ関数が出すハッシュ値の空間とユーザの名前空間が一致していないため，各サーバ間で割り当てられた名前空間の範囲の大きさにずれが生じるとも考えられる．

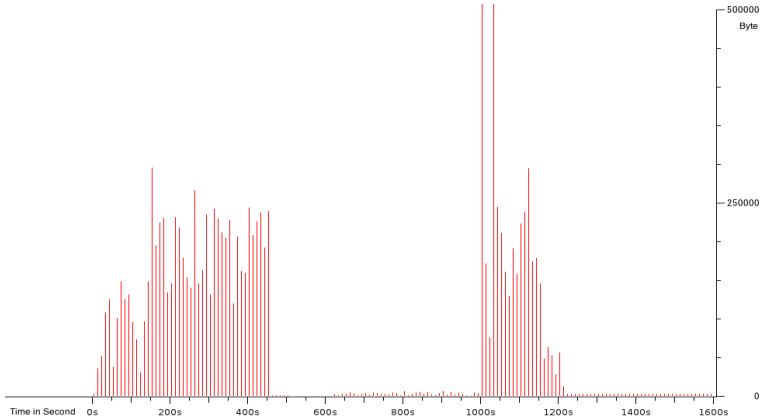

図 3 範囲委譲方式を利用した場合の狭帯域時のサーバの通信量
Fig. 3 Received byte in Narrowband network with scope

表 4 ユーザ作成の成功率
Table 4 Sucsess rate of user registration

サーバ名	分断状態	狭帯域状態	通常状態
サーバ 1	34.5%	100%	100%
サーバ 2	36.3%	100%	100%
サーバ 3	32.1%	100%	100%

6. まとめと今後の課題

本論文では，ネットワーク分断時にも利用可能な狭帯域のバックアップ回線の存在を前提とし，一貫性を維持しつつ少ない通信量で可用性を向上させるために，分散オブジェクトライブラリを拡張し，分断時に各サーバにかかる制限を緩和するための通信を行うことを提案した．分断時に各サーバごとに制限される操作範囲は異なるが，それらの範囲を各サーバ間で委譲しあうことで，各サーバが操作可能な範囲が広がり，分断時と比較して狭帯域時の方が可用性が向上する．また一貫性に関しては，権限の委譲の効力により，狭帯域時は分断時と同等である．さらに，分散オブジェクトライブラリを拡張して本論文で提案した手法を組み込み，動作実験を行うことで，狭帯域時と比較して可用性の向上及び，通信量の削減を確認した．

今後の課題としては，パケット形式の工夫による通信量の削減が挙げられる．また，今回実装を行った RUBiS においては，セマンティクス上かけることができない制限などがある．それらの制限に関しても狭帯域時における本論文の提案の有効性を示すために，RUBiS のようなオークションサービスとはセマンティクスが異なるサービスに実装を行うことが挙げられる．さらに，権限委譲をより効率的に行うために，個別委譲方式と範囲委譲方式の両方式における通信量とワークロードの関係について分析を行う点が挙げられる．またその分析結果及び，ウェブサーバの負荷分散などの技術を応用することで，個別委譲方式及び範囲委譲方式を柔軟に切り替え，より効率的な通信を行うことが可能になると考えられる．

参 考 文 献

1) Brewer,E.A.:Towards robust ditributed systems(2000).PODC'00(Invited talk).
2) DeCandia,G. Hastorum,D. Jampani,M. Kakulapati,G. Lakshman,A. Pilchin,A. Sivasubramanian,S. Vosshall,P. and Vogels,W :Dynamo:amazon's highly available key-value store, SOSP'07 : Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles, New York, NY, USA, ACN, pp.205-220(2007)
3) Benjamin C. Pierce. "Harmony: The Art of Reconciliation." TGC,April,2005.http://www.cis.upenn.edu/∼bcpierce/papers/harmony-tgc-talk-2005.pdf
4) Sivasubramanian,S. Alonso,G. Pierre,G. and Steen,M.v.: GlobeDB: Autonomic Data Replication for Web Applications, In Proc. of the 14th internationals conference on World Wide Web(WWW'05), pp.70-78(1999).
5) Gao,L. Dahlin,M. Zheng,J. and Iyengar,A.: Application specific data replication for edge services, WWW'03: Proceedings of the 12th international conference on World Wide Web, New York, NY, USA, ACM, pp449-460(2003).
6) 小長谷秋雄, 宮澤和徳, 品川高廣, 加藤和彦. "ネットワーク分断に対応した分散オブジェクトライブラリ" 情報処理学会論文誌, コンピューティングシステム, Vol.3, No.2, pp.99-112, 2010
7) Dan. Teodosiu, Nikolaj. Bjorner, Yuri. Furevich, Mard. Manasse, Joe. Porkka, "Optimizing File Replication over Limited-Bandwidth Networks using Remote Differential Compression", Microsoft Corporation, 2006.
8) Yui-Wah,Lee. Kwong-Sak,Leung. and M,Satyanarayanan.: Operation-based update propagation in a mobile file system: In Proceedings of the 1999 USENIX Technical Conference, Monterey, CA, 1999.
9) Team,R. : RUBiS, http://rubis.ow2.org/index.html