モバイルコンピューティングとワイヤレス通信　20－16
(2002. 3. 8)

モバイルからのアクセスを含む広域監視システムの検討

桜井 鐘治† 前田 慎司† 黒田 正博†

†三菱電機情報技術総合研究所
〒247-8501 神奈川県鎌倉市大船 5-1-1

E-mail: †{saku, shinji, marsh}@isl.melco.co.jp

あらまし CPUの処理速度やメモリ容量，さらにフラッシュメモリ等の不揮発性記憶領域の容量が限られた，いわゆるThin Clientにより監視対象機器で定期的に発生する測定データを収集し，データセンタにネットワークを通じて蓄積する広域監視システムモデルを考えた場合に，ネットワークにインターネットや携帯電話等無線ネットワークを利用する場合には，計測データを確実にかつ効率的に送信することが課題となる．本稿では，この課題に対してデータ同期機構を適用した広域監視システムのプロトタイプと評価結果について述べ，評価結果から課題として明らかとなったデータセンタのスケーラビリティについてこれを確保する方式を提案する．

キーワード 広域監視システム，負荷分散，自動再構成

A Study of Wide Area Management System including Mobile Access

Shoji SAKURAI†, Shinji MAEDA†, and Masahiro KURODA†

†Information Technology R & D Center, Mitsubishi Electric Corporation
Ofuna 5-1-1, Kamakura, Kanagawa, 247-8501 Japan

E-mail: †{saku, shinji, marsh}@isl.melco.co.jp

Abstract A Wide Area Management system, which includes thin clients connected by Internet or wireless communication with a data center, needs an efficient and certain communication mechanism. This paper presents our prototype implementation that has a method using data consistency object modeling to transfer measured data from thin clients to our data center. The evaluation of our prototype showed that performance of data center is poor in large-scale system. For this reason, this paper also presents our proposal for this performance problem.

Key words Wide Area Management, Load Balancing, Auto Reconfigure

1. はじめに

近年のインターネットおよび携帯電話の急速な普及に伴い，無線通信または有線通信を利用して広域に分散する機器を監視するシステムを安価に構築することが可能となってきている．さらに，携帯電話機の高機能化に伴い，監視対象機器で発生した特定のイベントを携帯電話へ通知することや，携帯電話から監視対象機器で発生した測定データを参照することも可能となってきている．

一方，監視対象機器の多くは，ＣＰＵの処理速度やメモリ容量，さらにフラッシュメモリ等の不揮発性記憶領域の容量が限られた，いわゆる Thin Client であり，監視対象機器で定期的に発生する測定データの全てをローカルに蓄えておくことは不可能である．このため，監視対象機器から無線通信または有線通信を使ってアクセスが可能なネットワーク上にデータを蓄積するサーバを設置し，監視対象機器で発生した測定データを管理する方式が考えられる．この場合データを蓄積するためのサーバは，アクセスする通信手段を柔軟に選択でき，通信費を抑制することが可能であるインターネット上のデータセンタに置かれることが一般的である．

このように監視対象機器で発生したデータを無線および有線通信を利用してインターネット上のデータセンタに蓄積するシステムにおいては，以下の３つの要件を満足することが必要である．

(1) 発生するデータの確実な転送
(2) 広域通信を考慮した効率的なデータ転送
(3) データセンタのセキュリティを確保した通信

リソースが限られた監視対象機器で発生するデータは，定期的に測定されるデータで監視対象機器のローカルなストレージ領域が溢れる前に確実にデータセンタに転送する必要がある．また無線通信等の広域通信を利用する際には，通信の途絶によりデータ転送が完了しないうちに切断されてしまうことが容易に起こり得る．このため，通信の不安定さを吸収しかつ効率的なデータ転送を行うことが必要である．さらに，データセンタはインターネット上に構築されているため，悪意のある第三者からの攻撃を防ぐためにインターネットとの接続はファイアウォールを介して行う必要があり，監視対象機器とデータセンタ間のデータの転送もファイアウォールを介してデータセンタのセキュリティを確保して行う必要がある．

データバージョン管理方式を用いたデータ同期機構[1]は，複数のサイトに分散したデータの整合性を確保する方式であるが，上記の３つの要件を満たしている．

また，無線通信及びインターネットを利用することにより広域からのアクセスの自由度が大きくなるため，従来の広域監視システムより多数の監視対象機器を１つのデータセンタで管理することも期待される．このためには，大規模なシステムに対応したスケーラブルなシステムアーキテクチャが４つめの要件として要求される．

本稿では，データ同期機構を基本とした広域監視システムについて，プロトタイプの開発とその評価，さらに４つめの要件として上げたスケーラビリティの確保に向けた方式の検討について述べる．

以下，2.では，広域監視システムのアーキテクチャについて示し，3.と4.では，データ同期機構を適用したプロタイプ開発及びその評価について述べる．5.では，スケーラビリティについての必要な要件とこれを実現するための方式を提案する．最後に6.でまとめと今後の課題を示す．

2. 広域監視システムのアーキテクチャ

本アーキテクチャは，(1)データセンタ内に配置され，ファイアウォールを介してインターネットと接続されるセンタサーバ，(2)WAN(PSTN や ISDN の有線公衆網または携帯電話網)を介してデータセンタと接続され計測機器等による測定データをセンタサーバに対してアップロードするユーザサイト，(3)WAN(主に携帯電話網)を介してデータセンタと接続されユーザサイトで発生した特定イベントの通知を受信しユーザサイトで発生した測定データの参照を行う管理者サイト，の三つのコンポーネントから構成される．

図１に示すように，１つのデータセンタ内には複数のセンタサーバを配置することができ，そのそれぞれに対して測定データのアップロードを行うユーザサイトが存在する．ユーザサイトとセンタサーバの間では測定データのアップロード以外にユーザサイトでのデータの計測条件（計測の開始／停止，計測間隔，同期間隔）を変更する際に通信を行う．また，センタサーバには各監視対象のユーザサイト毎にユーザサイトで特定のイベントが発生した際にこれを通知する管理者サイトが登録されており，イベ

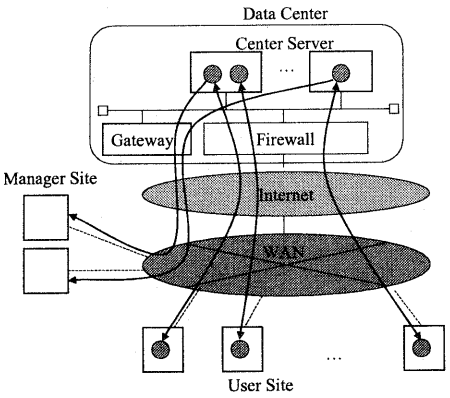

図 1 広域監視システムのアーキテクチャ

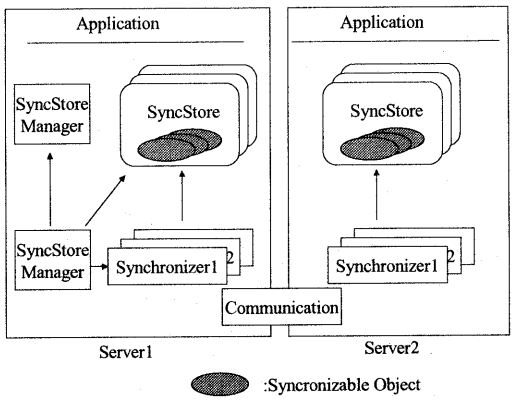

図 2 データ同期機構の構造

ントが発生した際に携帯電話網を使って通知を行う．さらに管理者サイトから特定のデータに対しての参照要求を受信した際にはこれに対して該当するデータを送信する．管理者サイトにはローカルにデータは記録されず，このためデータはセンタサーバから管理者サイトの向きにだけ送信される．管理者サイトとセンタサーバ間の実際の通信はゲートウェイを介して行われる．

3. プロトタイプ開発

3.1 データ同期機構

データ同期機構は，複数のサイトでデータレプリカ（以下，レプリカと呼ぶ）をもち，それらレプリカ間で整合性を保つ機構である．

本機構は，図2に示すように，アプリケーションから見れば，データを保存したり取り出したりするコンテナとしての SyncStore，その SyncStore に入れるデータとしての Synchronizable な Java Serializable オブジェクト，（以下，単にオブジェクトと呼ぶ）からなっている．オブジェクトの SyncStore への投入あるいは SyncStore からの取出しは，SyncStore の put()/get()メソッドで行う．オブジェクトは Reconcilable サブクラスか Diffable サブクラスの Synchronizable オブジェクトとして表現される．Reconcilable なオブジェクトは，同期を行う相手 SyncStore 内に存在する同一のオブジェクト識別子をもったオブジェクトとの間で，タイムスタンプを比較し，相手オブジェクトが自オブジェクトより新しいタイムスタンプを保持している場合，自オブジェクトを相手オブジェクトで置き換える．Diffable なオブジェクトは，同じく相手 SyncStore 内の同一識別子をもったオブジェクトとの間でタイムスタンプを比較し，一方のオブジェクトの最新タイムスタンプまでの更新ログを，他方のオブジェクトに適用する．

SyncStore 間の同期を行う場合，ネットワークの性質にあった Synchronizer を用いることになる．図2中 Syncronizer1 はファイアウォールを通過する HTTP を使用する Synchronizer であるが，Syncronizer2 はソケットを利用する Synchronizer である．

3.2 データ同期機構の適用

プロトタイプでは，データ同期機構をユーザサイトとセンタサーバとの間の通信に適用し，計測データのアップロードと計測条件の設定を実施した．

計測条件としては，ユーザサイト毎の条件として，センタサーバとの同期間隔が存在し，計測点毎の条件として，計測の開始／停止，計測間隔，イベント通知の上限値／下限値，イベント通知先メールアドレス，が存在する．

計測条件は，主にセンタサーバ側で設定され，ユーザサイト側に適用される．本広域監視システムではセンタサーバとインターネット間にはファイアウォールが存在するため，これを通過する HTTP の Synchronizer を用いる．また，ネットワークにはユーザサイト側からダイヤルアップ接続による回線も含むため，同期は常にユーザサイト側から開始される．したがって，センタサーバ側で変更した計測条件は，条件を変更する前の同期間隔に従いこの後次にユーザサイト側から開始される同期によりセンタサーバ側からユーザサイト側に適用される．このた

め，変更した計測条件が有効になるのは次回の同期終了後以降となる．

確実なデータ転送を行うには，1つのユーザサイトに対してプライマリとセカンダリ，さらにターシャリの複数のセンタサーバを同期相手として登録し，プライマリがダウンした際にはセカンダリに，セカンダリがダウンした際にはターシャリにフェイルオーバーさせる構成が可能である．なお，プロトタイプでは，フェイルオーバーの機能は実装していない．

計測データは，各計測点につき1時間毎の計測データを1つの時間データオブジェクトとしてSyncStore に投入する．このため，例えば計測間隔が10分の場合には，1つの時間オブジェクトは生成された後に5回更新されることになる．

本プロトタイプでは，追加するデータが数値データであり，個々の測定データはサイズが小さく，一般的な計測条件としては，“計測間隔<<同期間隔”の下で使用すること前提としている．

図3に測定データのサイズを4byteとした際の，Diffrableオブジェクトと Reconcilableオブジェクトによるデータ転送量の比率を，D／R(同期周期,計測周期)として示す．“計測間隔<<同期間隔”の条件下では，Diffable なオブジェクトを用いて更新ログにより同期を行うよりも，Reconcilable オブジェクトを用いて同期を行う方が通信量は小さくなる．

また，Diffable なオブジェクトを用いた場合にはデータとは別に更新ログを保持するための領域が必要となるため，この領域を必要としないReconcilable オブジェクトの方が Thin Client であるユーザサイトには適している．これらより，本プロトタイプでは Reconcilable を用いて計測データのアップロードを行っている．

ユーザサイトの各計測点について，計測された値が計測条件で設定された上限値を上回るか下限値を下回った場合には，ユーザサイトの各計測点につい登録されているこれを通知する管理者サイトに対してイベントの発生が通知される．図4の(a)(b)にプロトタイプでの管理者サイトにおけるイベント通知を受信した際の画面を示す．

図4の(c)に示すように、管理者サイトのユーザは，受信したイベント通知より，センタサーバにアクセスし，イベントを発生した計測点のさらに詳細な計測データを参照することができる．

セキュリティについてはチャレンジ･レスポンスを使用している．同期開始時に，ユーザサイトから

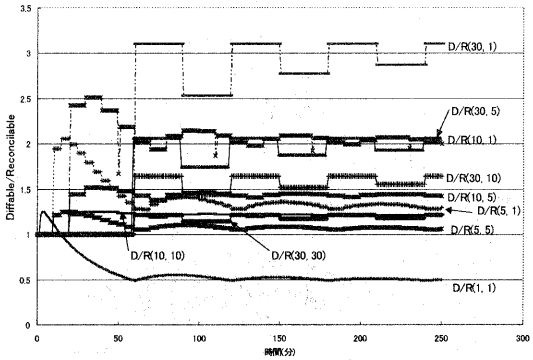

図 3 Reconcilable と Diffable のデータ量の比較

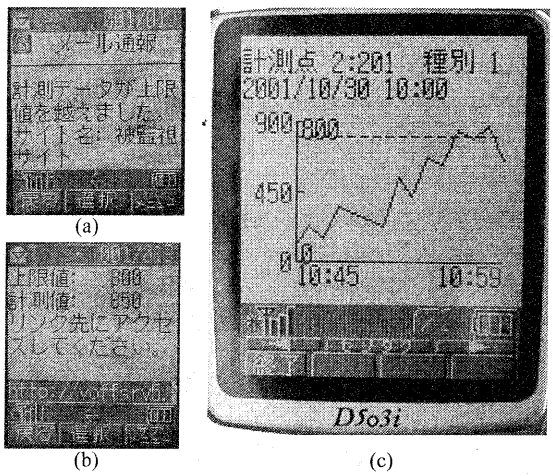

図 4　管理者サイトへのイベント通知とデータ参照

接続されたセンタサーバからチャレンジ値を送信し，ユーザサイト側でこれをもとにレスポンス値を生成し返信する．センタサーバ側で生成したレスポンス値とユーザサイト側で生成したレスポンス値が一致した後にはじめて計測データの同期を行う．

4. プロトタイプの評価

4.1 評価環境

プロトタイプのスケーラビリティを評価する環境のについて記述する．

プロトタイプは，ユーザサイトに，ノートPC(Pentium133Mhz,メモリ 32MB)センタサーバにPC サーバ(PentiumII 300Mhz,メモリ 96MB)，管理者サイトに Java 搭携帯電話を使用して構築した．

4.2 評価結果

プロトタイプの評価結果を図5と図6に示す。

計測点を 1, 4, 64, 255 の 5 通り，計測間隔を 2.5, 5, 10 分の 3 通り，同期間隔を 15, 30 分の 2 通りの計 30 通りについてユーザサイトとセンタサーバ間で同期処理を行った際のセンタサーバの CPU 負荷を測定し，計測点数と CPU 負荷の関係を表したグラフを図 5 に示す．CPU 負荷は同期処理が行われた時刻の近傍 5 分間における平均である．

図 5 より，CPU 負荷は計測点数に対して比例して増加し，計測間隔の影響を受けないことが分かる．これは，計測間隔の変更により各計測点の計測データオブジェクトのサイズは変化するが、センタサーバが処理するオブジェクト数は変わらないため同期処理の負荷には影響を与えていないものと考えられる．計測点数を m とすると m に対する CPU 負荷 L1 は，式①で表される．

$$L1 \cong 0.036 \times m + 3.13 \quad \cdots\cdots\cdots\cdots ①$$

測定結果は同期時刻近傍の 5 分間の CPU 負荷であり、同期間隔 ts を短くすると単位時間内にセンタサーバが同期するオブジェクトの数がこれに反比例して増加することが予想される．このため実際には①の式は、

$$L1 \cong 0.036 \times m \times \frac{5}{ts} + 3.13 \quad \cdots\cdots\cdots\cdots ②$$

となるものと予想される．

次に，計測点を 255，計測間隔 2.5 分，同期間隔 15 分として，ユーザサイト数を 1 から 3 まで変化させて同期処理を行った際のセンタサーバの CPU 負荷を測定し，ユーザサイト数と CPU 負荷の関係を表したグラフを図 6 に示す．これも CPU 負荷は同期処理が行われた時刻の近傍 5 分間における平均である．

図 6 より，CPU 負荷はユーザサイト数に比例して増加していることが分かる．このことも，センタサーバと同期するオブジェクトの数が増加することによるものと考えられる．

4.3 広域監視システムの問題点

データ同期機構を用いた広域監視システムでは，センタサーバと同期を取るように設定されたユーザサイトの数は頻繁に変更されることは少なく，比較的長い時間幅で一定であると考えられる．センタサーバと同期を取るように設定された各ユーザサイトは，設定された同期間隔でセンタサーバとの同期を実行する．センタサーバとの同期が完了したユーザサイトは，前回の同期開始から設定された同期間隔の経過後に，次の同期の実行を開始するため，本システムにおけるセンタサーバとユーザサイトの関係は，ユーザサイトをジョブの発生源（呼源）としたセンタサーバに対する有限呼源待ち行列(Finite Source Queue)[1]と見なすことができる．

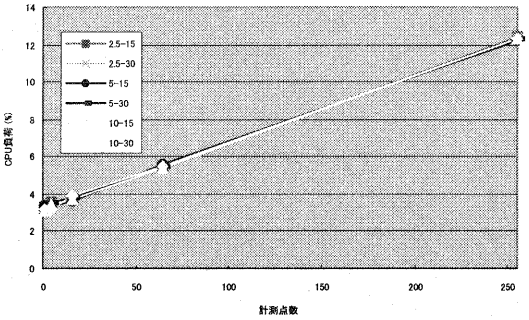

図 5 計測点数の変化に対する CPU 負荷の変化

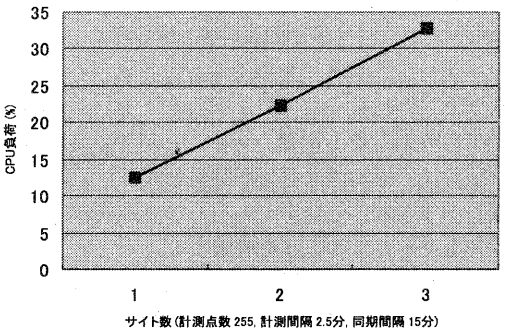

図 6 ユーザサイト数の変化に対する CPU 負荷の変化

ユーザサイト数 N，各ユーザサイトにおける同期完了から次の同期開始までの待ち時間の平均を $1/\lambda$ (秒)であるとする．各同期のセンタサーバにおける処理時間を $1/\mu$ (秒)，センタサーバの CPU 利用率を ρ とすると，センタサーバの処理能力は，センタサーバに接続するユーザサイト数 N によらず一定であるので，センタサーバとの同期処理を完了し次の同期開始まで待ちの状態に移行するユーザサイトの時間当たりの平均は $\mu\ \rho$ となる．したがって各ユーザサイトが同期を実行し次の同期を実行する回数の平均は，単位時間当たり $\mu\ \rho/N$ となる．

センタサーバでの同期の平均応答時間(待ち時間+処理時間)を T で示すと，各ユーザサイトがセンタサーバとの間を 1 回巡回するのに要する時間の平均は，$T+1/\lambda$ であるので，

$$T + \frac{1}{\lambda} = \left(\frac{\mu\rho}{N} \right) \quad \cdots\cdots\cdots\cdots ③$$

N が非常に大きい場合には，センタサーバの利用率 ρ は 1 に近くなることが予想される．③に ρ=1 を代入すると次の式④が得られる．

$$T \cong \frac{1}{\mu}N - \frac{1}{\lambda} \quad \cdots\cdots④$$

式④より，N が非常に大きい状態では，N が 1 増加すると T が $1/\mu$(処理時間分)だけ増加し，急激に応答性能が悪化することが理解できる．

5. スケーラビリティの確保

5.1 スケーラビリティ確保のための方式

スケーラビリティを確保するためには，N が非常に大きい状態では，$1/\mu$ を抑えるか，λ を小さくし T の立ち上がりを遅らせることが必要である．

前者は，平均の同期処理時間を短くすることを意味し，後者は，センタサーバへの同期要求の平均到着率を小さくすることを意味する．

データセンタに設置するセンタサーバの台数を 1 台とすると，センタサーバへの同期要求の平均到着率を小さくすることは，ユーザサイトの同期間隔を大きくすることに他ならず，システム利用上の制限となるため現実的でない．このため，実際には，データセンタに設置するセンタサーバを複数台とし 1 台あたりへの同期要求の平均到着率を小さくすることが必要となる．

以降では，後者の方式により，システムのスケーラビリティを確保する方式を検討する．

5.2 スケーラビリティについての要件

データセンタに複数のセンタサーバを設置する場合には，センタサーバ間で負荷の偏りが生じることが予想される．システム全体でリソースを有効に利用するためにはセンタサーバ間で負荷の移動を行い各センタサーバでの負荷を平準化することが必要である．

このセンタサーバ間での負荷の移動を行うにあたっては，以下の要件を満足することが必要である．

(1) ユーザサイトで発生したデータの確実なセーブ
(2) 負荷移動におけるオーバヘッドの抑制
(3) 自動構成変更

ユーザサイトの同期の負荷は，計測条件を変更することにより変動するが，実際の運用においては比較的長い時間幅で一定であると考えられ，新たなユーザサイトの登録／削除か計測条件を変更した場合にのみ，結果として負荷が変動する．

5.3 システムアーキテクチャ

2 台のサーバを用いてシステムのスケーラビリティを確保する場合には，サーバに対するディスク装置の関係により，(1) 1 台のデュアルポートディスク装置を 2 台の出サーバから使用する，(2) 2 台のデュアルポートディスク装置を 2 台のサーバから使用する，(3)ディスク装置を 1 台の有する 2 台のサーバをネットワークで接続して使用する，(4) 2 台のサーバと 2 台のネットワークディスク装置をネットワークで接続して使用する．の 4 つの構成が考えられる[3]が，本稿では，スケーラビリティを確保するための問題点が CPU ネックであることより(3)の構成を検討する．なお，ディスク装置は単体で RAID とミラーリングによる冗長化構成が取られており，一度書き込まれたデータはシステムがダウンした場合にも正常に保存されているものとする．

5.4 プロトコル設計

負荷情報の共有

センタサーバ間での負荷の移動は，負荷の高いセンタサーバから負荷の低いセンタサーバへユーザサイトの接続先を変更することで行う．この負荷の移動処理では少なくとも移動元か移動先のどちらかのセンタサーバが他方のセンタサーバの負荷情報を把握している必要がある．このためには，(1)特定のサーバ（ディスパッチャ）が全てのサーバの負荷情報を管理する方法と(2)全てのサーバの負荷情報を各サーバが記憶する方法と(3)負荷の高いサーバの情報を他のサーバが記憶する方法または(4)負荷の低いサーバの情報を他のサーバが記憶する方法の 4 つに分かれる．(1)は，ディスパッチャに高い信頼性が必要になる．(2)はクラスタベースメールシステムの Porcupine[4]で採用されている方式であるが負荷の高いサーバでも定期的に負荷情報の管理のための処理を行う必要がある．(3)はシステムの負荷が全体的に高くなった際に極端な性能劣化を招く恐れがある．このため本稿では(4)を基本とした負荷移動プロトコルを検討する．なお，複数のセンタサーバは同一のデータセンタ内の同一 LAN に接続されていることを前提とし，ブロードキャストする負荷情報の通知が行えるものとする．

負荷の定量化

ユーザサイトとの同期によるセンタサーバの負荷は，ユーザサイトの測定条件により変化する．

このことより，ユーザサイトとの同期によるセンタサーバの負荷は次の式⑤で示されると予想される．

$$L=\sum_{i=1}^{M}\left(\frac{1}{ti}+\alpha\right)\times\frac{1}{ts}+\beta \qquad \cdots\cdots⑤$$

ここで，M はユーザサイトにおける計測点数，ti は計測点 i における計測間隔，ts は同期間隔，α はオブジェクトの同期に要する基本オーバヘッド，β はサイトの同期に要する基本オーバヘッドである．

プロトタイプの評価結果より，ti による影響は無視できるとすると，式⑤は

$$L=M\alpha\frac{1}{T}+\beta \qquad \cdots\cdots⑥$$

で表され，プロトタイプの場合には α には 0.18(= 0.036 * 5)，β には 3.13 がそれぞれ当てはまる．

接続先センタサーバの変更

ユーザサイトの接続先センタサーバを変更する処理の概要を図 7 に示す．接続先センタサーバの変更は，負荷の低いサーバからブロードキャストする負荷情報により全てのセンタサーバに通知される．これには送信元のセンタサーバの負荷情報が含まれる．負荷情報を受信したセンタサーバのうちで現在負荷が高いものから，負荷情報の送信元に対して，現在の負荷情報を含む移動要求が送信される．移動要求を受信したセンタサーバでは，最も負荷が高いセンタサーバに対して移動許可を送信し，これを受信したセンタサーバでは，負荷情報の差と定量化したユーザサイトとの同期負荷より接続先を変更するユーザサイトを決定し，移動先センタサーバに対して移動実行を送信する．これには，接続先を変更するユーザサイトの接続情報が含まれる．移動実行を受信したセンタサーバでは，受け取り確認として完了通知をブロードキャストする．これには，接続先を変更したユーザサイトの ID と現在の接続先センタサーバの ID を含む．なお，接続先の変更となったユーザサイトには，接続先の変更情報が完了通知を受け取った後の次回の同期で通知される．また，ユーザサイトの新たな接続先となったセンタサーバの負荷変動は，さらにこの後の新しいセンタサーバとの初めての同期の後で確定する．この間，移動に関与しなかった他のセンタサーバからの移動要求は全て廃棄される．

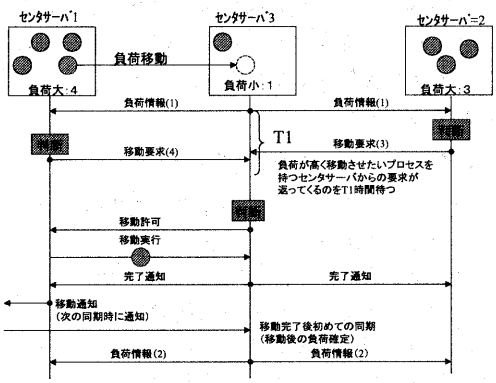

図 7 接続先センタサーバの変更処理

ユーザサイトの情報の管理

ユーザサイトの情報は，HLT(Home Location Table)と DLT(Data Location Table)の 2 つのテーブルを用いて管理する．HLT は，ユーザサイト ID と現在の接続先であるセンタサーバとの組を登録したテーブルで全てのセンタサーバで，完了通知に含まれるユーザサイト ID とセンタサーバ ID を元に更新する．なお，完了通知のブロードキャストにより更新をするため，これの受信に失敗したセンタサーバは更新されずに古いままとなるが，この場合には，古い情報を使ってアクセスされたセンタサーバより最新の情報を提供することにより更新を行う．なお，移動実行を送信するセンタサーバで完了通知の受信に失敗した際には，移動実行を再送する．DLT は，ユーザサイトの計測データについてどの時間のデータがどのサーバに蓄積されているかを示す．DLT は移動実行に含まれ，移動元のセンタサーバから移動先のセンタサーバに送られる．

新規センタサーバの追加

データセンタのセンタサーバの負荷が全体的に高いと判断された場合には，新規センタサーバをシステムに追加することができる．この場合，新たに追加されるセンタサーバは起動後に，負荷情報を現在負荷 0 でブロードキャストするだけでシステムに組み入れられる．

センタサーバの切り離し

センタサーバの定期的メンテナンスや機種交換のためにセンタサーバをシステムから切り離す場合に

は，現在の負荷を強制的に最大にセットすることにより現在このセンタサーバを接続先としているユーザサイトの接続先を他のセンタサーバに変更することができる．

センタサーバのダウン時の処理

センタサーバのダウンが発生した場合には，ダウンしたセンタサーバを接続先としていたユーザサイトの同期が失敗する．このため，ユーザサイトでは，セカンダリのセンタサーバに接続先を変更し，データのアップロードを行う．この場合，センタサーバ側では接続先の変更の処理を行っていないユーザサイトの接続を受け付けることになるため，これによりセンタサーバ側では，ユーザサイトの現在の接続先となっているセンタサーバの異常を検出し，センタサーバ ID をユーザサイト ID と共に WCS(Watching Center Server)へ登録する．ダウンから復旧したセンタサーバは，新規に追加されたセンタサーバと同じく現在負荷 0 で負荷情報をブロードキャストすることでシステムに再び組み入れられる．これを受けたフェイルオーバー先のセンタサーバは現在負荷最大で，移動要求を復旧したセンタサーバに送信し，接続先の切り戻しを行う．

なお，管理サイトへのイベント通知を生じる閾値を越えた計測データのアップロード直後にセンタサーバがダウンする場合には，管理サイトへのイベント通知が最悪ダウンしたセンタサーバの普及以降にまで遅くなる場合も想定され，これを防ぐためにフェイルオーバー直後には，最後のアップロードをセカンダリのセンタサーバに対して再度実行できるよう同期 2 回分のデータを保持するリソースがユーザサイト側に必要である。

管理者サイトとの通信

管理者サイトとの通信は，センタサーバからのイベント通知と管理者サイトからのデータ参照時に行われるが，ユーザサイトの接続先が変更された場合に，前者は新しく接続先となったセンタサーバから行われる．一方後者については，前の接続先のセンタサーバに対して行われるが，古い情報を使ってアクセスされたセンタサーバよりの HLT の情報に基づき最新のセンタサーバの情報を提供することにより管理者サイトの接続先の更新を行う．

6. まとめ

本稿では，広域監視システムの必要要件を検討し，これに対しデータ同期機構を広域監視システムに適用したプロトタイプの開発について述べた．データ同期機構を使用することにより，発生するデータの確実な転送，広域通信を考慮した効率的なデータ転送，データセンタのセキュリティを確保した通信が行えるが，プロトタイプの評価結果からセンタサーバにおける同期処理の CPU 負荷はデータ量にほとんど影響されず単位時間あたりの同期回数に比例すること明らかとなった．このことは大規模なシステムを構築する際に 1 つのデータセンタで構築できるシステムのスケーラビリティが問題となることを意味する．

このため，本稿では，複数のセンタサーバ間でユーザサイトの接続先を変更し負荷移動を行うプロトコルの提案を行った．本プロトコルにより複数のサーバ間での負荷移動とサーバの構成変更さらにサーバダウン時のフェイルオーバーやその復旧が行える．

ただし，現状はユーザサイトからの計測データのアップロードについてこれを停止しないことと，接続先変更処理の負荷を押えることを最優先としている．このため，アップロードした計測データへの参照は，センタサーバがダウンしている場合には停止するケースが存在する．また，センタサーバの処理能力が個々に異なるヘテロジニアスな環境に適用した場合には，CPU の利用率だけでの接続変更先の選択では，接続先の変更後に負荷状態の逆転が起きる場合もあり，センタサーバの処理速度を考慮した負荷情報の定量化が必要である．

今後は，これら問題の解決方法を検討するとともに，今回提案のプロトコルの実装及び評価を行っていく予定である．

文　献

[1] 黒田正博,井上淳,渡辺尚,水野忠則，“楽観的データ整合性モデルを用いた放浪型メッセージングシステム”,信学論(B),vol.J82-B,no.5,pp.827-838,Jan.1999

[2] 亀田壽夫, 紀一誠, 李頡, 性能評価の基礎と応用, pp.45-49, 共立出版，1998

[3] A.D. Birrell, A. Hisgen, C. Jerian, T. Mann, G. Swart, The Echo Distributed File System, SRC Research Report 111, System Research Center, Digital Co., Palo Alto, CA, 1993

[4] Yasushi Saito, Brian N. Bershad, Henry M. Levy, "Manageability, Availability and Performance in Porcupine: A Highly Scalable, Cluster-based Mail Service", 17th ACM Symposium on Operating System Principle, Kiawah Island Resort, near Charleston, SC, Dec 1999