解説

Business Continuity（事業継続）実現に向けた情報システム技術

応般

西澤 格[*1]　藤原 真二[*2]　山本 政行[*3]　矢崎 武己[*1]

[*1] 日立製作所 中央研究所　[*2] 日立製作所 IT プラットフォーム事業本部　[*3] 日立製作所 横浜研究所

事業継続と情報システム

◆ 広域災害と事業継続の重要性

日本では，1995 年の阪神・淡路大震災，2007 年の新潟中越沖地震，2009 年の新型インフルエンザパンデミック，2011 年の東日本大震災が，米国では，2000 ～ 01 年のカリフォルニア大停電，2001 年の同時多発テロ，2005 年のハリケーンカトリーナが発生した．日本や米国のほかにも，2010 年のアイスランドでの火山噴火，2011 年のタイ大洪水など，全世界で災害が発生している．経済がグローバル化し，サプライチェーンが複雑化する中で，ある地域の災害が全世界の経済に及ぼす影響が拡大している．たとえば，東日本大震災による経済的被害は約 17 兆円に及ぶと推定されている[1]．

一方，情報システムは高度化し，システムの停止がビジネスに及ぼす影響は拡大しており，深刻なデータ損失を発生させた企業の 43% は廃業に追い込まれているとの報告がある[2]．この流れを受けて，日本では内閣府によって事業継続（Business Continuity，以下 BC）のガイドラインが策定されるようになっている[3]．また，国際標準化機構（ISO）や英国規格協会（BSI）では BC に関する規格制定も進んでいる．たとえば，ISO の規格である ISO/IEC 27031 では，情報システムが組織の BC のために求められるレベルの対応能力を適切な状態に保つための取り組みが規定されている[4]．

BC の計画にはオフィスや人員の確保など非システム要素も含まれるが，本稿では議論の範囲を IT に限定する．BC 実現のための情報システムに対する要件として，文献 3）ではバックアップの重要性を，そして文献 4）では事業が拠り所とするデータ保護の重要性を強調している．

◆ ディザスタリカバリ

システムのバックアップを作成し，災害からデータを保護するための代表的な技術として，ディザスタリカバリ（Disaster Recovery，以下 DR）を挙げることができる．DR システムを構築する際の重要な指標としては，RPO（Recovery Point Objective，目標復旧地点）および RTO（Recovery Time Objective，目標復旧時間）が用いられる．RPO は，どのタイミングまで遡ったデータを回復するかを示す指標である．たとえば，1 日前のデータが回復できればよい場合，RPO は 1 日となる．データ保護の観点からは RPO はできるだけ短いことが望ましい．一方，RTO はシステムの復旧に要する時間を示す指標である．システムを迅速に復旧させるという観点から，こちらも短い方が望ましい．DR システムにおける RPO と RTO の関係を，**図 -1** に示す．

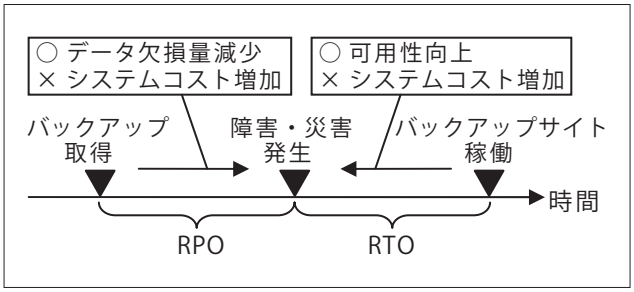

図 -1　DR システムにおける RPO と RTO

RPO に影響を及ぼす要因としては，バックアップの間隔が挙げられる．一方 RTO に影響を及ぼす要因としては，リストア対象のデータサイズ，データを格納しているデバイスからの読み出し速度，バックアップ先のサイト（以下，リモートサイト）でのリストア方式が挙げられる．RPO と RTO を短縮するためには，システムを構成する計算機，ネットワーク，ソフトウェアへの性能要求が高まり，かつシステム構築・運用コストが増加するという課題がある．業務が異なればデータ管理の対象も多岐にわたり，かつ前記 RPO，RTO に対する要求も異なる．加えて，システム構築・運用にかけられる費用にも制限があるため，要求とコストのバランスをとりながらシステムを設計する必要がある．

◆ 本稿の構成

通信，金融，公共等の社会インフラを支えるミッションクリティカルな業務では，データ欠損が許容できず，かつシステム復旧時間も可能な限り短くする必要がある．すなわち，前述した RPO と RTO はいずれも 0 に近づけることが望ましい．

本稿では，ミッションクリティカルな業務の事業継続を支える情報システム技術のうち，データ管理技術と通信技術にフォーカスする．データ管理技術としては，現在基幹業務のデータ管理の中核を担っているデータベースと，オフィス文書やメール等，業務推進に必須のファイルを取り上げる．そして，データベースを対象とした DR システム構築に必要なレプリケーション技術[☆1]を「データベースのレプリケーション技術」で，そしてファイルを対象とした DR システム構築に必要なバックアップ，リストア技術を「ファイルバックアップとリストア技術」で説明する．さらに「大陸間レベルのレプリケーションを実現する WAN 高速化技術」では，DR システムで必須となる，低遅延で遠隔サイトへのデータのレプリケーションを実現するためのネットワーク高速化技術を説明する．最後に「まとめ」で，本稿をまとめるとともに，今後の課題を示す．

☆1 本稿では，データベースではシステム全体としての複製を作成し，災害時に複製先でシステムが立ち上がることを想定して，バックアップではなくレプリケーションという語を用いている．

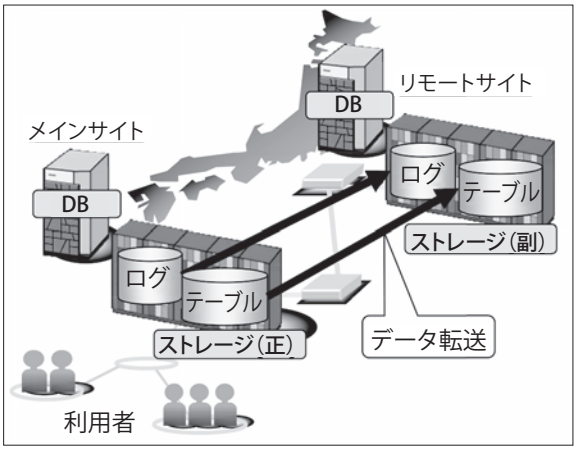

図 -2　データベースのディザスタリカバリ

データベースのレプリケーション技術

◆ 方式概要

前述したように，広域災害時にはデータベース（以下，DB）の DR が重要となる．DB の DR では，データ保護の観点から RPO=0 の実現を目的としている．**図 -2** に DB を対象とした DR システムの概念図を示す．本システムでは，RPO=0 を達成するため，業務を実行しているメインサイトのストレージ（正）に格納されているテーブルデータ（以下，単にテーブルとする）とデータベースのログ（以下，単にログとする）の更新分を，リモートサイトのストレージ（副）へリアルタイムで複製を作成（レプリケーション）する．レプリケーション時のメインサイトのストレージ（正）とリモートサイトのストレージ（副）のデータ転送には，ストレージのリモートコピーが利用される．

図 -3 に従来方式として，(a) 同期転送方式と (b) 非同期転送方式の概念図を示す．リモートサイトが遠距離の場合，(a) の同期転送方式ではメインサイトとリモートサイトのテーブルの同期をとるために，DB の Write 要求完了（Step3）はストレージ間のデータ転送（Step2）の完了を待つ必要があり，オンライン性能確保が課題となっていた．

一方 (b) の非同期転送方式では，メインサイト

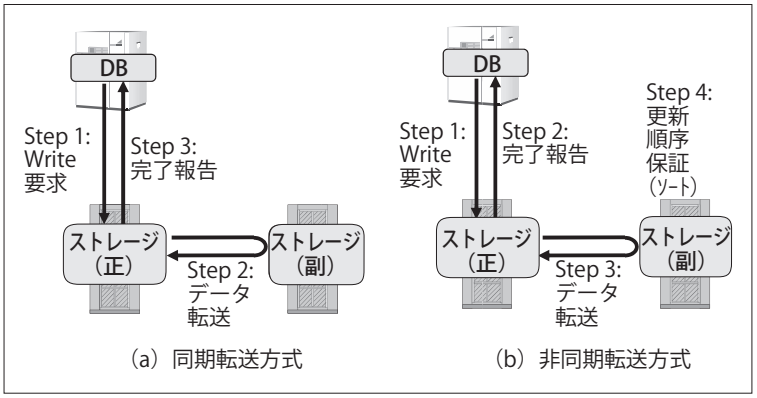

図 -3　従来方式

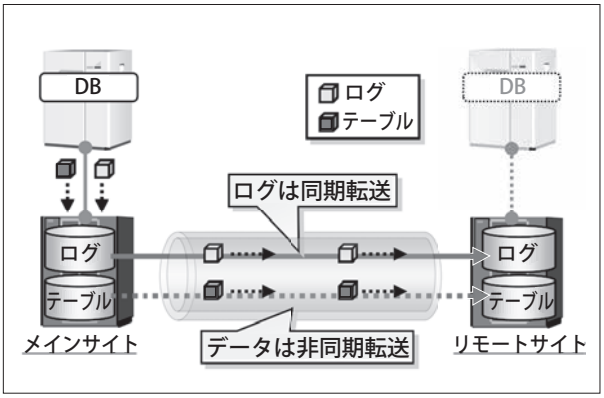

図 -4　ハイブリッド転送方式

での DB の Write 要求は，リモートサイトのストレージに反映（Step3）される前に完了（Step2）するため，メインサイトの DB の更新の一部がリモートサイトに反映されない「データ欠損」が生じ，RPO=0 が保証できないことが課題となっていた．

そこで我々は，ストレージと DB を連携させたハイブリッド転送方式を，文部科学省のリーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」のストレージ・データベース融合技術にて，東京大学と共同で開発した[5]．本方式では，通常時のオンライン性能を維持しながら，災害時のテーブルのデータ欠損なしを保証するために，DB のログを同期転送，テーブルを非同期で転送する（**図 -4**）．ログを同期転送することで，メインサイトとリモートサイトで DB の最新の更新情報を同期させることができる．また，非同期で転送されるテーブルデータを利用することにより，リモートサイトでの DB 更新の負荷を削減することができる．

災害発生時のリモートサイトへの切替えでは，ストレージのリモートコピー機能と連携した DB のトランザクション処理により，データ欠損なしでの DB 回復を保証する．さらに，ログ同期転送の間に発生した他の更新処理のログをバッファに蓄積し，次回のログ同期転送時にまとめて転送することにより，平常時のオンライン性能への影響を最小化することができる．

◆ 評価と応用例

図 -5 に回線シミュレータを用いたハイブリッド転送方式の性能評価を示す．評価には業界団体 TPC（Transaction Processing Performance Council）によって策定された，卸売り会社のトランザクション処理システムをシミュレートして性能を評価する，TPC-C 相当の販売管理モデルを用いた．グラフの縦軸の TPS は，Transaction Per Second の略で，1 秒あたりのトランザクション処理数を表す．TPS はデータベースの性能を表す指標として広く用いられている．評価の結果，メインサイトの処理にリモートサイトへの通信処理が影響を受けない非同期転送のトランザクション処理性能を 100% とした場合，ハイブリッド転送方式では回線遅延 3ms の場合に 88%，回線遅延が 5ms の場合に 84% の性能を確保することができた．

ハイブリッド転送方式は，広域災害やテロなどに対してもデータ欠損が許されない重要なシステムで，かつオンライン性能も確保する必要がある場合に適用する．たとえば，1 日あたり最大 18 万件のデータを処理するシステムにおいて，400km 離れた遠隔地にリモートサイトを設置している実事例に適用されている．

ファイルバックアップとリストア技術

◆ 方式概要

急増するオフィス文書等のファイルの格納先として，NAS（Network Attached Storage）に代表されるファイルストレージ（以下，FS）は，企業内の各拠点や部門で普及してきている．FS の代表的な

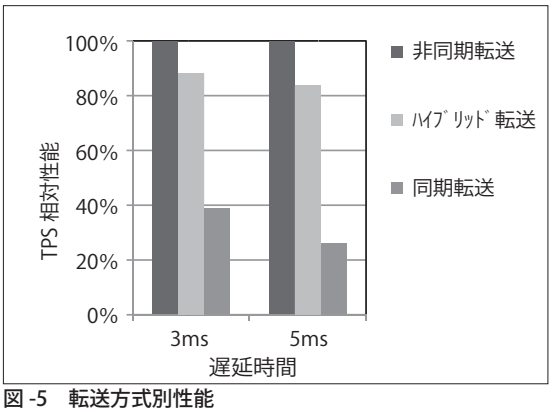

図 -5　転送方式別性能

	拠点別運用方式	集中運用方式
耐災害性	× 拠点被災に弱い	○ 拠点被災に強い
運用負担	× 拠点で管理要	○ 集中管理
運用性能	○ 拠点内 LAN 経由	× 拠点間 WAN 経由

表 -1　従来の運用方式と課題

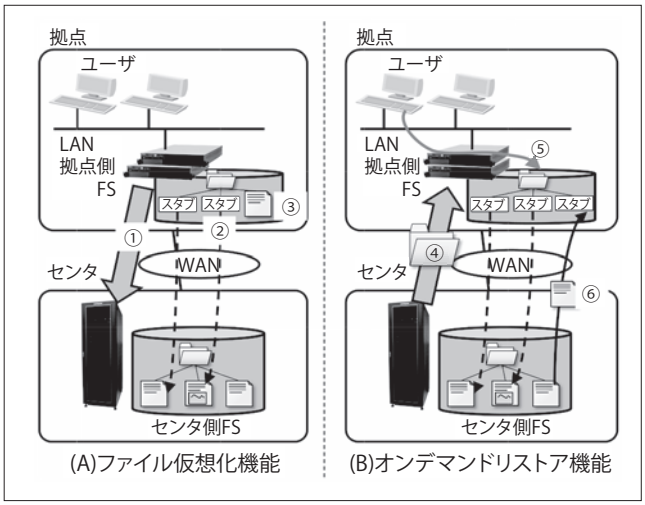

図 -6　ファイル仮想化とオンデマンドリストア

バックアップ運用としては，拠点ごとにバックアップを取得する「拠点別運用方式」と，センタ等の 1 拠点にファイルを転送してバックアップを集中取得する「集中運用方式」が挙げられる．

拠点別運用方式は，平時のファイルアクセスやリストアが拠点内 LAN 経由で高速に実行できるメリットを持つが，拠点の災害に対してファイルを保護する耐災害性の保証が必要，拠点ごとにバックアップ運用や容量管理等の運用負担が発生する課題がある．一方，集中運用方式は，拠点被災に対する耐災害性の確保および管理をセンタのストレージで集中して対応できるメリットがある．しかしながら，ファイルアクセスやリストアが拠点間 WAN（Wide Area Network）を経由するため，運用時の性能に課題がある．**表 -1** に両方式の利害得失をまとめる．

我々は，拠点側 FS とセンタ側 FS を併用することにより，拠点被災に対する耐災害性を確保しつつ拠点側の運用負担を軽減し，平時の高速ファイルアクセスや拠点被災時の迅速なサービス再開を実現した．開発した方式は，バックアップ時のファイル仮想化機能と，リカバリ時のオンデマンドリストア機能から構成される．以下，各機能の概要を説明する．

ファイル仮想化機能は，ファイル格納先を仮想化し，拠点ユーザには拠点側 FS へアクセスさせつつ，ファイルをセンタ側ストレージに集中バックアップする機能である（**図 -6**（A））．まず，拠点側からセンタ側 FS へ定期的に新規ファイルや更新ファイルをバックアップする（図 -6 ①）．ユーザが拠点側 FS からバックアップ先のファイルにアクセスできるように，バックアップしたファイルのスタブファイルを作成する（図 -6 ②）．スタブファイルとは，ファイル名，サイズ，アクセス制御情報など，ユーザからの見た目はまったく同じで，中がない特殊なファイルである．ユーザが拠点側 FS のスタブファイルにアクセスすると，拠点側 FS はセンタ側 FS から自動的に実体を取得し，ユーザは目的のファイルへアクセスできる．スタブファイルにアクセスするたびに，毎回センタ側 FS から実体を取得するのはオーバヘッドが大きいため，高頻度でアクセスされるファイルは拠点側 FS に実体をキャッシュし，性能を確保する（図 -6 ③）．

次に，図 -6（B）に示すオンデマンドリストア機能について説明する．従来，拠点被災等からの復旧で新たに拠点側 FS を構築するには，センタ側 FS からすべてのデータのリストアが完了するまでサービスを再開することができなかった．本機能では，必要なデータから段階的にデータをリストアすることで，迅速にファイル共有サービスを再開する．まず，センタ側 FS にバックアップしているデータから，最上位フォルダのみを拠点側 FS にリストアして，ファイル共有サービスを再開する（図 -6 ④）．

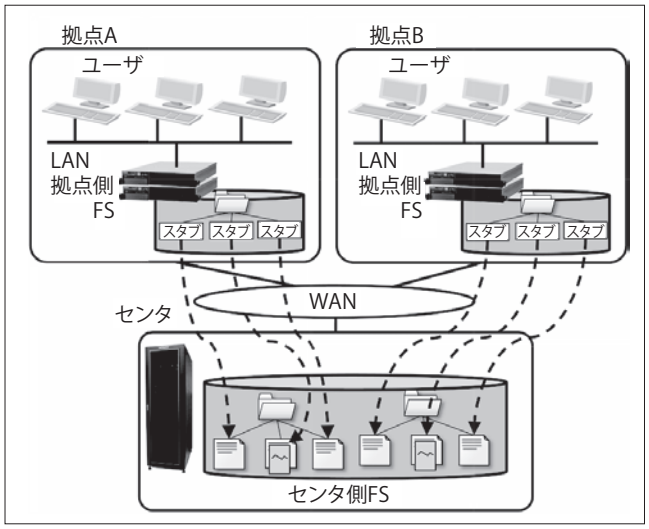

図-7　クラウドストレージサービス

ユーザがFSのフォルダにアクセスすると，直下のフォルダとスタブファイルをセンタ側FSからリストアする．リストアしたフォルダにユーザがアクセスした場合も，同様に直下のフォルダとスタブファイルをリストアする（図-6 ⑤）．スタブファイルに高頻度でアクセスが発生すると，仮想化機能の場合と同様に，ファイルデータを拠点側FSでキャッシュし，性能を確保する（図-6 ⑥）．

◆ 応用例と評価

近年のクラウドコンピューティングのサービス形態の1つに，ストレージ資源（容量やバックアップ運用等）をサービスとして提供するクラウドストレージ（Cloud Storage，以下CS）[6]がある．現在のCSは，(1) ユーザがCSサービスに依存したファイルアクセスプロトコルを利用しなければならない，(2) 低速のWAN経由でファイルをCSにアクセスしなければならないという課題がある．

ファイル仮想化機能とオンデマンドリストア機能を活用したCSサービス（**図-7**）により，ユーザは各拠点側FSが提供するファイルアクセスの標準プロトコル（NFSやCIFS等）を用いたファイル共有を利用するだけで，センタ側にあるクラウドストレージを意識する必要はない．さらに，拠点側での平時高速アクセスも実現できる．我々は，企業内の多拠点のデータ一元バックアップ運用だけでなく，提案方式を活用したCSサービスの提案も進めている．

提案方式と従来の集中運用方式とをRPOとRTOの観点から評価する．RPOに関しては，両方式ともあらかじめ定義したバックアップ間隔でセンタ側FSにバックアップする運用により，同等となる．一方RTOに関しては，提案方式は従来の集中運用方式よりも短縮できる．その理由は，従来の集中運用方式ではセンタ側FSから拠点側FSに全データのリストアが完了するまでサービスを再開できないのに対し，提案方式ではオンデマンドリストア機能により，ルートディレクトリとサブディレクトリのスタブ情報をリストアした時点で，サービスを再開できるためである．

大陸間レベルのレプリケーションを実現する WAN 高速化技術

◆ 方式概要

金融の勘定系データに代表される欠損が許されないデータに関しては，広域災害に備えて遠隔サイトへのデータのレプリケーション，可能であれば大陸レベルのレプリケーションの実現が望ましい．しかしながら，遠隔サイトへのレプリケーションでは大きな通信遅延が発生し，現在普及している通信プロトコルTCPでは，送受信端末間の通信帯域が低下し，データのレプリケーションに時間を要する．その結果，企業データのデータレプリケーションに，たとえば数日の時間を必要とし，RPOが長くなるといった課題がある．そこで，我々はTCPに準拠しつつ，通信遅延が大きな環境下にて通信帯域を向上する，協調型TCPを開発した．以下，開発技術の詳細と評価結果，およびその応用例を述べる．

データ欠落のないデータ送信を実現するプロトコルとしてはTCPが一般的に用いられている．TCPでは，受信端末が送信端末からのパケットを受信すると，受信端末は応答確認（ACK）を送信端末に送信する．送信端末は，ACKによりパケットの送達を確認し，送達確認のないパケットを再送信することでデータ欠落のないデータ送信を実現する．送信端末の通信帯域を制御するため，ACKを受信す

ることなく送信可能なバイト数（congestion window，以下 cwnd）と，その最大値（winsize）が規定される．送信端末は cwnd 分のパケットを送信した後，往復遅延時間（Round Trip Time，以下 RTT）を経過するまで新たなパケットを送信できないため，RTT が大きくなるにつれて帯域が低下する（**図 -8** 左図）．また送信端末は，パケット廃棄を検知すると cwnd を一定割合に減少させる．通常のネットワークでは，通信帯域の平均が回線帯域を下回っていても，ほかの端末の突発的なパケット送信によりパケット廃棄が発生する場合があり，通信帯域が低下する（図 -8 右図）．

そこで，我々はこれらの課題を解決する協調型 TCP を開発した．協調型 TCP では，ACK を待たずにパケットを送信することで，RTT の影響を受けない通信を実現する（**図 -9** 左図）．さらに，送信端末は通信帯域の制御に cwnd や winsize を使用せず，パケット廃棄率の時間変化によりネットワークの混雑状態を判定し，動的に通信帯域を制御する（図 -9 右図）．パケット廃棄に応じた最適な通信帯域に制御することで，ネットワークでのパケット廃棄に影響を受けない通信を実現する．

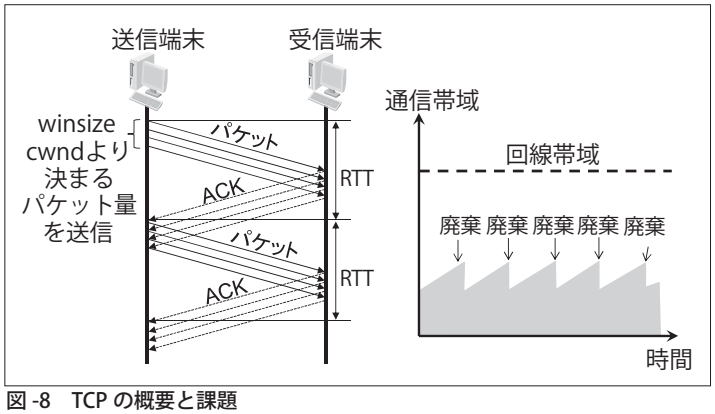

図 -8　TCP の概要と課題

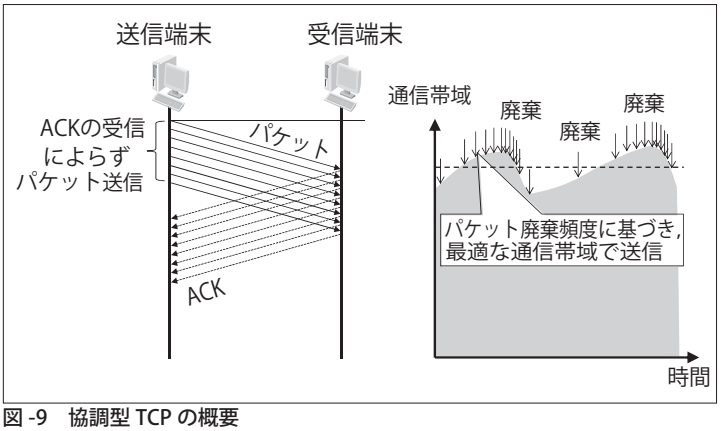

図 -9　協調型 TCP の概要

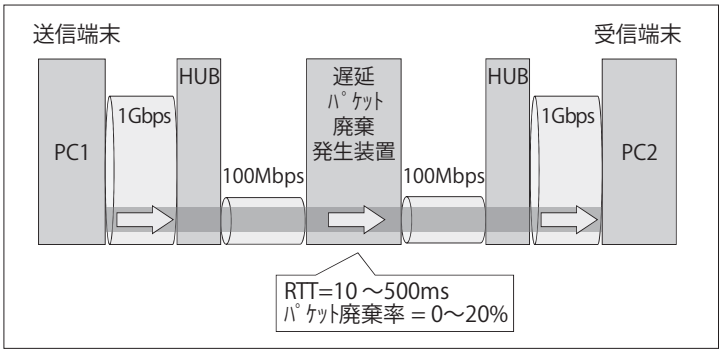

図 -10　協調型 TCP の評価環境

◆ 評価と応用例

協調型 TCP を，**図 -10** に示す WAN を模擬した実験環境にて評価した．本環境は TCP と協調型 TCP を実装した送信端末と受信端末，および WAN を模擬するための遅延・パケット廃棄発生装置で構成される．

送信端末と受信端末間に 1 セッションを設定し，ftp によりパケットを転送した場合の評価結果を**図 -11** に示す．図 -11 左図は，パケット廃棄率を 0% とした際の RTT と通信帯域の関係を，そして右図は RTT を 200ms と固定した際のパケット廃棄率と通信帯域の関係を示す．左図によると，通常の TCP では RTT に依存して帯域が減少するのに対し，協調型 TCP では常に 90Mbps 以上を達成している．一方右図によると，通常の TCP では高品質なネットワークで実現し得る 0.01% の廃棄率において帯域が 5Mbps 程度となるのに対し，協調型 TCP ではパケット廃棄率が 1% 以下の環境下において 90Mbps 以上の帯域を達成している．以上により，協調型 TCP では通信遅延の増加による通信帯域の劣化を大幅に抑止できることが分かる．

開発した協調型 TCP は，特に通信遅延が大きな WAN の高速化に有効であり，その応用例として，

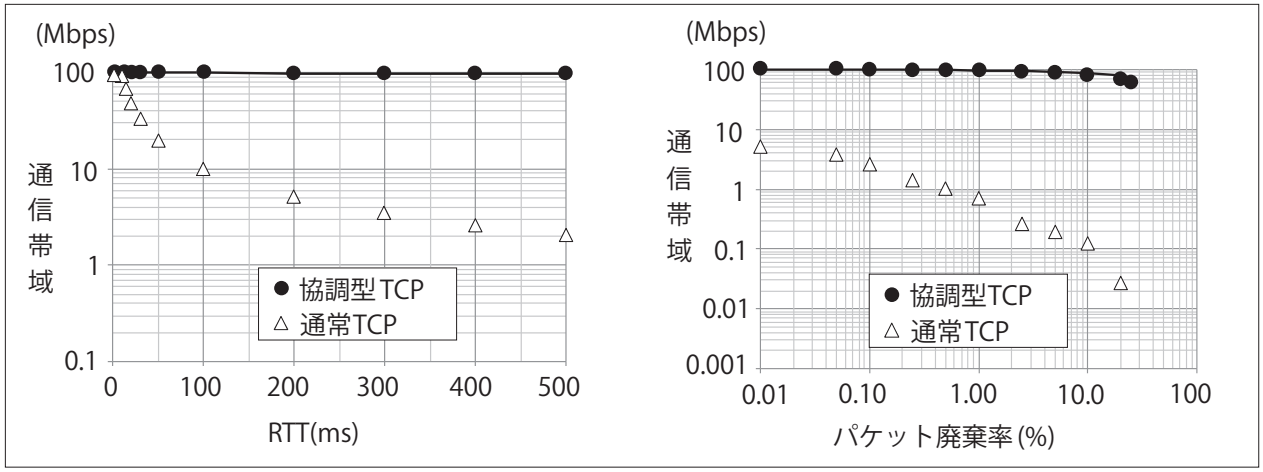

図-11　評価結果

WAN経由でサービスを提供するクラウドサービスの応答性向上が挙げられる．特に通信遅延が応答性に大きく影響する，大量データを一度に通信するクラウドサービスやクラウドアプリケーションへの適用が有効と考える．

製造業，金融などの企業においては，分散した企業拠点に配置されていた設計データ，金融データなどをクラウドに集約して企業拠点がWAN経由で情報を共有するデータベースの一元化によって情報通信システムの運用コストの低減を図る動きがある．通信帯域の低下は企業の各拠点からデータをアクセスする際の応答性（転送時間）に大きな影響を与える．協調型TCPの適用によりWANでの通信を高速化することで，企業拠点からのデータベースの高速な共有を実現し，企業の生産性を向上することが可能となる．

まとめ

本稿では，ミッションクリティカルな業務の事業継続に向けた情報システム技術として，データベースのレプリケーション技術，ファイルバックアップとリストア技術，そして大陸間レベルのレプリケーションを実現するWAN高速化技術を取り上げ，技術の概要，およびその適用例を解説した．

事業継続の実現に向け，ITシステムへの要求はますます高まる傾向にある．今後，高性能化と高機能化に加えて，システム開発・運用コストの削減，およびより使いやすいシステムを目指した運用管理容易化技術の開発を継続し，技術・製品を通じて社会に貢献していきたい．

参考文献

1) 内閣府：平成23年度年次経済財政報告 (2011).
2) Schmidt, K. : High Availability and Disaster Recovery, Springer (2010).
3) 内閣府：事業継続ガイドライン 第二版 ―わが国企業の減災と災害対応の向上のために― (2009).
4) ISO/IEC 27031 Information Technology - Security Techniques - Guidelines for Information and Communication Technology Readiness for Business Continuity.
5) 鈴木芳生，他：同期リモートコピーを用いたデータ欠損のないDRシステムの性能安定性の評価，情報処理学会論文誌データベース，Vol.1, No.1 (2008).
6) 吉田　浩：クラウドストレージ標準化の最新動向，情報処理，Vol.52, No.6 (2011).

(2012年3月30日受付)

西澤　格（正会員） itaru.nishizawa.cw@hitachi.com
1996年東京大学大学院工学系研究科博士課程修了．同年（株）日立製作所中央研究所入社．2002～03年米Stanford大客員研究員．ACM会員．現在，中央研究所にてデータ管理ミドルウェアの研究開発に従事．

藤原真二（正会員） shinji.fujiwara.yc@hitachi.com
1990年京都大学大学院工学研究科情報工学専攻修士課程修了．同年（株）日立製作所中央研究所入社．1998～99年米Stanford大客員研究員．ACM，IEEE，電子情報通信学会各会員．現在，ITプラットフォーム事業本部にてデータベースの研究開発に従事．

山本政行（正会員） masayuki.yamamoto.jw@hitachi.com
1996年京都大学大学院工学研究科情報工学専攻修士課程修了．同年（株）日立製作所システム開発研究所入社．現在，横浜研究所にてストレージシステムの研究開発に従事．

矢崎武己（正会員） takeki.yazaki.gq@hitachi.com
1995年東京大学大学院総合文化研究科修士課程修了．同年（株）日立製作所中央研究所入社．2002～03年米Stanford大客員研究員．電子情報通信学会各会員．現在，中央研究所にてクラウドネットワーキングの研究開発に従事．