

戦略的創造研究推進事業 CREST

研究領域「情報社会を支える新しい高性能情報処理技術」

研究課題「ディペンダブルで高性能な  
先進ストレージシステム」

## 研究終了報告書

研究期間 平成 15 年 10 月～平成 21 年 3 月

研究代表者：横 田 治 夫

東京工業大学 学術国際情報センター・教授



## § 1 研究実施の概要

### (1) 研究の背景と目的

情報技術の目覚ましい発達により、個人や組織のシステムが保持するデータの種類や数・容量が爆発的に増加している。これからの情報社会においては、そのような膨大なデータを確実に蓄積し、安定して供給することが強く求められている。

これまでの情報システムでは、そのような多種・大量のデータの蓄積・供給のために、専門家が関わって緻密に管理してきた。しかし、一人の人間の管理能力には自ずと限界があり、それが多種・大量のデータを保持するシステムの構築や運用における管理コストの増大を招いている。特に、システム中に故障や負荷の偏りが発生した場合に、情報提供サービスの品質の保持するための管理がますます困難となってきている。さらに、利用者が必要とする情報を探し出し、有効に活用するための格納方法や検索方法もデータの多種・大量化に対応する必要も生じている。

本研究では、これからの情報社会に求められる情報システム基盤としてのストレージ機能を提供することを目的に、

- ・ データの数・容量が増加しても容易にシステムを拡張し性能を向上させ、
- ・ システム規模が拡大しても低管理コストでコンテンツを安定して提供し、
- ・ 故障や負荷の偏りが発生しても提供するサービスの品質を保持し、
- ・ 蓄積された多種多様なコンテンツの特徴を活かした管理・検索を可能とする

ための要素技術を確認することを目指して研究を行ってきた。そのような技術が提供されることによって、ストレージ管理のコストが抑えられ、必要とする情報が的確に提供される、より快適な情報社会を実現することができる。

### (2) 従来技術の問題点

これまでの技術で大量データを情報システムに蓄積するためには、高信頼化のためにパリティ計算を用いた RAID (Redundant Array Inexpensive Disks) を構成要素に SAN (Storage Area Network) や NAS (Network Attached Storage) といったネットワークストレージを構築し、その中で集中管理によるデータ分散配置やアクセス管理が行うことが主であった。そのようなシステムにおいては、集中管理部分へのアクセス集中による性能低下、あるいは集中管理部分が故障した場合に全体のサービス停止といった性能と信頼性におけるアキレス腱が存在した。また、故障が発生した場合に RAID 特有のデータ再構築処理によるサービス品質の低下、アクセス集中による性能の低下といった問題もあった。このため、システム中に故障やアクセス偏りが発生した場合に、管理者のスキルに依存した制限されたサーバ側でのストレージ管理を行う必要があった。さらに、蓄積されたコンテンツの検索はサーバ側で行われることから、格納されたコンテンツの特徴を活かしきれず、的確に必要なとする情報を提供できないという問題点もあった。これらの様子を図1に示す。

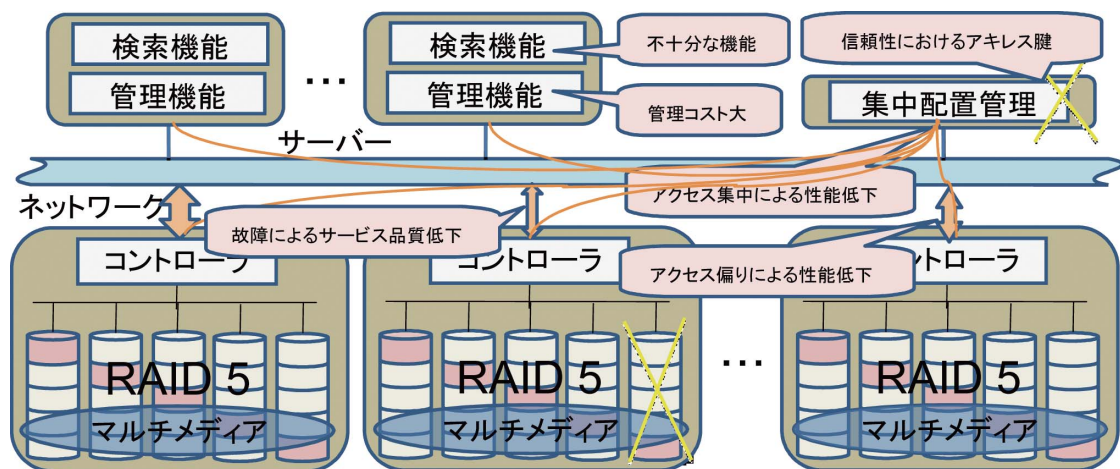


図1 従来の技術による情報蓄積システムの構成とその問題点

### (3) 本研究のアプローチとその効果

本研究では、情報社会に求められる情報システム基盤として、前述の問題点を解決するストレージ技術の提供を目的としてきた。システム拡張性を持たせ管理コストを削減するため、従来はサーバで行われてきた管理・検索処理をストレージ側で分散して行うアプローチを取ることにした。

データ配置管理をストレージ側で分散して行うことにより、故障や負荷偏りがあった場合にサービス品質保持のためストレージ間で自律的にデータを移動することが可能となり、サーバ側での管理コストを下げるができる。また、分散制御の下で、RAID のような格納データのコード化を行わないことにより、故障発生時や修復時にもバンド幅を確保でき、データ移動も低コストで可能になる。さらに、ストレージ側で暗号化処理を行うことで、低コストで安全性の高いセキュリティ管理を行うことが可能になる。一方、管理機能だけでなく検索機能もストレージ側に取り込むことにより、蓄積されたコンテンツの近くで検索処理を行うことができるようになり、XML 等コンテンツの特徴を活かした検索や管理が可能になる。本研究のアプローチのイメージを図2に示す。

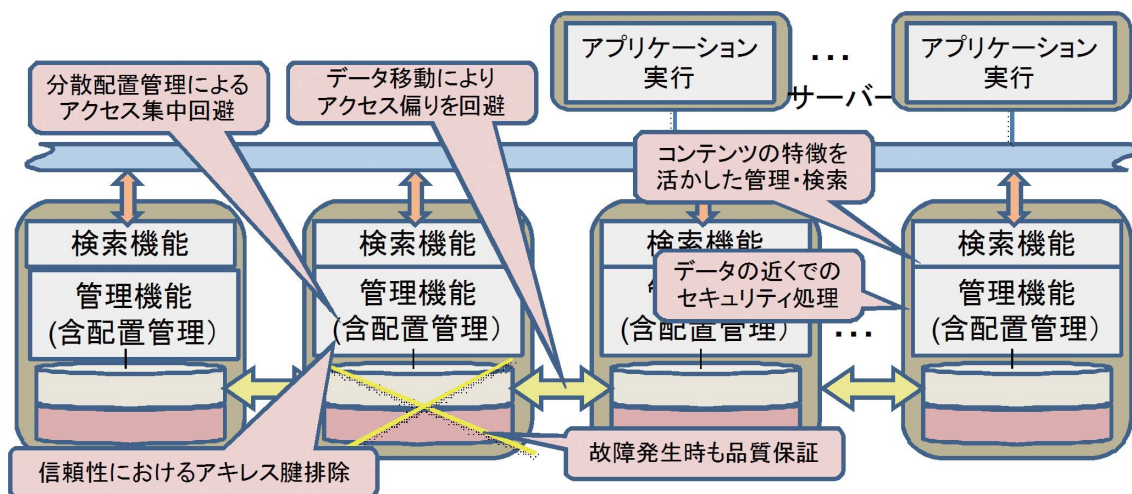


図2 本研究のアプローチのイメージ

### (4) 提案技術と成果

実際に前述のアプローチを実現するためには、様々な技術を新たに導入する必要がある。まず、分散して格納されたデータに対して、データが移動してもシステム全体に影響を与えることなくローカルに処理しながら、高速のアクセスを可能とするための検索・更新性能の優れた分散ディレクトリ構造が必要となる。さらに、データ移動のコストが少なく、故障が発生しても十分なバンド幅を確保できる高信頼なデータ配置が必要となる。

このため、まず検索性能と更新性能が両立可能な分散ディレクトリとして Fat-Btree を採用した。さらに、データ移動時や故障発生時に有利なデータ配置としてプライマリ・バックアップデータ配置を採用した。それら Fat-Btree とプライマリ・バックアップ構成の特徴を活かし、先進ストレージ構築に必須の以下のような技術を提案し、160 ノードの実験環境等を使ってその効果を実証してきた。

- 分散環境において複数利用者の同時多数アクセスに対応するためには、分散ディレクトリ用の高効率な同時実行制御手法が必要である。特に分散ストレージにおいては、同時実行制御のためにストレージノード間の通信が多発すると、アクセス性能が極端に劣化し、実際に分散ストレージを利用することが不可能になる。このため、従来の単一ノード用のディレクトリに対する同時実行制御とは全く異なった、通信コストをできるだけ抑えた同時実行制御手法を提案した。多数のストレージノードを用いた実験によって、提案手法がこれまでの手法に比べて非常に性能が高く、システムの拡張性に優れていることを実証した。
- そのような分散ストレージにおいては、各ストレージノード間でデータの一貫性を保つために分散コミットプロトコルを取り入れることも必要である。従来の分散コミットプロトコルでは、同時実

行制御と同様に通信コストが高く、分散ストレージとして使い物になる性能を出すことができなかった。このため、Fat-Btree とプライマリ・バックアップデータ配置の特徴を活かした新たな効率的な分散コミットプロトコルを提案した。多数のストレージノードを用いた実験によって、提案プロトコルが従来のプロトコルと比較して非常に効率的であることを実証した。

- c) 分散ストレージのノード数を増やして、ノード数に応じた性能を確保するためには、各ノードのアクセスや容量のバランスを取ることが極めて重要である。しかし、一般的なデータ配置手法では、アクセスもしくは容量の一方をバランスさせるともう一方のバランスが崩れるという問題点がある。これに対して、プライマリ・バックアップ構成の特徴を活かして、アクセスバランスと容量バランスを両立させるデータ配置を提案した。このデータ配置手法は、さらに故障が発生した後のリカバリ時間とその間のサービス品質を向上させるという効果もある。バランス両立の効果とリカバリ時の時間短縮およびサービス品質向上の効果をシミュレーションで確認した。
- d) ストレージ中に同じファイルの複数のバージョンを格納しておくことも多い。そこで、差分バージョンを有効利用したアクセスバランスと容量バランスの両立方法の提案を行った。現在主流のスナップショット管理においてバージョンングを行うようなストレージシステムにおいて、バランス確保のためには有効な手法である。シミュレーションによりその効果を確認した。
- e) 上記のような均衡化処理を行うためには、ストレージノード間のデータの移動が必ず発生する。特にアクセスバランスを取る場合、アクセスの頻度が高いストレージノードからデータを相対的にアクセス頻度の低いストレージノードに移動することになるが、データ移動自身がアクセス負荷をかけるために、アクセス頻度の高いストレージの負荷がさらに上がってしまい、一時的にサービス品質を劣化させる原因となりうる。特に、近年ストレージの用途として重要視されている動画提供サービスのような場合に、データ移動によって動画の提供が途切れることがあることは望ましくない。そこで、プライマリ・バックアップ構成の特徴を利用して、サービス品質の低下をできるだけ抑えるノード間データ転送手法に関する提案を行い、シミュレーションによりその効果を確認した。
- f) 安全に安定してデータを提供するためにはセキュリティ管理も必要である。安全性を高めるためには、情報がストレージ内に暗号化して格納されていることが望ましい。しかし、複数の利用者が情報を共有する場合に、メンバーに変更があると再暗号化が必要となる。そのような再暗号化はサービス提供の一時停止を招く可能性がある。そこで、サービスを継続して提供するために、プライマリ・バックアップ構成を利用した再暗号化手法を提案し、その効果を評価した。
- g) ストレージ中に格納されたコンテンツの特徴を活かした検索・管理を行うためには、コンテンツ自身やコンテンツのメタデータを記述する XML を効率よく扱うことが必要となる。まず木構造である XML を効率よく格納し、木構造を前提に必要な情報を取り出すために、XML の各ノードの効率的なラベル付け手法を提案し、評価を行った。次に、ラベル付を前提に、複数のキーワードに合致する部分木を高速に取り出すための索引構造を提案し、評価を行った。また、類似した部分木を探し出し、情報を統合する手法に関しても提案・評価を行った。さらに、XML を分散して格納して、高速に検索する手法に関しても提案・評価を行った。
- h) 提案するストレージ装置の利用範囲を拡大するためには、広域分散を考慮したストレージネットワーク技術も必要となる。広域分散環境で安全に利用するための VPN を想定し、信頼性と性能を向上させる手法に関しても検討を行った。

以上述べたように、従来はサーバで行われてきた管理・検索処理をストレージ側で分散して行うアプローチを実現させるために必要となる一連の技術を提案し、実験等によってその有効性を示し、トップレベルの国際会議や各種論文誌等でその成果を多数公表し、高い評価を得てきた。

さらに、提案技術の組み合わせによって、実際にこれまでにないストレージ環境が実現できることを示すために、2 種類の自律ディスクプロトタイプハードウェアとその上のプロトタイプソフトウェアを開発した。情報処理学会の「わくわく IT」等においてそのプロトタイプを用いて、Windows からプロトタイプ中に格納されたデータにアクセスさせながら、均衡化や故障を発生させた時のリカバリのデモンストレーションを行い、提案技術の効果をよりわかり易く公開してきた。

また、国内外の技術動向調査を行うと同時に、研究進捗状況の確認とトップレベルの専門家からのコメントの収集を目的に、2004、2005、2007 年の 3 回の国際ワークショップを開催した。



## § 2 研究構想及び実施体制

### (1) 研究構想

新しい情報社会では、多種多様の膨大な情報を確実に蓄積し、必要な情報を安定して高速に取り出すことが極めて重要となる。本研究では、その要件を満たすためのディペンダブルで高性能なストレージシステムの構成技術の提供を目的に研究を進めてきた。

そのような先進的なストレージシステムを構築するため、ストレージシステム内の各構成要素であるストレージ装置を高機能化してクラスタを構成し、そのストレージクラスタ内での自律的な制御によって、ストレージに対するアクセス負荷の均衡化や容量の均衡化と故障対策を行う自律ディスクのアプローチを研究開始時から採用した。

このアプローチを実現するに当たり、当初、

- ・ ストレージクラスタ内のアクセス負荷バランスと容量バランスの両立
- ・ システム内負荷均衡化・耐故障処理と通常サービスの品質の両立
- ・ 格納されるコンテンツの特徴を生かした柔軟なデータ管理

といった点にまず主眼を置き、その実現のための技術の確立と、それらに基づく先進ストレージシステムの実用化を目標とした。

このため、先進ストレージ研究統轄・推進グループ、高度メディア蓄積・管理研究グループ、システムアーキテクチャ研究グループ、コンテンツ対応メタデータ研究グループが互いに協力して研究を推進し、その提案内容の位置づけを明確化するためにストレージ技術動向調査グループが国内外の技術動向の調査を行うこととした。その中でも、先進ストレージ研究統轄・推進グループは上記3項目全体に関する検討を行い、高度メディア蓄積・管理研究グループとコンテンツ対応メタデータ研究グループが格納されるコンテンツの特徴を生かした柔軟なデータ管理に焦点を当てることにした。

研究実施に当たり、プロジェクト期間を3つのフェーズに分け、第一フェーズで基本アイデアの検討・提案、第二フェーズで実験・評価とそれに基づく改良を行い、第三フェーズで実用を想定した展開を行うこととした。実験環境としては、提案する手法の100台以上の台数効果を確認するために、接続形態が一様になるように構成した160ノードからなる自律ディスクシミュレータを準備した。また、実現する先進ストレージシステムの有効性を目に見える形で示すために、プロトタイプシステムのハードウェアとソフトウェアの開発を開始した。

研究の進展に伴い、アクセス負荷バランスと容量バランスの両立のための画期的なアイデアが生まれると同時に、分散アクセスを効率的に行うための同時実行制御手法や分散アトミックコミットの効率的な手法も考案され、上記実験設備を用いてその効果を実証することができた。また、放送局におけるコンテンツ対応メタデータに対する検討も進み、コンテンツ対応メタデータ研究グループは当初計画の検討が終了したため、平成19年度からはグループとしての活動は終了することとなった。

一方、多種多様の膨大な情報を確実に蓄積し、必要な情報を高速に取り出すという観点から、コンテンツおよびコンテンツを記述するためのXMLの重要性が増してきたため、新たな研究項目としてストレージにおけるXMLの格納および検索を研究対象に加えることにした。また、安全に安定してデータを提供するためにはセキュリティ管理の必要であることから、それまでの提案手法と組み合わせたストレージとしてのセキュリティ管理も研究対象に含めることにした。さらに、分散ストレージとしての適用範囲を広げるために、広域分散環境に関する検討の必要性が生じてきたため、平成18年度から、ストレージネットワーク研究グループを新たに加え、広域分散環境でストレージを安全に利用するためのVPNを想定し、信頼性と性能を向上させる手法に関しても検討を行うことにした。

チーム内の各グループの研究進捗の確認等を行うために、チーム内ミーティングを頻繁に開催するとともに、2004年、2005年、2007年の3回にわたり海外において国際ワークショップを開催した。国際ワークショップ等を介して、積極的に研究内容を広く公表し、海外のトップレベルの専門家

からのコメントの収集を行い、研究の位置づけを明確化することに努めた。

また、提案技術を組み合わせた実用的な先進ストレージシステムのイメージを示すために、試作したプロトタイプシステムを用い、CREST シンポジウムや他の機会を使って積極的に先進ストレージシステムのデモンストレーションを行ってきた。

## (2) 実施体制

グループ名	研究代表者又は主たる共同研究者氏名	所属機関・部署・役職名	研究題目
先進ストレージ研究統轄・推進グループ	横田治夫	東京工業大学・学術国際情報センター・教授	ディペンダブルで高性能な先進ストレージシステムの研究の統括・推進、および先進ストレージシステムにおけるデータ管理機能の検討、実装、評価
高度メディア蓄積・管理手法研究グループ	宮崎 純	奈良先端科学技術大学院大学・情報科学研究科・准教授	先進ストレージシステムにおける高度メディア蓄積・管理手法に関する研究
ストレージネットワーク研究グループ	小口 正人	お茶の水女子大・理学部情報科学科・教授	先進ストレージシステムにおけるストレージネットワークの性能とディペンダビリティに関する研究
システムアーキテクチャ研究グループ	太田 光彦	富士通株式会社・ストレージシステム事業本部・部長付	自律ディスクプロトタイプ機の開発試作及び自律ディスク実用化に向けた機能実験環境構築。自律ディスクの実環境実験と評価
コンテンツ対応メタデータ研究グループ	和泉 吉則	NHK 放送技術研究所・主任研究員	先進ストレージシステム利用に向けた放送局用コンテンツ対応のメタデータに関する研究
ストレージ技術動向調査グループ	中村 隆喜	日立製作所・システム開発研究所・研究員	先進ストレージシステム研究遂行のためのストレージ技術動向調査

### § 3 研究実施内容及び成果

#### 3.1 先進ストレージ構築技術に関する研究(東京工業大学 先進ストレージ研究統轄・推進グループ)

##### (1)研究実施内容及び成果

本研究では、ディペンダブルで高性能な先進ストレージシステムを実現するため、各ストレージ装置に処理能力を持たせた自律ディスクを発展させた形で研究を行ってきた。以下では、まず提案手法の前提となる自律ディスクの基本概念を説明する。

続いて、自律ディスクの要素技術であるFat-Btree を前提とした効率的な同時実行制御手法、隣接ノードへのバックアップを有効利用した分散アトミックコミットプロトコル、さらにプライマリとバックアップのアクセス頻度に着目したアクセスバランスと容量バランスを両立させるデータ配置手法、バージョン管理を利用したアクセス・容量バランスの両立手法、バックアップへのアクセス移譲によりサービス品質の低下をできるだけ抑えるノード間データ転送手法、プライマリとバックアップで異なるキーによる暗号化を行うことで利用者メンバーの変更時にアクセス停止を回避するセキュリティ手法、ストレージに XML を格納する場合に XML の各ノードの効率的にラベル付けを行う手法、ストレージ中に格納された XML に対して複数のキーワードに合致する部分木を高速に取り出すための索引手法、さらに、ストレージ中の類似した XML 部分木を探し出し情報を統合する手法、といった先進ストレージを構築するための必要となる提案手法とその効果について述べる。

##### (a) 研究の準備

自律ディスクは、インテリジェント化したストレージ装置をネットワークで結合してストレージクラスタを構成する。ストレージ装置内で自律的に管理処理を行い、必要に応じてストレージ装置間で通信し合い、適宜データ移動を行うことで、システム内のアクセス負荷やデータ容量のバランスを取ってサービスの質を保証すると同時に、データの冗長配置による故障対策や効率の良い障害回復を行うことを特徴としている。

このため、どのストレージ装置にアクセスしても、集中管理を行わずに目的とするデータに高速に到達することが可能な分散ディレクトリを採用している。また、冗長データ配置としては、障害発生時や障害回復時にも安定したデータ提供を行うために、RAID のようなパリティ計算を行わず、プライマリとバックアップを論理的に隣接したストレージ装置でチェーンを構成するように配置するチェインド・デクラスタリングを採用している。さらに、柔軟な管理を行うために ECA ルールで動作管理を行っている。図 3 に構成例を示す。

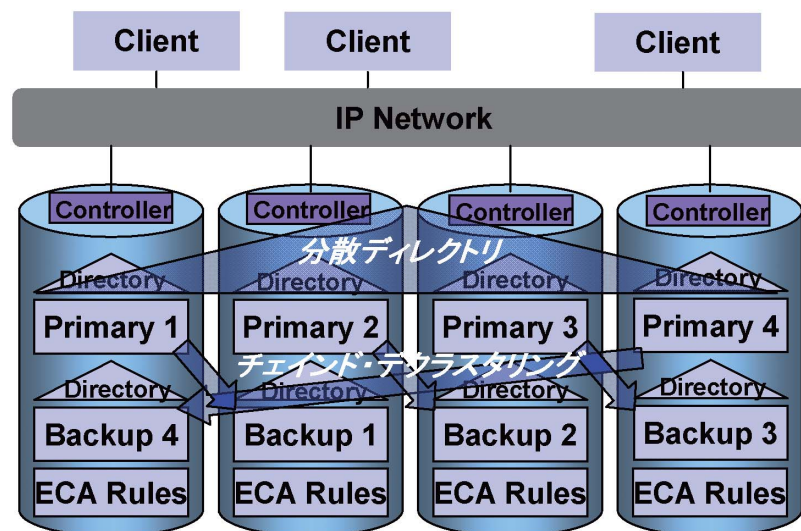


図 3 自律ディスクの構成例



自律ディスクでは、分散ディレクトリとして Fat-Btree 構造を採用していることを特徴としている。Fat-Btree は、従来の Btree 構造をベースに、更新の多い葉の部分はコピーを持たず、更新頻度の低い根に近い部分にのみコピーを持つことで、どのストレージ装置にアクセスしても Btree の根から葉までのアクセスパスが確保できると同時に、更新処理における同期コストを極力減らすことができる(図 4)。

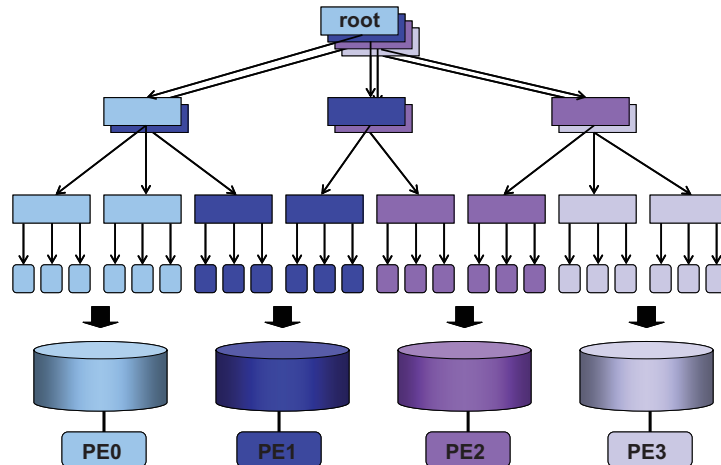


図 4 Fat-Btree による分散ディレクトリ

本研究では、この自律ディスクのアプローチを基に、ディペンダビリティや性能を向上させるために様々な手法の提案を行ってきた。それらの手法の性能改善の効果とその拡張性(スケーラビリティ)の評価を行うために、実験環境として、160 台のノードの接続形態が一様になるように構成した自律ディスクシミュレータを用意した。図 5 に自律ディスクシミュレータの外観を、表 1 にその諸元を示す。



図 5 自律ディスクシミュレータ

表 1 シミュレータの諸元

# of Nodes:	160
CPU:	Athlon XP-M 1800+
Memory:	DDR SDRAM 1GB
HD:	30GB, 5400rpm, 2.5in
Network:	Gbit Ether × 2
OS:	Linux 2.4.20
Java VM:	Sun J2SE 1.5.0_03

(b) 複数利用者ストレージ利用のための効率的な並列ディレクトリ同時実行制御

自律ディスクでは、分散ディレクトリとして Fat-Btree を採用しているが、Btree では複数の利用者から同時にアクセスを想定して同時実行制御が必須である。従来から、単一 Btree に対する効率的な同時実行制御手法として、B-OPT や ARIES/IM 等が提案されていたが、それらを分散 Btree に適用すると、木構造を変更する処理に対する同期コストが大きく効率が悪かった。このため、楽観的なラッチ処理をインクリメンタルに行う INC-OPT 方式を提案し、その効果を示したが、リスタート時のラッチのコストが大きいという問題が残っていた。

そこで、並列 Btree の構造変更可能箇所をマークすることでリスタートを軽減して同時実行制御の効率化を図る MARK-OPT 方式を提案し、自律ディスクシミュレータを用いてその効果を実証した。図 6 に MARK-OPT 方式のイメージを示す。本手法は特許申請を行い、論文は電子情報通信学会英文論文誌に採録されている。

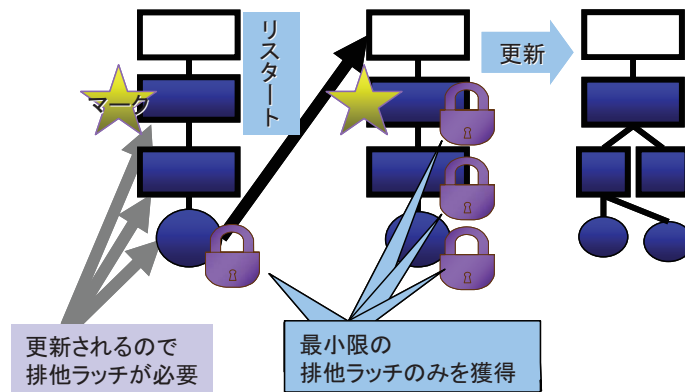


図 6 MARK-OPT 方式のイメージ

さらに、INC-OPT 方式や MARK-OPT 方式を含む従来の Btree 同時実行制御方式では、安定して Btree をトラバースするために、子ノードのラッチを確保してから親ノードのラッチを開放するラッチカップリング手法が用いられるが、並列 Btree におけるラッチカップリングは通信コストやラッチ保持時間の面から性能劣化の大きな要因であった。そこで、ラッチカップリングを行わないことによる問題点を解決する手法を取ることで、性能改善を行う LCFB (Latch Coupling Free parallel Btree concurrency control protocol) 手法、および LCFB 手法と Blink 手法を組み合わせ多 LCFB-Blink 手法を提案し、同様に自律ディスクシミュレータを用いてその効果を実証した。図 7 に従来手法と LCFB 手法の処理の流れの違いを示す。

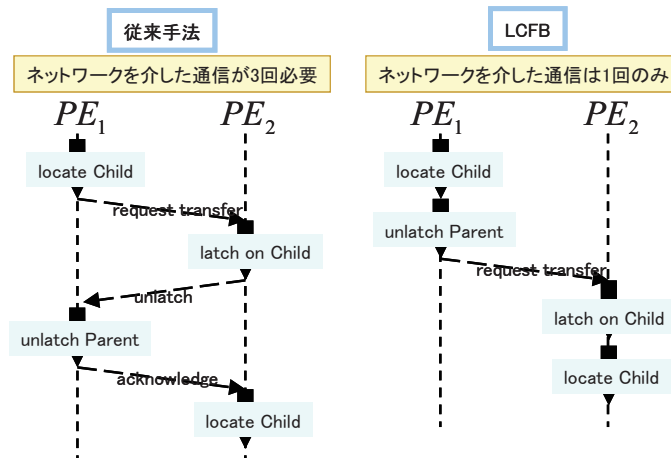


図 7 LCFB と従来手法との違い

自律ディスクシミュレータを用い、16 台のクライアントから 128 台の自律ディスクにアクセスして、従来および提案同時実行制御手法を適用した時の更新の比率を変化させた場合のスループットの推移を図 8 に示す。また、自律ディスクの台数を変化させたグラフを図 9 に示す。いずれの場合も ARIES/IM、INC-OPT、MARK-OPT、LCFB、LCFB-Blink の順に性能が良いことがわかる。このような分散ディレクトリ用の同時実行制御は、複数ユーザが利用する実用的な分散ストレージには必須である。

LCFB についても特許申請を行った。また論文は、当該分野で非常に高く評価されている国際会議である EDBT2008 に採択され(採択率 16.7%)、発表を行った。

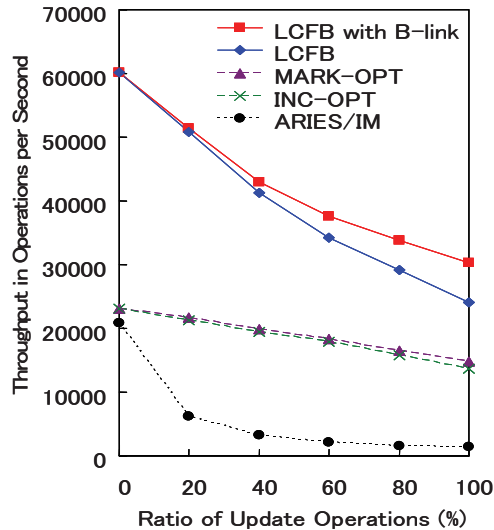


図 8 更新比率に対する各手法の違い

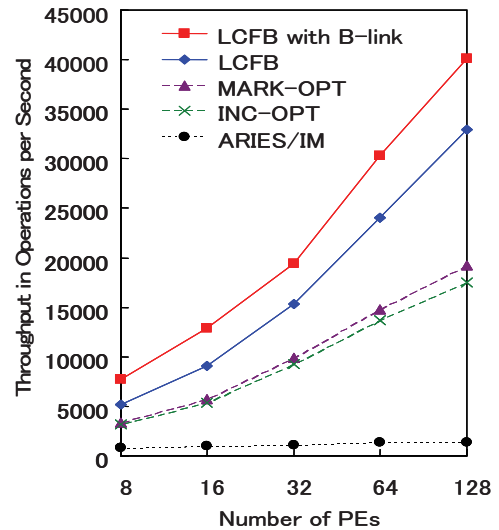


図 9 自律ディスク台数に対する違い

### (c) 矛盾なくストレージに格納するための効率的な分散コミットプロトコル

Fat-Btree のトラバースでは複数のストレージ装置が関与するため、一貫性を維持するために、それらの装置間でトランザクションとして分散コミットを行う必要がある。しかし、従来の分散コミットプロトコルである 2PC (2 Phase Commit) や EP (Early Prepare) は通信コストや同期コストの関係から、性能向上のボトルネックとなる。そこで、自律ディスクのチェーン・デクラスタリングによる複製配置を有効利用した BA-1.5PC (Backup Assist 1.5 Phase Commit) プロトコルを提案し、性能向上を実現した。

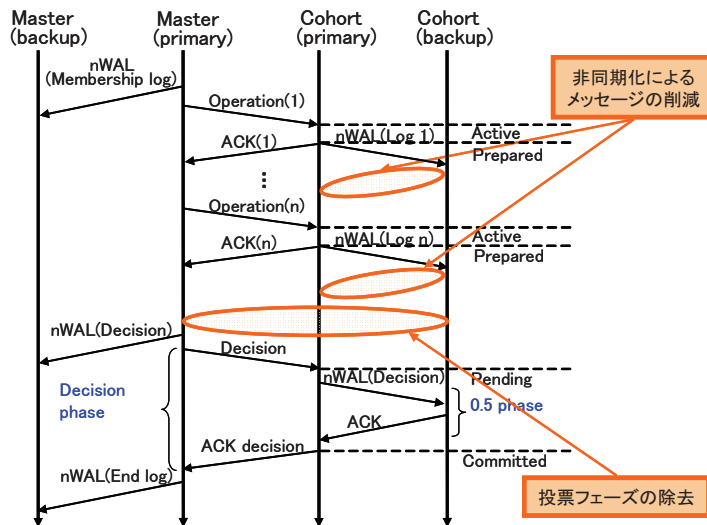


図 10 BA-1.5PC の流れ

BA-15.PC では、EP と同様に 2PC のプリペアコミットに対する応答を各操作の応答に暗黙的に内包すると同時に、EP のメッセージ転送前のログの強制書き込みをバックアップへの非同期 n-WAL によってコストを抑えている。非同期 n-WAL で除去した複製ストレージ間の同期は、決議フェーズに行うことで整合性を取っている。全体の処理の流れを図 10 に示す。

自律ディスクシミュレータの 64 ノードを用いた場合の実験の結果を図 11 に示す。実験の結果、提案手法の BA-1.5PC は、2PC および EP と比較して性能を大幅に改善し、オブジェクト数が増えるほどその差が拡大することを示している。このようなコミットプロトコルは複数ノードからなる実用的な分散ストレージには必須である。本件も特許申請を行うと同時に、当該分野でよく知られる国際会議である PRDC2006 に採択され、発表を行った。

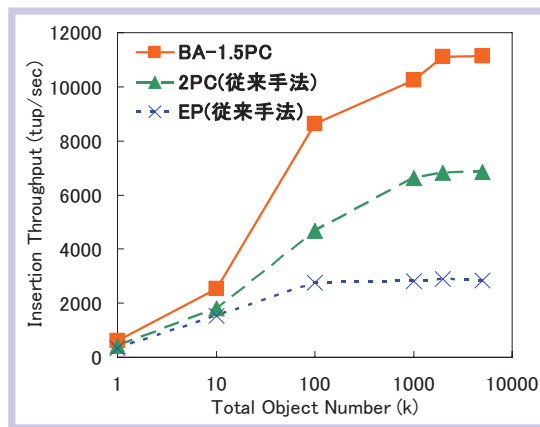


図 11 分散コミットプロトコルの性能比較

(d) バックアップデータを利用したアクセス・容量バランスの両立手法

アクセス負荷と容量バランスの両立に関しては、信頼性向上とサービス品質を保つために採用しているチェインド・デクラスティング方式を拡張して、プライマリデータでアクセス負荷バランスを取り、アクセス頻度の低いバックアップデータの配分で容量バランスを取る Adaptive Overlapped Declustering (AOD) 法を提案した。

AOD では、バックアップデータを左右の論理隣接ストレージ装置に分割し、分割比率を適合的に変更することで、容量バランスを取る(図 12)。つまり、アクセスの負荷分散はプライマリで行い、アクセス頻度が高いディスクのプライマリの容量は少なくなる。そのような場合に、分割されたバックアップの容量比率を多くすることで、ディスクスペースの利用効率を上げることができる。

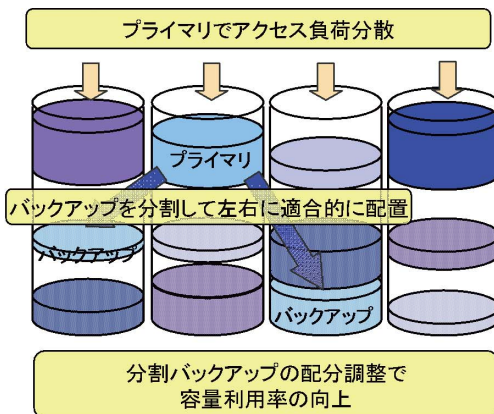


図 12 アクセスと容量バランスの両立

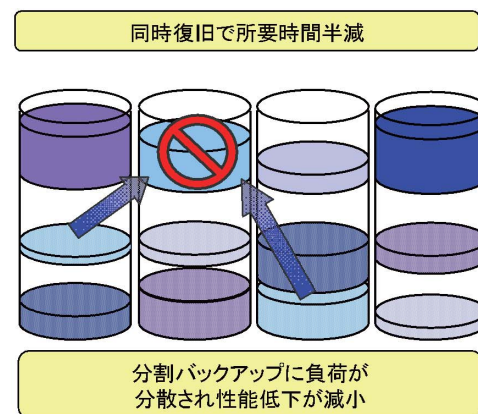


図 13 バックアップ再生時間の短縮



さらに、AOD は、両側のバックアップからプライマリを再生することができることから、容量利用率向上の効果だけでなく、バックアップ再生のときの時間短縮と負荷軽減効果もある(図 13)。

シミュレーションによって未使用領域率と復元時間における効果を確認した。シミュレーションの結果を図 14 と図 15 に示す。図 14 は、いずれかのディスクの容量が一杯になったときに試用されていないディスクスペースを、偏りの度合いを変化させて示したものである。従来手法と比較して、AOD の未使用部分の比率が少なく、効率よく使えていることを示している。また、図 15 は、故障後に新しいディスクを置き換えた場合の、経過時間とその間のサービスのレスポンス時間を示している。従来手法と比較して、短い時間でバックアップが再生され、再生の間の性能低下が低くなっていることが分かる。

このようなデータ配置手法は、実用化においてコストを抑えてシステムを構築するには、必要な技術である。本件も特許申請を行うと同時に、当該分野でトップクラスの国際会議の一つである ICDE2005 に採択され(採択率 13%)、発表を行った。

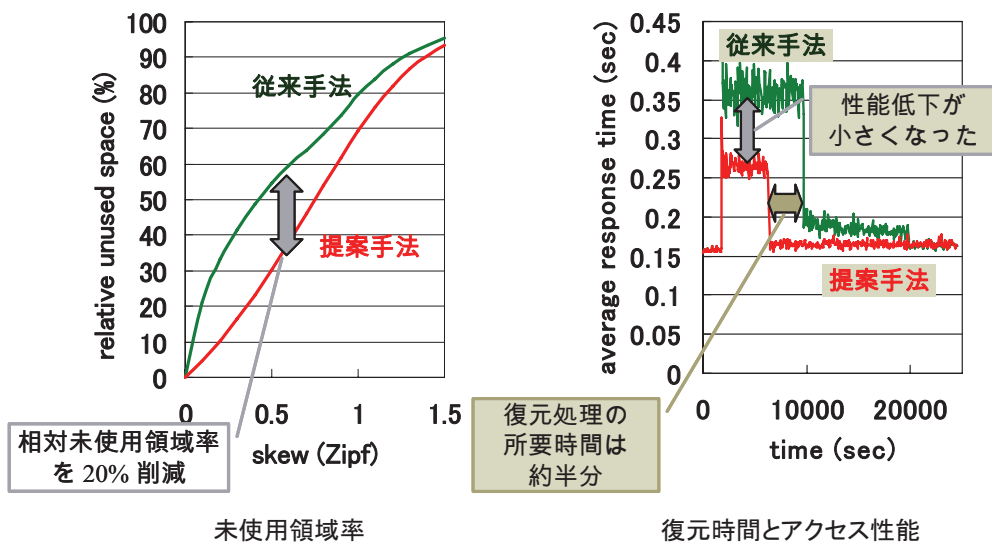


図 14 容量バランスの効果

図 15 バックアップ復元における効果

(e) バージョン管理を利用したアクセス・容量バランスの両立手法

ストレージ中に同じファイルの複数のバージョンを格納しておくことも多い。そこで、差分バージョンを有効利用したアクセスバランスと容量バランスの両立方法である COBALT (Combination Of Btree And Linked-list Transfer)の提案を行った。

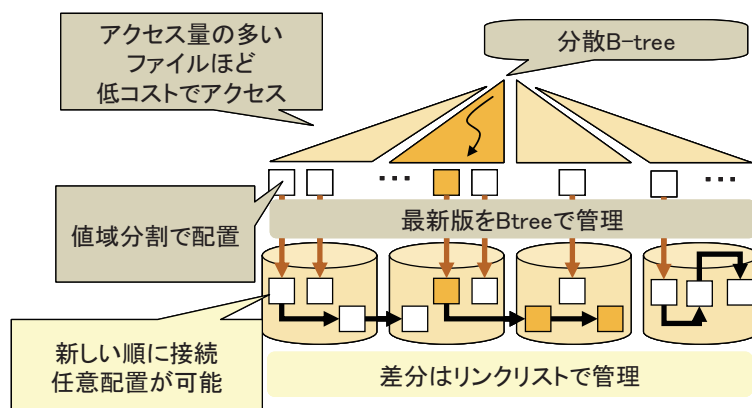


図 16 COBALT のアクセス構造

COBALT ではストレージに格納されるデータのバージョン間のアクセス頻度に着目して、アクセス負荷と容量バランスを両立させる。アクセス頻度の高い最新バージョンはこれまで同様分散 Btree で管理しながらアクセス負荷分散に用い、比較的アクセス頻度の低い古いバージョンデータを容量均衡化に利用する。このために、分散 Btree と連結リストの組み合わせで効率よく管理する(図 16)。

シミュレーションによってアクセス負荷のバランスを保ちながら容量のバランスが確保される効果を確認した。図 17と図 18に、従来手法と比較した実験開始後の経過時間とアクセス負荷および容量の偏りの推移を示す。Btree のみと比較して同様のアクセス負荷の均衡を保ちながら優れた容量均衡化の効果を実証している。

スナップショット管理においてバージョンングを行うようなストレージシステムにおいて、バランス確保のためには有効な手法である。本件も特許申請を行い、論文は電子情報通信学会の和文論文誌に採録された。

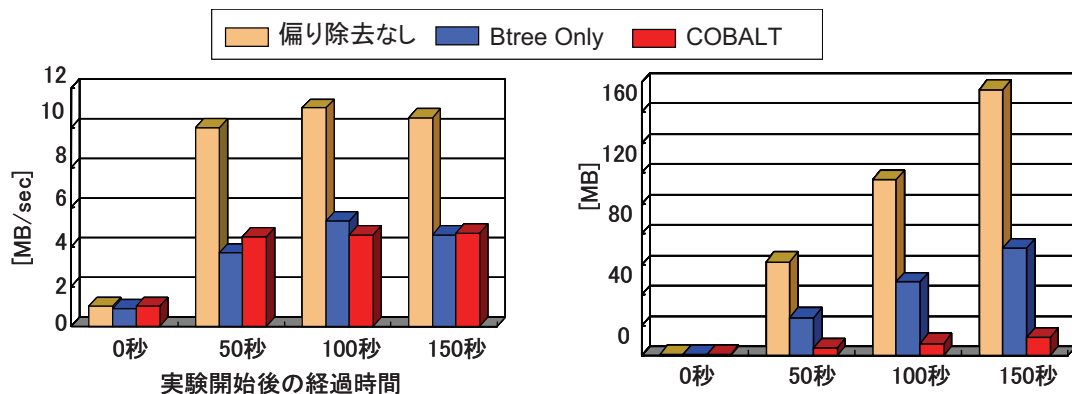


図 17 アクセス負荷の偏り

図 18 容量の偏り

#### (f) サービス品質を考慮したノード間データ移動手法

ストレージに対するアクセスのサービスの質を保証するためには、通常サービス処理とデータ管理処理の優先度管理が重要となる。通常サービス処理の優先度を下げるとレスポンスが悪化するが、通常サービス処理の優先度を上げすぎるとデータ管理処理が働かなくなってしまう。特に、自律ディスクではデータ管理のためにストレージ装置間でデータ移動が行われることから、データ移動処理と通常サービスを両立させる必要がある。

一般に分散ストレージシステムにおけるストレージ装置間でのアクセス負荷均衡化のためのデータ移動は、一時的に負荷の高いストレージ装置の負荷がさらに高くなり、クライアントへのサービスとしてのデータ提供の性能が保障できないという問題が生じる。この問題を解決するために、クライアントへのサービスの性能保証のため、Replica-assisted Migration (RM)と呼ぶ手法を提案した。

RM 手法では、信頼性確保のための複製データを利用して、クライアントのアクセス要求に対して複製データへのアクセス回送を行うと同時に、データ移動の対象をその時の負荷の状況からオリジナルデータか複製データかを動的に切り替えるデータ移動経路選択を行う。このような2種の複製データ利用法をを組み合わせることによって、データ移動時にも定められた性能要件を満たしてクライアントへの十分なサービスの提供を行うことが可能となる(図 19)。

これに加えて、RM 手法において、各ノード最大許容負荷を目標とした複製利用制御アルゴリズム、ストレージノードキャッシュを考慮した回送制御、現実的なアクセス負荷を対象とした複製利用制御手法を提案し、その効果を評価した。

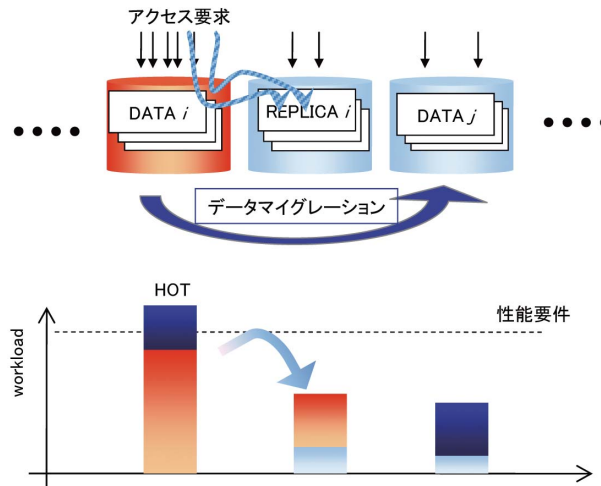


図 19 RA の処理

読み出しのみ、読み書きを行う人工的アクセスパターンと、大規模 WEB サーバおよびファイルサーバの 4 種のワークロードを用いた実験を行い、提案手法が性能保証に有効であることを実証した。図 20 は読み出しのみ、図 21 は読み書きを行う人工的ワークロードに対する実験結果である。

本件に関しても特許を出願し、論文は情報処理学会のジャーナルに採録された。

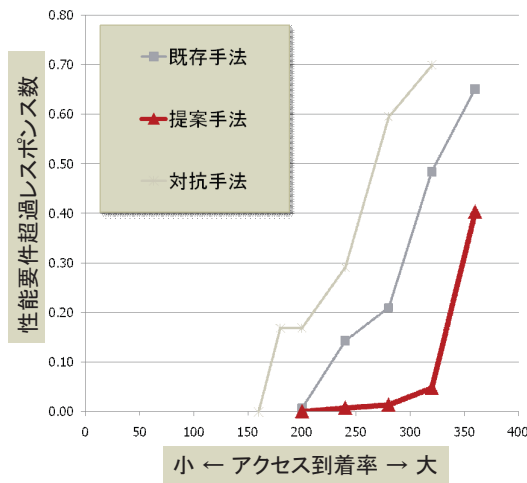


図 20 読み出しのみを行うワークロード

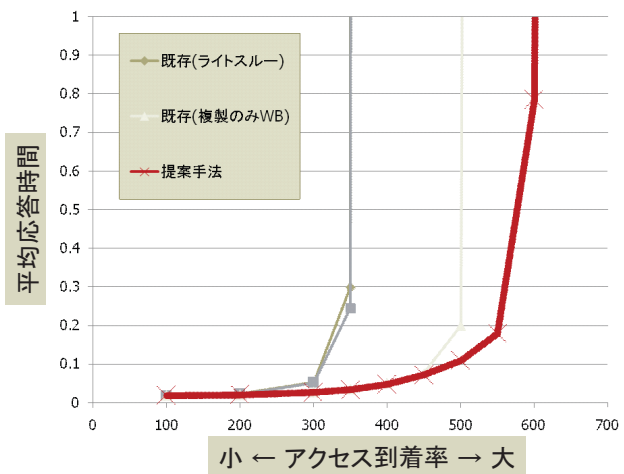


図 21 読み書きを行うワークロード

### (g) 高信頼分散共有ストレージのためのセキュリティ手法

情報セキュリティの重要性が増大する中で、ネットワークストレージでは悪意あるユーザによる盗聴から伝送路上のデータを保護することが必須となる。データの保護には主に暗号が用いられる。伝送時のみ暗号を用いる encrypt-on-wire 方式に比べ、データを予め暗号化して格納する encrypt-on-disk 方式は伝送時の効率がよい。しかし、encrypt-on-disk 方式ではユーザのアクセス権失効に伴いデータを再暗号化する必要が生じる。この再暗号化の方法は、直ちに再暗号化を行う active revocation とデータ更新時まで再暗号化を遅延する lazy revocation があるが、active revocation は性能面でのコストが高く、lazy revocation は古い鍵で暗号化された脆弱な状態が残るといった問題点がある。

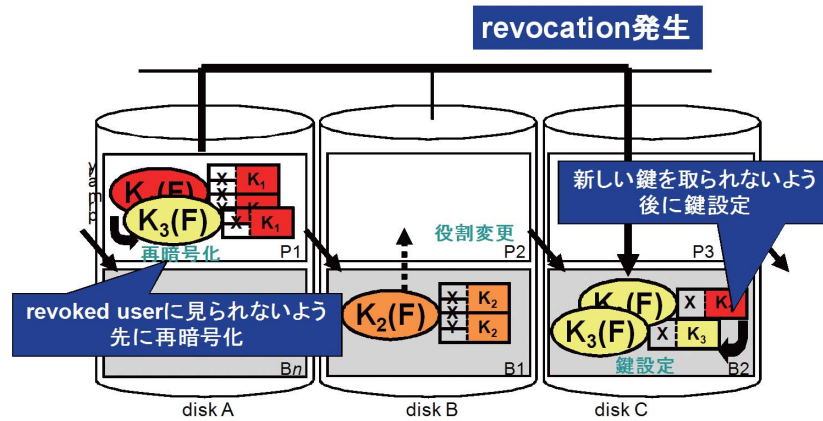


図 22 RBA-Rev の処理の流れ

自律ディスクのような、データの複製を他のディスク装置に置く分散ストレージにおいてプライマリ・バックアップデータ配置を前提として、安全性の高い active revocation の性能面での改善を行った RBA-Rev と BA-Rev の提案を行い、その性能への影響を評価した。RBA-Rev と BA-Rev では、アクセスされない複製データを予め異なる暗号鍵で暗号化して準備しておくことによって、revocation 象ファイルへアクセスできない状態の継続を抑え、同時に再暗号化処理を行うディスク数が減少して、数のディスクで大きく性能が低下することを抑えることができる(図 22)。その効果は配置に制約がない場合において特に大きいことを確認した(図 23)。

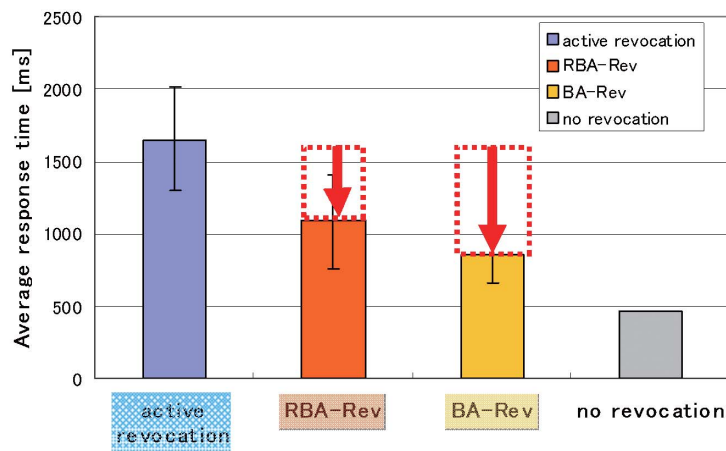


図 23 revocation 時の性能への影響

また、通常の encrypt-on-disk 方式では一時的に平文が生成される為、情報漏洩の可能性がある。分散ストレージの構造と特定の暗号化方式の性質を利用した、平文を生成しない再暗号化手法を提案し、性能面とセキュリティ面で評価し、有用性を示した。これらのセキュリティ技術は、今後のストレージ実用化において欠くことのできない技術である。

#### (h) コンテンツ単位の情報ライフサイクル管理

コンテンツの特徴を活かしたストレージ管理として、ECA ルールを用いた柔軟な情報ライフサイクル管理を行う RB-ILM (Rule Based Information Lifecycle Management) を提案した。



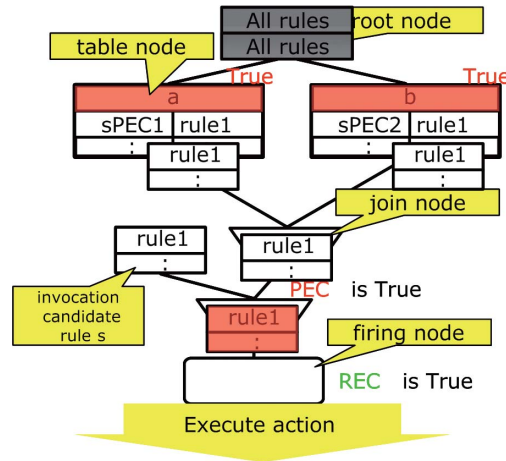


図 24 弁別ネットワークの例

コンテンツ単位での ECA ルールを用いることで柔軟性を実現するとともに、コンテンツ数増加によりルール数が増えた場合の対応として、ルールを分類し弁別ネットワークを用いたルール処理の高速化を図っている。提案する弁別ネットワークの例を図 24 に示す。従来のルールベースシステムの弁別ネットワークでは、弁別ネットワーク上をデータが移動するが、本提案手法では大量のルールに対応するために、ルールの条件を予め評価できる PEC と、実行時でなければ評価できない REC に分類し、PEC の部分を弁別ネットワーク上に流すことで効率化を図っている。図 25 に示すシミュレーションの結果は、提案手法がルールの増加に対して効果的であることを示している。

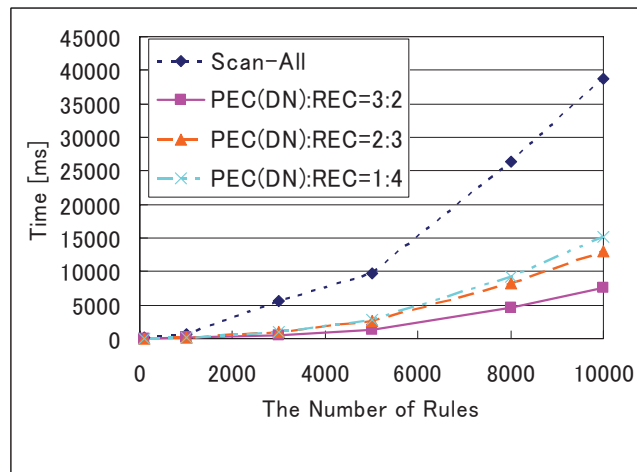


図 25 シミュレーション結果

(i) 格納されたデータを扱うためのディペンダブルな Web サービス連携

分散して格納されたデータを扱うための枠組みとして Web サービス連携を利用することを想定し、その場合に処理全体をディペンダブルにするための手法に関しても検討を行った。各データ格納ノードが Web サービスの枠組みで情報を提供するとした場合に、利用できる Web サービス提供ノードを自律的に選択し、障害が発生した場合には同等のサービスを提供するノードにリトライするためのトランザクションモデルとアーキテクチャ、および信頼性と性能を解析する QoS モデルからなる FENECIA を提案した。本件に関しては、論文が RIDE2004、WIRI2005、PRDC2005 に採択されると共に、国際的に非常に評価の高い論文誌である VLDB Journal に採録された。

#### (j) ストレージにおけるXML 部分木検索高速化手法

ストレージに格納された XML に対して、その部分木を高速に検索することが、XML を利用する上では必要となる。XML の部分木を扱うため、まず XML の各ノードにラベリングをする手法をとることが多い。ストレージにおいては、部分木の検索性能と同時にそのラベルの容量利用率も問題となる。このため、検索性能と容量利用率のよいラベリング手法である DO-VLEI を提案した。本件に関しては、ICDE2005 のポスターに採択された。

また、XML 部分木に対して高速で検索を行うためには、キーワードに対する索引が必要となる。従来の索引手法では、キーワード数が増えると性能が非常に悪くなる傾向にあった。このため、キーワード数が増えても検索性能のよい索引手法として、Dewey Order に基づく XML ノードラベリングを使った B+tree による索引手法である KBDLM と、それにフォールスドロップ対策を行ったスーパーインポーズドコードを組み合わせた新たな索引手法である KBSI を提案し、その検索速度に対する効果を従来手法である SA (Stack Algorithm)と比較して確認した(図 26)。本件に関しては、DEXA2008(採択率 19%)に採択された。

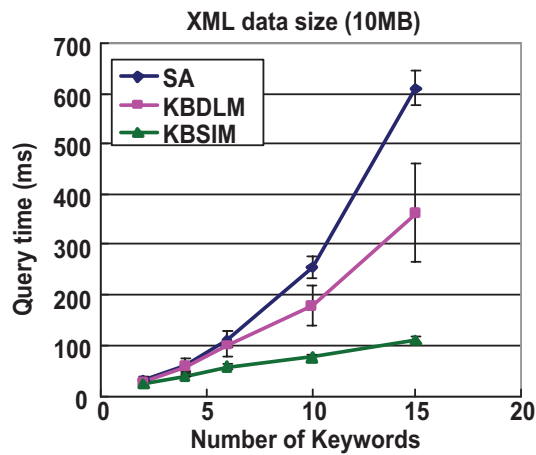


図 26 XML 部分木検索手法の実行時間の比較

さらに、類似した XML 部分木の同一性を高速に判定し、情報として統合することも XML ストレージ管理を行う上では必要となる。従来技術としては、木の編集距離を用いた類似部分木検出手法があるが、処理時間がかかり精度が悪いという問題があった。そこで、葉ノードの内容を重視した新たな類似検出手法である LAX を提案し、その性能と精度の両面に対する効果を示した(図 27)。本件に関しては、BNCOD2005(採択率 15%)、WAIM2008(採択率 25%)に採択された。

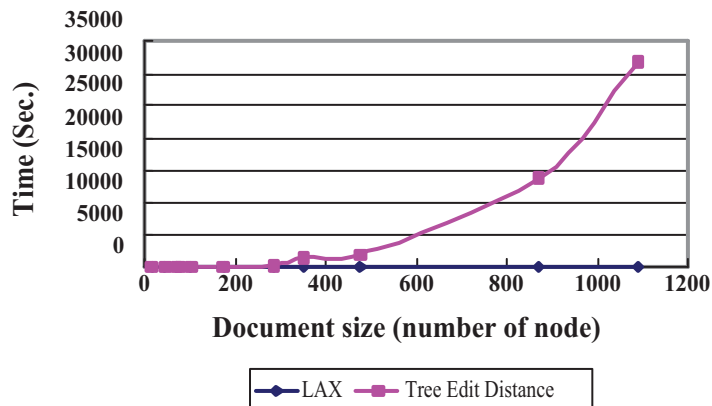


図 27 XML 部分木類似性判定手法の実行時間の比較

### 3.2 先進ストレージシステムにおける高度メディア蓄積・管理手法に関する研究(奈良先端科学技術大学院大学 高度メディア蓄積・管理手法研究グループ)

#### (1)研究実施内容及び成果

W3C で標準化された XML は、組織間のデータ交換形式としてデファクトスタンダードとなり、計算機上でのデータ永続化形式としても利用が広がっている。例えば、DNA やたんぱく質のバイオ情報等の大規模な科学データベースは、各研究機関独自の形式からデータ交換が容易な XML に移行しつつある。また、オフィス文書や、ビデオデータの索引等も XML が記述形式として利用されており、今後ますます多くデータが XML 形式としてストレージに格納されることは間違いない。

しかしながら、大規模 XML データを高速に検索できるだけでなく、効率よく維持管理も可能にすることは容易ではない。米国 ZapThink 社の調査では、「コンテンツの共有・再配信・再統合の実現のために開発者は XML の柔軟性と拡張性を利用しているが、この XML の性質は現在のストレージにとって困難な存在」と位置付けており、XML データのストレージへの格納や格納された XML データの問合せ処理技術の進展が強く望まれている。

現在は、ホスト計算機で XML データが処理されるが、XML データ処理の煩雑さが、ホスト計算機の処理負荷を押し上げる原因となっている。もし、これらの XML データの効率の良い格納、ならびに処理の一部もしくは全体が、自律ディスクを利用してストレージレベルで実現できるならば、より効率的な XML データ処理が可能となると考えられる。そのため本研究では、XML のデータ格納方式として、A) XML データをそのままストレージに格納するネイティブ XML データストレージ、B) XML データを関係データモデルに写像してストレージに格納する方式、の全く異なる二つのアプローチに基づいて、それぞれ XML データの格納方式と XML データ問合せ処理について研究を行った。

まず、ネイティブ XML データストレージに関して、先行研究として Natix、OrientStore 等が挙げられるが、我々は XML 問合せの典型例であるユースケースから、最適な格納方法を研究した点で実用的かつ高速なデータアクセスが可能であり、問合せ処理エンジンのメモリ消費量も非常に少ない。この点において世界で随一の高性能な XML ストレージである。また、これを分散処理可能なよう拡張した問合せエンジンは、世界で初めて参照渡しを利用したものであり、XML ストレージから結果出力まで一貫してパイプライン処理可能である。これにより、これまで世界で最高速とされていた値渡しを利用した MonetDB/XRPC よりも 20 倍の高速な XML データ処理が可能となった。

一方、関係データモデルに基づく XML データストレージに関して、これまで XML データの分割、分散配置に関しては、Bremerらの方式があるが、問合せ処理に関しては議論されていなかった。我々は、XML データの分割および最適な分散配置ならびにデータ移動方法だけでなく、分割された XML データの最適な問合せ処理に至るまで、一貫した枠組みを与えた点で他の類似研究に追従を許していない。

上記の二つのアプローチにより、多くのデジタルメディアで利用される XML データの自律ディスクへの格納、維持管理、さらに自律ディスクと連携した XML データの効率の良い処理方法について明らかにした。ネイティブ XML データストレージとその問合せ処理については、(a) Xbird: XML データのストレージへの格納方式と、それに連携する高性能・省メモリ XQuery 問合せ処理系を開発した。また、(b) Xbird/D: remote proxy と参照渡しを利用した、高性能分散 XML 問合せ処理系を開発した。一方、関係データモデルに基づく XML データストレージとその問合せ処理では、(c) PMX: 関係ストレージを搭載した自律ディスククラスタ上で XML データの問合せ処理を行うことを目的として、関係データベースに写像した XML フラグメントを自律ディスクの各ノードに配置するためのコストベースの最適配置手法と問合せ処理手法を開発した。また、(d) GMX: 自律ディスククラスタ上における XML データの問合せ処理の高速化を目的として、Holistic Twig Join を分散実行するためのデータ分割戦略を開発するとともに、並列処理アルゴリズムを開発した。以下に、それぞれの項目についてその概要と成果を述べる。

### (a) Xbird: XML データのストレージへの格納方式とその XQuery 問合せ処理

XML データの二次記憶への格納方式として、XML データの内部表現である Document Table Model (DTM)の一形式を利用する方式、ならびにそのデータ格納方式と連携して動作する XQuery 問合せ処理系 Xbird を提案した。DTM は、実行性能の効率化と記憶領域の最小化を目的とした主記憶上での XML データの内部表現形式であり、主要な XQuery/XSLT 処理系で既に採用されている。DTM は基本的に表形式で表現され、XML 木の各ノード固有の整数値のラベル(表の索引)を用いて、XML データ維持のための情報を配列要素内に格納する。

DTM はプリミティブデータ型で XML の木構造を表現できるという特徴があり、従来の DOM モデルのようなオブジェクト生成に起因する実行効率の悪化、ならびにオブジェクトが消費するメモリ領域を抑制することができ、研究目標の一つである問合せ処理エンジンの省メモリ化が可能となるばかりでなく、高速化にも結びつく。DTM に基づく我々のアプローチは、これら実装手法における一次記憶上のデータ表現に対して二次記憶への拡張を行う場合に、透過性の観点から優れている。

図 29 は、図 28 の XML データを文書順に深さ優先順でラベル付けした XML 木を DTM として表したものである。図 29 の DTM の 1 列目はノード種別とノードの各種属性、2 列目、3 列目には親ノード、弟(next-sibling)ノードをそれぞれ保持する。4 列目には DTM 表とは別に管理する文字列の ID、あるいはローカル要素名と名前空間との組み合わせ(QName)を保持する。なお、両図中の E は要素ノード、T はテキストノードを指す。

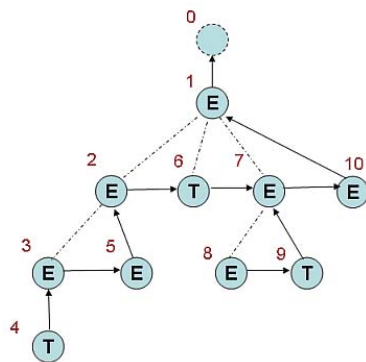


図 28 深さ優先文書順ラベル付 XML 木

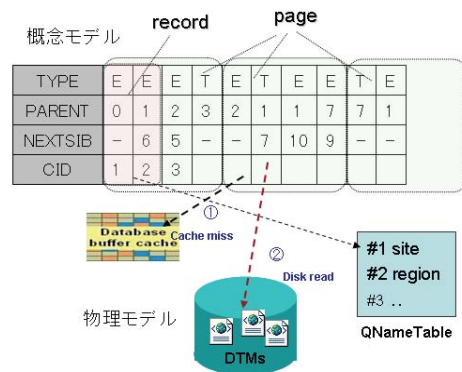


図 29 Document Table Model

DTM 表に対して効率のよいアクセスを行うために、DTM 表の二次記憶への格納は、DTM 表をブロック化し、二次記憶のアクセス単位であるページに各ブロックを入れることで行う。この二次記憶に格納された DTM 表を DTMs ファイルと呼ぶ。問合せ処理中に要求されるページは、DTMs ファイルに対してページングを行うことにより実現する(図 29)。

ページングは、基本的にデータベースシステムのページバッファ管理と同等であるが、一度に読み込むデータサイズについてはデータベースのページサイズに基づくよりも、実際の XML データアクセスにおけるページ要求パターンに即したものである方がより効果的である。実際の問合せの典型的なページ要求パターンは、概ね文書順となっていることが予備実験により判明した。すなわち、ページの読み込みの際はエクステントと呼ばれる複数の連続ページを単位として、一括して連続ページを読出せば良い。一方、データの更新を考慮すると、データの更新は通常非常に小さな領域に対して行われるので、データの書き込み時はページ単位で行うほうが良い。そのため、提案方式では、読出しはエクステント単位、データの書き込みはページ単位とした。

以上の XML データの二次記憶方式およびそのアクセス方法を実現する省メモリ XQuery 問合せ処理エンジン Xbird を Java で実装し、XML データベースの標準ベンチマークツールである XMark を用いて評価を行った。比較対象には、同じく Java で実装された商用 XQuery プロセッサである SAXON-SA を利用した。評価を行った計算機は、CPU PentiumD



2.8GHz、メモリ 2GB、および 7200 回転の SATA ディスクから構成され、OS として Windows XP が稼動している。JVM のヒープのオプションとして、-Xms512m -Xmx1024m を与えた。なお、ページサイズを 2KB、エクステントサイズを 64KB とした。

実験結果を表にしたものを図 30 に示す。表中の DNF は計算が一定時間以内に終了しなかった場合である。提案方式は「固定長 DTMs」であるが、一次記憶のみで二次記憶を用いない「DTM」方式、DTMs の拡張として DTM 表をページ内に格納する際に LZ 圧縮して格納し、I/O の際には必要サイズのデータのみを二次記憶をアクセスする「可変長 DTMs」方式も参考のため図 30 に示した。

図 30 に示すとおり、提案方式は問合せ処理において、高性能であることが明らかである。さらに注目すべき点は、SAXON-SA が XMark のスケールファクタが 4 以上の XML データに対してメモリ不足のために動作不能となるにも関わらず、提案方式はスケールファクタ 10(約 1GB)でも効率よく動作することである。つまり、提案方式は XML データを効率よく二次記憶に格納し、その XML データに対して高性能かつ省メモリの XML 問合せ処理を実現している。

	XMark113MB (ScaleFactor: 1)				XMark340MB (ScaleFactor:3)			
	DTM	可変長DTMs	固定長DTMs	Saxon-SA	DTM	可変長DTMs	固定長DTMs	Saxon-SA
Q1	9.938	3.531	3.609	9.875	28.64	10.704	8.93	26.875
Q2	9.922	4.296	3.891	10.203	28.593	11.75	9.265	27.968
Q3	9.89	5.219	4.687	10.328	30.281	14.359	12.47	27.641
Q4	9.812	5.453	5.266	10.375	29.641	15.282	12.797	27.219
Q5	9.203	3.719	3.953	10.25	29.172	10.14	8.953	25
Q6	10.968	10.156	8.578	9.906	32.825	29.375	22.703	27.328
Q7	12.969	15.562	14.62	11.047	39.203	61.906	45.453	27.734
Q8	10.672	5.657	5.515	11.344	30.5	14.875	13.203	27.875
Q9	10.953	8.937	9.25	914.672	31.31	25	22.422	DNF
Q10	16.813	23.157	27.422	13.422	49.704	78.453	73.796	36.828
Q11	19.328	16.907	16.938	766.344	89.687	87.297	84.172	DNF
Q12	15.703	12.984	12.656	351.812	67.563	60.156	57.656	DNF
Q13	9.781	3.437	3.563	10.719	28.781	9.64	8.391	30.484
Q14	11.719	12.563	11.484	10.781	34.47	37.922	30.218	32.234
Q15	9.765	3.25	3.156	9.828	28.969	8.578	7.719	26.719
Q16	9.812	3.39	3.391	10.5	29.15	9.406	8.11	25.797
Q17	10.109	4.156	3.984	10.469	29.891	11.63	9.578	27.016
Q18	10.25	4.969	4.75	10.672	28.922	13.422	11.93	27.36
Q19	11.532	9.594	9.797	11.5	33.657	26.922	24.844	34.625
Q20	10.312	5.344	5.156	10.313	30.219	14.765	13.156	27.938

図 30 Xbird(固定長 DTMs)と SAXON-SA との問合せ処理時間の比較(単位:秒)

#### (b) Xbird/D: remote proxy と参照渡しを利用した高性能分散 XQuery 問合せ処理

一般的なストレージシステムの問題の一つとして、計算はホストで行うため、計算に必要なデータ全てをストレージとホストの間で転送する必要があることが挙げられる。この問題を解決するために、Active Disk や Intelligent Disk 等のストレージ側で簡単な計算を行わせ、ホストへの転送データ量を削減する手法が提案された。データベース処理はある条件でデータを絞り込む操作が多く、その計算は単純であるため、ストレージ側でのデータの絞り込む手法と親和性が良い。また、自律ディスククラスタの問題として、データの位置透過性を実現する引き換えとして、内部ネットワークのトラフィックの増大が挙げられる。これは、データアクセスごとに内部ネットワークを利用してデータの物理位置を検索するためである。これを解決するためには、ストレージからのデータ転送をアクティブ転送にすれば良い。これにより、要求されたデータが格納されているストレージデバイスへ一回だけルーティングし、そのストレージデバイスからホストへコネクションを張れば、それ以降のデータ転送は内部ネットワークを介さずに直接ホストとデータ転送が可能となる。

上記二点を同時に解決するための XQuery 処理系 Xbird/D の提案を行った。Xbird/D は Xbird をリモート問合せ可能なよう拡張したものである。この拡張は、分散 XQuery 問合せ中の各オペレータを遅延評価でストリーム処理を行うことができ、Xbird の DTM に基づく

XML データ格納方式と親和性が高い。また、各オペレータは任意の計算ノードに配置することができ、計算能力を持つストレージデバイスにデータを絞り込むオペレータを配置し、アクティブ転送を行うことにより、上記の二つの問題を解決できる。これにより、Xbird/D は自律ディスククラスと連携して、高性能な XQuery 問合せ処理を実現できる。

Xbird/D では、問合せのオペレータ間並列を実現するために remote proxy を利用した。ベースとなる Xbird プロセッサは、Volcano モデルに基づき、イテレータ木による反復問合せ処理を行う。パイプライン化された問合せ処理により、オペレータはループ処理や軸を辿る処理を実行する。このイテレータに基づく実行モデルは、問合せ式の遅延評価を可能にし、Xbird/D の分散処理においても重要な役割を担う。ここで、remote proxy を利用した pass-by-reference (参照渡し) 実行戦略がどのように実現されるかを述べる。Remote proxy は、分散オブジェクト通信のための代理(proxy) デザインパターンの拡張であり、Xbird/D 特有の機能ではないが、これまでに分散 XML 問合せ処理に用いられた例はなく、また remote proxy を利用することだけでオペレータ間並列を実現することできない。そこで、remote proxy とリモートブロッキングキューを利用した非同期の中間結果の実体生産方式を組み合わせた。これにより、データ実体の供給過多・不足を回避し、システムリソースを効率よく活用することができる。

ここで、Xbird/D における remote proxy の効果を測るために、値渡しの RPC に基づく最も高速な分散 XQuery プロセッサの一つである MonetDB/XRPC バージョン 4.18.1 との性能比較を行った。実験環境には 4 台の計算機を利用した。ここでは、それぞれの計算機で実行される XQuery プロセッサを PE1...PE4 として表記する。それぞれの計算機は、PE2 が Athlon 64 X2 2.4GHz の CPU を装備しているのを例外として、1Gbps の Ethernet で繋がれ、CPU として Pentium D 2.8GHz、2GB のメモリ、SuSE Linux 10.2 を OS、JDK 1.6 を実行環境として装備する。Xbird/D の評価には、図 31 に示す問合せを用い、同等の問合せを MonetDB/XRPC の評価に用いた。データセットには、XMark のデータ生成器にスケールファクタ 10 を設定して生成した 1.1GB の XML 文書を利用し、生成した文書は図 31 の \$col 変数に束縛した。生成した XML 文書は、PE3、PE4 で実行中の MonetDB と Xbird のデータベースインスタンスに予め取り込んだ。

```

declare function bdq:select1() {
  execute at $PE3 {
    fn:collection($col)/site/closed_auctions/closed_auction
  }
};
declare function bdq:select2() {
  execute at $PE4 {
    fn:collection($col)/site/open_auctions/open_auction
  }
};
declare function bdq:reduce() {
  execute at $PE2 {
    ( fn:subsequence(bdq:select1(), 1, 1000)
      | fn:subsequence(bdq:select2(), 1, 1000) )
  }
};
declare function local:filter() {
  for $a in bdq:reduce()
  where $a/seller/@person >= "person10000"
  or $a/buyer/@person >= "person10000"
  return $a
};
local:filter() (: execute at PE1 :)

```

図 31 分散 XQuery 問合せ

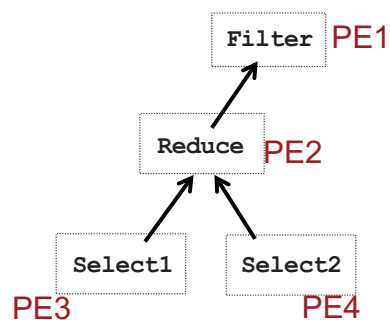


図 32 問合せオペレータの配置

実験結果の概要が図 33 である。実験では、remote proxy(Proxy)、演算結果を直接親ノードに返す remote proxy(Proxy+Forward)、Xbird/D の値渡し版(Value)、ならびに MonetDB/XRPC (XRPC) の四つの戦略の比較を行った。図 33 より、我々の remote proxy を利用したリモート参照渡しの実装は実行時間に大きな改善を示した。これは、PE3 と PE4 における不必要な計算を除去した結果であり、分散 XQuery 処理に遅延評価を適用した結果である。値渡しを利用したリモート問合せの評価は、完全な結果を一度に計算・生産す

るが、それらの全てが後の計算で利用されていない。この理由により、我々の remote proxy 戦略(Proxy) が値渡しに比べて9倍高速となった。さらに、直接親ノードに計算結果を返す方式(Proxy+Forward)は PE2 における冗長な符号化/復号化を除去し、通信オーバーヘッドを削減し、我々の Java で実装したプロトタイプシステムは、競合実装である C++で実装された XRPC に比べて22倍の性能向上を示すに至った。また、remote proxy を利用したシステムのみが、マルチユーザ環境をシミュレートした 30 スレッドから総計 160 の並行した問合せを 160 秒で処理することができた。この時の最大問合せ処理時間は 53.75 秒であり、平均処理時間は 36.75 秒であった。値渡しを利用した実装(Value と XRPC) では、頻繁なスワップイン/スワップアウトが発生し、そのリソース活用の貧弱さを露呈した。

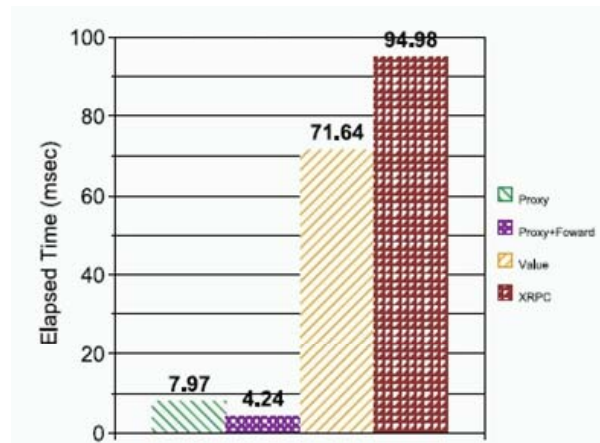


図 33 分散 XQuery 問合せの実験結果

### (c) PMX: 関係データモデルに基づく大規模 XML データ分散処理方式

XML データの格納・検索手法には、さまざまな手法が存在し、それぞれ長所短所が知られているが、本研究では関係データベースを利用する手法を用いる。すなわち、自律ディスクの各ノード上において、関係データベース、あるいはそれに準じた機能を有するシステムが動作していることを仮定する。これには、関係データベースが確立された技術であり、さまざまな技術的蓄積があること、利用可能なソフトウェアプロダクトが、商用、オープンソースを含めて多数存在することなどが利点として挙げられる。

関係データベースは本質的にフラットな関係表しか扱うことができないのに対し、XML は木構造を有するため、そのままでは XML データを格納することができない。そのため、木構造を何らかの方法で関係表に写像する必要がある。本研究では、われわれの研究グループで開発した、経路アプローチの一種である XRel を利用する。このようにして格納された XML データを、自律ディスククラスタ上で分散処理するためには、いくつかの技術的課題が存在する：1) XML の木構造を考慮した XML データの分割法、2) 分割した XML データ(XML フラグメント)を、自律ディスククラスタの各ノードに配置する方法、3) 問合せの分散処理手法。

図 34 に提案手法の概要を示す。上で述べたように、XML データは経路アプローチによって関係表へ変換する。次に、分散配置のために得られた関係表のフラグメント化を行うが、XML は木構造を有するため、フラグメント化は単純には行えない。そこで本研究では、XML データ集合から導出される構造概要(DataGuide)上で XML データに対するフラグメントを形式的に定義する方法を開発した。また、システムに対する問合せワークロードを元に、そのワークロードの元で最適なフラグメント化と自律ディスクへの配置を行うために、提案システムにおける問合せ処理のコストモデルを開発した。このコスト関数を元に、複数のフラグメントをコスト最適になるように複数のノードに配置する。これはある種の組合せ最適化問題ととらえることができる。この問題に対しては、遺伝的アルゴリズムを用いて、準最適な自律ディスクへのフラグメントの配置を求めることとした。

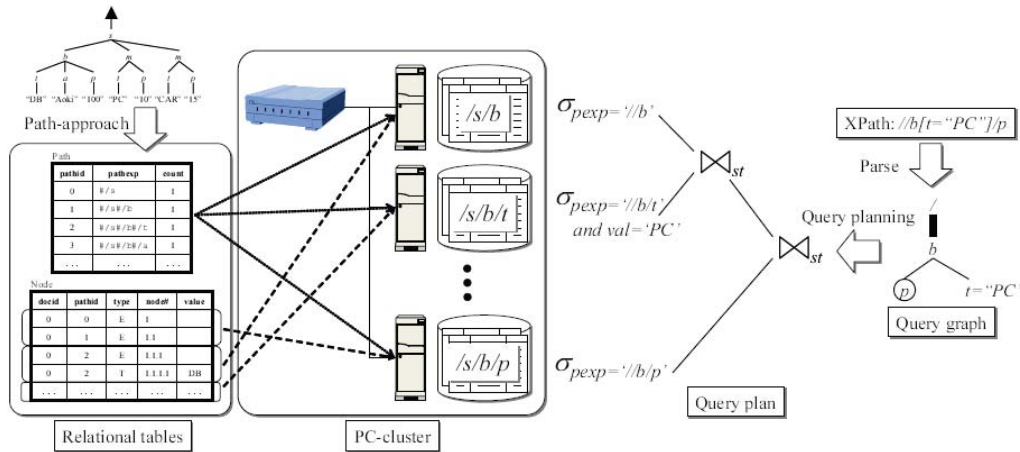


図 34 PMX の概要

問合せ処理は以下のように行われる。まず、得られた問合せを構文解析し、論理問合せプランを得る。論理問合せプランを葉からたどり、各オペレータを処理可能な(データを保持している)自律ディスクノードへ転送する。これを繰り返すことによって、問合せを分散処理することができる。

以上の手法をプロトタイプシステムに実装し、実験による性能評価を行った。100MB および 1TB の Xbench データに対して、単純に文書を分散する分散手法を比較手法として比較を行った。図 35 の結果からわかるように、提案手法はベースラインの比較手法に対し良好な問合せ処理性能を有することが示された。特に 1TB のデータでは、比較手法は実用的な時間内に問合せ処理を終えることができなかったのに対し、提案手法は実用的な時間で処理を終えることができる。

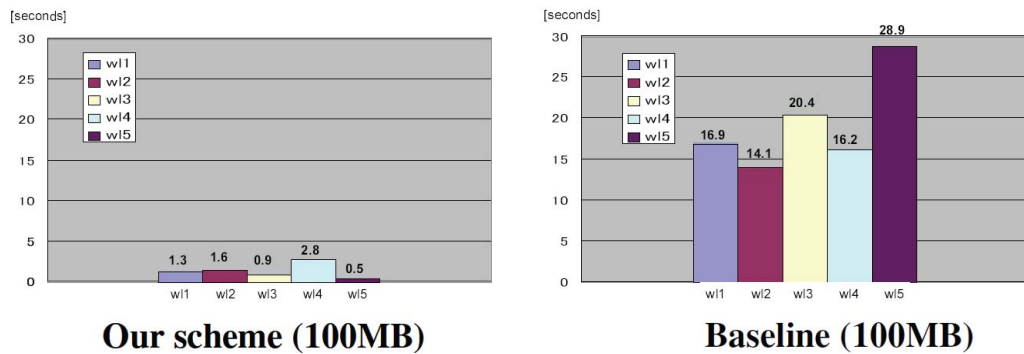


図 35 PMX の性能評価

#### (d) GMX: Holistic Twig Join に基づく大規模 XML データ処理方式

XML 問合せは一般に twig と呼ばれる複数の枝を有する小さな木構造によって与えられる。このとき XML 問合せとは、XML データ中で twig にマッチする全ての出現を列挙することに他ならない。twig 問合せを高速に処理するため、これまでいくつかのアルゴリズムが提案されているが、それらを総称して Holistic Twig Join と呼び、XML 問合せを高速化するために欠かせないアルゴリズムであるが、これまで Holistic Twig Join を分散処理する研究はほとんど存在しない。そこで本研究では、Holistic Twig Join を自律ディスククラスタ上で分散処理する方式について検討を行った。

Holistic Twig Join では、問合せとして与えられた twig に対し、XML データ中における全ての埋め込み可能な箇所を列挙することが目的となる。XML データは、あらかじめ同じ要



素名を持つノード毎に分類され、文書順にソートされた形で保存されているとする(これをノードストリームと呼ぶ)。Holistic Twig Join は、twig に関連する要素に対応するノードストリームを並行して先頭から順に走査することで解を探索する。ここでわかるように、Holistic Twig Join 処理を分散化するには、複数のノードストリームを適切に分割し、自律ディスクノード上に配置することが必要である。ここでいう適切な分割とは、1) 自律ディスクノード間の負荷が均衡化されている、2) 問合せ処理時に、自律ディスクノード間で余計なデータ転送が発生しない、ことなどを指す。

この目的のため、本研究では GMX (Grid-Metadata Model for XML) と呼ばれる手法を開発した(図 36 参照)。

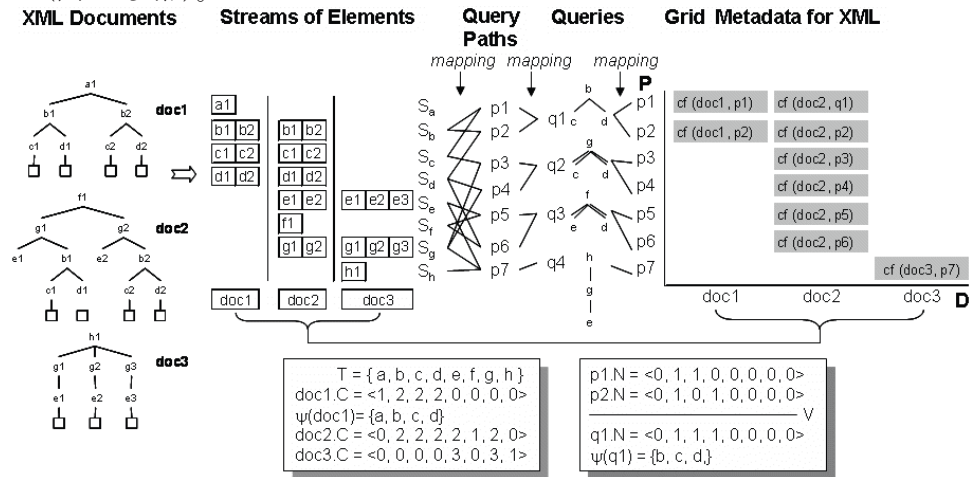


図 36 GMX の概要

GMXでは、XMLデータ集合およびtwig問合せの間の関連に着目し、XMLストリームのフラグメント化を行う。すなわち、問合せに出現する要素名の種類に応じて、問合せ同士の間にある種の相関を見ることができる。同様に、XMLデータ間の相関も出現する要素名、属性名、傾向によって与えることができる。さらに、問合せとそれにマッチする(し得る)文書という観点から、文書と問合せの間にも相関を与えることができる。これらを GMX と呼ばれるある種の多次元キューブ上に表現し、多次元キューブ上の各次元に対して操作を加えることにより、ある問合せに対して相関の高い(部分)ノードストリーム集合を、あるいは(部分)ノードストリーム集合に対して関連のある問合せ集合を導出することができる。

得られた(部分)ノードストリーム集合は、各自律ディスクノード間の負荷が均衡化するように分散配置される。問合せ処理時には、処理しようとする問合せが、各自律ディスクノードが保持しているノードストリームに関連するかどうかを調べた上で、関連する自律ディスクノードに問合せを転送し、処理を実行する。

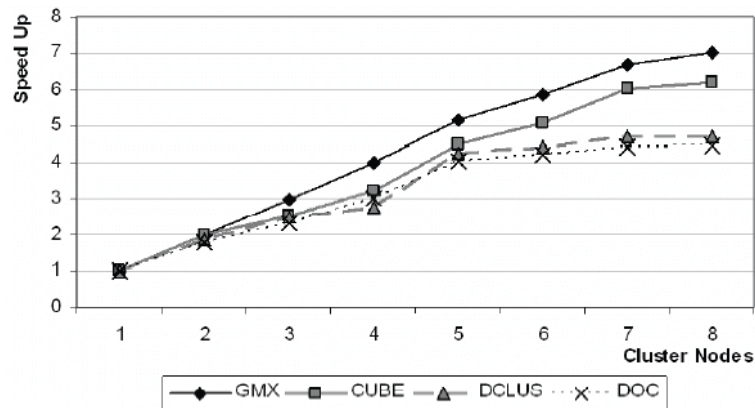


図 37 GMX の性能評価



提案手法を実装し、その有効性を実験により検証した。実験では、17種類のDTDからなる194文書に対し、GMXによる分割を行った上で自律ディスクノード上に配置した。図37の自律ディスクノード数を変化させたときの問合せ処理時間を見ると、提案手法がもっともノード数に対して線形となり、良好な問合せ性能を持つことが示された。

## (2)研究成果の今後期待される効果

XMLデータの特徴として、構造情報を持つことが挙げられる。オフィス文書ではしばしばキーワードによる文書検索が行われるが、構造情報(例えばタイトルや段落等の文書構造)を付加することにより、キーワードだけでは表現できないリッチな検索や、精度の高い検索が可能となる。また、ビデオデータの索引に利用されるMPEG-7もXMLで記述されており、例えば大量の映像アーカイブから、複雑な検索要求をXMLデータへの問合せとして映像を容易に取り出すことが可能となる。更に、DNA、タンパク質、天文データに代表されるXMLで記述された大規模な科学データについても、XMLデータの格納ならびに問合せが効率よく行うことが可能となる。

すなわち、自律ディスクのストレージとしてのスケーラビリティと、提案手法によるXMLデータのストレージへの格納・高性能分散XML問合せ処理のスケーラビリティが相乗的に働き、個人レベルのリッチなデスクトップサーチから、企業のストレージやデータセンタに格納されるオフィス文書やコンテンツ、科学技術データベース、さらにはインターネット上の検索エンジンに至るまで応用可能な技術である。

今後、更なるスケーラブルかつ高性能なXMLデータ処理のために、自律ディスクのグリッドやP2Pネットワークへの展開が必要である。また、その上での効率の良いXML処理のために、並列処理スケルトン、例えばMapReduceを利用したXQuery問合せ処理の検討が必要である。なお、XQuery問合せ処理のMapReduceに関しては、本研究プロジェクトの一環として現在研究中である。

### 3.3 先進ストレージシステムにおけるストレージネットワークの性能とディペンダビリティに関する研究(お茶の水女子大学 ストレージネットワーク研究グループ)

#### (1)研究実施内容及び成果

近年、インターネット技術の進展やマルチメディアアプリケーションの普及などにより、ユーザが蓄積し利用するデータ容量が爆発的に増加している。これに伴いストレージの増設、管理コストの増大が問題となっている。そこでSAN(Storage Area Network)が登場し、広く用いられるようになった。SANはサーバとストレージを物理的に切り離し、各ストレージとサーバ間を相互接続してネットワーク化したもので、これにより各サーバにばらばらに分散していたデータの集中管理が実現された。

一般にSANとしてはファイバチャネルを用いるFC-SAN(Fibre Channel - SAN)が広く利用されている。しかしFC-SANはファイバチャネルを用いているため高価となり、また接続距離に制約がある。一方、SANにIPネットワークを利用したIP-SANとしてiSCSIが期待されている。iSCSIは、これまでDAS(Direct Attached Storage)で使われてきたSCSIコマンドをTCP/IPパケット内にカプセル化することにより、サーバ(Initiator)とストレージ(Target)間でデータの転送を行う。今後インターネット技術の発展により、ギガビットクラスの回線の実現が期待され、iSCSIの有用性もさらに高まると考えられる。

現状において、SANは主にサーバサイト内のみでしか使用されていない。しかし遠隔バックアップ等を目的として、離れたサイトのサーバとストレージをSANで接続することが強く期待されている。そこで本研究では、VPN(Virtual Private Network)を利用することにより、ローカル環境で使用されているiSCSIを用いて広域ネットワーク上でリモートアクセスを行うことを検討した。

また、近年大規模なストレージシステムは多数のストレージノードをネットワーク結合した並列ストレージシステムとして構成されている。システムはストレージ仮想化機構により仮想化された単一の巨大ボリュームとして提供される。このような高機能並列ストレージシステム

を利用する場合、システムへのアクセスネットワークがボトルネックとなってしまうと考えられる。そこで本研究ではネットワークの性能と信頼性を高めるために、VPN 広域ネットワーク内を複数経路で接続し、iSCSI 並列ストレージアクセスの特性を解析、評価した。

**(a) VPN 複数経路接続 iSCSI ストレージアクセス：iSCSI 複数コネクション**

IP-SAN の代表的なプロトコルに iSCSI がある。iSCSI は SCSI コマンドを TCP/IP パケットでカプセル化する規格で、iSCSI により SAN を IP 機器だけで構築することが可能となる。一方で図 38 のように複雑な階層構成をとることになり、下位のプロトコルの限界性能を超えることはできない。また、iSCSI には長距離アクセスの実現が期待されているが、ギガビットクラスの太い回線を用いた場合の遅延帯域積の問題も指摘されている。そこで iSCSI における性能や信頼性を向上する手法の実現が求められている。

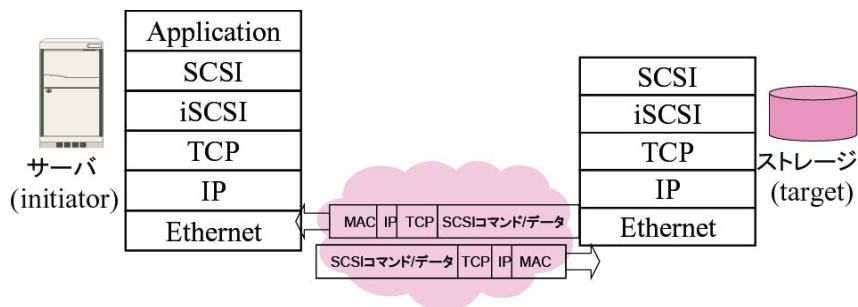


図 38 iSCSI

iSCSI は様々なチューニングを行うことが可能である。本実験で用いたニューハンプシャー大学が提供する UNH-iSCSI の実装では、1 つの iSCSI セッション内に複数の TCP コネクションを確立するように設定することができる。さらにこのコネクションをポート番号と対応付けることができる。つまり、図 39 に示すようにターゲットの 1 つの IP アドレス、1 つの iSCSI ドライブにポート番号の異なる複数のコネクションを接続することが可能である。本研究では iSCSI のこの特性を利用して、iSCSI 複数経路アクセスを実現した。

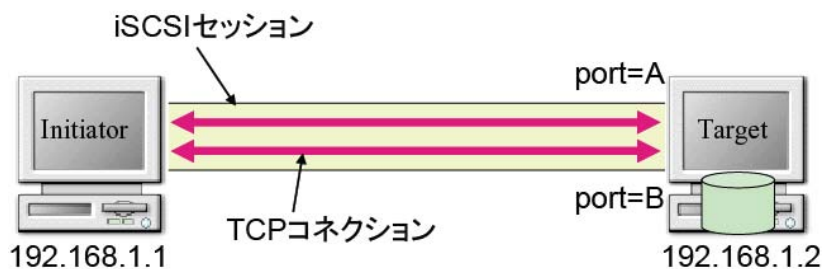


図 39 iSCSI 複数コネクション

**(b) VPN 複数経路接続 iSCSI ストレージアクセス：VPN 複数経路接続**

VPN は、インターネットや通信事業者が持つ公衆ネットワークを使って、拠点間を仮想的に閉じたネットワークで接続する技術である。安価であるという公衆網のメリットを活かしつつ、機密性の低さを暗号化等の別の方法で補うことにより、「実質的な専用網」を実現できるということが VPN の利点である。一方、専用網と異なりネットワークの品質は保証されない場合が多い。

本研究では非常災害対策などを目的とした iSCSI による遠隔バックアップなどの評価を行うため、VPN ルータで接続したリモート環境にネットワークストレージを設置し、iSCSI を広域ネットワーク環境に適用することを想定して実験を行った。この場合、広域ネットワーク内の VPN 越しにアクセスを行うため、VPN ルータを通ることによってネットワークの帯域幅が

制限され、スループットが著しく低下することが起こり得る。

さらに広域ネットワーク内は不安定な通信路であることが想定される。そこで本研究では、VPN 広域ネットワーク内を複数経路で接続することを考えた。これにより、データ転送の性能や信頼性、ネットワークの耐障害性なども向上すると考えられる。

ただし iSCSI 複数経路の構築は、アプリケーションなど上位層に対しては透過的に実現したい。本実験で用いた VPN ルータ Fujitsu Si-R570 はマルチルーティング機能を有している。マルチルーティング機能を使用すると、ポート番号などの情報を利用して同じあて先 IP アドレスを持つネットワークへ複数の経路を用いて送信することが可能となる。それぞれの通信内容に応じて通信経路を分離することが出来るため、片方の回線をバックアップ用に用いたり、音声データは専用線を用いそのほかの通信は公衆網を用いたりなど設定することができる。本研究ではこの機能を利用し、iSCSI 複数コネクション設定と対応付けることにより、コネクションごとに異なる経路を構築することを可能にした。また本研究で用いた VPN ルータは L3 における VPN を利用できる IP アクセスルータで、VPN の暗号化プロトコルには IPsec の DES, 3DES, AES を用いており、3DES 時には 500Mbps の暗号化速度を実現する。ただし、本研究の手法はこの環境に限らず、複数経路接続の設定ができるルータを用いて構築した VPN であれば適用することができる。

#### (c) 並列ネットワークストレージシステム: 並列ストレージシステム

近年、大規模なストレージシステムは多数のストレージノードをネットワーク結合した並列ストレージシステムとして構成される。そしてデータ郡をファイル、エクステント、ページ、ブロック等の粒度で分割を行い、分散格納する。データ配置管理の複雑さを利用者から隠蔽し、アクセス性能を最適化するため、システムはメタデータサーバや分散ディレクトリなどによるストレージ仮想化機構により仮想化され、単一の巨大ボリュームとして提供される。各クライアントからのアクセスリクエストは仮想化機構を通じて適切なストレージノードへと送られて処理される。並列ストレージシステムでは、ノード故障によるデータ喪失を防ぐため、他のノードのバックアップデータを保持する構成をとるものが多い。

並列ストレージシステムの研究において、これまでに可用性やスケラビリティに優れた高機能並列ストレージシステムである自律ディスクが提案されている。自律ディスクのデータ配置戦略において、データマイグレーションに起因するノード負荷上昇を効果的に分散するために複製データを利用し、マイグレーション移動元移動先ノードが性能要件を満たさない状態に陥ることを防ぐ手法 Replica-assisted Migration が提案されている。自律ディスクでは、Primary-Backup の one-copy 複製により管理する障害復旧用バックアップデータを Chained Declustering 複製配置戦略に基づきストレージノード間で保持し合うことで信頼性を高めている。

#### (d) 並列ネットワークストレージシステム: iSCSI 並列ネットワークストレージシステム

前節で述べた並列ストレージシステムでは、複数のストレージデバイスが束ねられ単一の巨大なボリュームを持つストレージとして提供される。これを iSCSI のようなネットワークストレージで Target として利用しようとした場合、実際には複数の Target が存在するが、サーバ側からは単一のデバイスを持つストレージとして見えるように構成することが可能である。

Target に並列ストレージのように性能の高いストレージシステムを用いた場合、ネットワークストレージのボトルネックはネットワーク部分となる可能性が高い。従ってネットワーク部分の能力を高めることが、全体の性能を上げるために重要である。このためには、(b) で述べた VPN マルチルーティングのように複数経路接続の手法が有効である。

従って本研究では図 40 に示すように、Target に並列ストレージシステムを利用した際に、ネットワークを VPN 複数経路接続とすることにより性能向上を図る手法を提案し、その性能評価を行う。各 Target は単一のデバイスを持ち、サーバも単一のインタフェースからアクセスを行っているにも関わらず、ネットワーク部分は複数経路化され、Target も複数の Target の集合から構成される並列ストレージシステムとなっている。

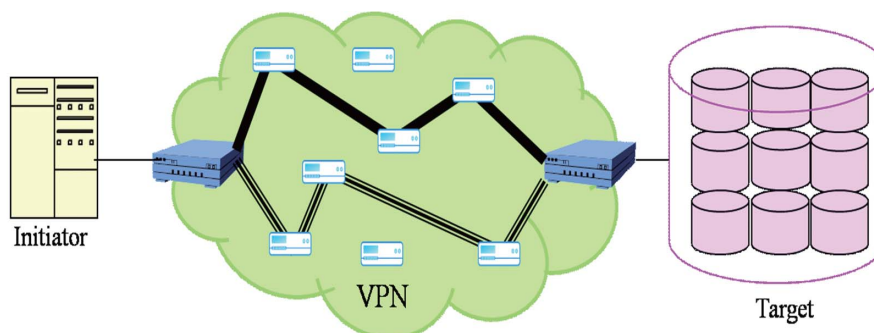


図 40 iSCSI 並列ネットワークストレージシステム

#### (e) 本研究の評価内容と関連研究

本研究では以上に紹介した知識と技術を利用し、VPN 複数経路接続時における iSCSI 並列ストレージアクセスの性能評価を行う。関連する研究としては、以下のようなものが挙げられる。

iSCSI と TCP の関係性を評価した研究としては、iSCSI のソフトウェア実装を使用した場合と TCP の性能を向上させる TOE (TCP Offload Engine) や HBA (Host Bas Adaptor) を使用した場合との比較などが行われている。またファイル操作やベンチマークプログラム実行時における iSCSI と NFS との比較も報告されている。

iSCSI ストレージアクセスにおいて TCP 輻輳ウィンドウを制御する研究としては、輻輳ウィンドウ値を動的にコントロールする手法がある。この手法は、まず Target の OS のカーネルに輻輳ウィンドウモニタ関数を挿入し、これによりモニタした輻輳ウィンドウの変化を観察して、Initiator にその値を通知する。通知を受けた Initiator は輻輳ウィンドウの値に基づきブロックサイズを再指定して、シーケンシャルリードアクセスを行うというものである。この手法を適用し輻輳ウィンドウを限界値で一定に保った場合には、高遅延環境において最大 28% のスループットの向上が確認されている。またトランスポート層を標準的な TCP Reno から他の TCP に置き換えて iSCSI に適用した際のシミュレーション結果なども報告されている。

一方 iSCSI 複数コネクションに関する研究として、広域 IP 網を介した長距離アクセス向けに iSCSI および関連プロトコルレイヤのプロトコルチューニングの検討が行われ、その有効性が確認されている。また TCP 複数コネクションにおいて、各々の TCP 制御情報をコネクション間で共有することにより複数コネクション間の公平性を保つ Fair-TCP を、iSCSI 複数コネクションに適用した結果も報告されており、情報を共有することによって複数コネクションにおける制御がより適切なものとなり、性能向上が見られた。しかし iSCSI 複数コネクションを複数の経路に割り当てて利用する手法に関しては、これまで検討が行われてこなかった。

#### (f) 複数経路並列ストレージアクセス実験システム

本実験では、VPN ルータを用いて複数経路を構成し、iSCSI 並列ストレージアクセスを実行した時の、性能と輻輳ウィンドウを評価するために図 41 に示す実験環境を構築した。iSCSI ストレージアクセスを行う Initiator とストレージを提供する 2 台の Target の間に VPN ルータを 4 台を挟み、複数経路アクセスが実行できるように構築した。さらにそれぞれの経路に、遠距離アクセスを想定して人工的な遅延装置である FreeBSD Dummynet を挿入した。

Initiator と Target には、OS は Linux2.4.18-3、CPU は Intel Xeon 2.4GHz、Main Memory は 512MB DDR SDRAM、NIC は Intel Pro/1000MT Server Adapter on PCI-X (64bit、133MHz)、iSCSI は UNH IOL reference implementation ver.3 on iSCSI Draft 18 を用いた。Dummynet1 には FreeBSD4.9-RELEASE、Dummynet2 には FreeBSD6.2-RELEASE を用いた。また VPN ルータには Fujitsu Si-R570 を用いた。これは 3DES 暗号化速度最大 500Mbps を実現する。



ここで iSCSI ストレージの 2 台の Target を用いて RAID0 を構成し iSCSI 並列ストレージシステムを構築した。この実験システムにおいて VPN ルータのマルチルーティング機能を用いて遠隔ストレージアクセスの通信制御を行った。以下では iSCSI リードアクセス、すなわち Target から Initiator へデータが転送される場合について説明しているが、逆も基本的に同じである。

まず VPN 単数経路通信の場合は、図 42 に示す通り同一経路上を 2 つのセッションが張れるように設定した。一方 VPN 複数経路通信の場合は図 43 に示す通りセッションごとに経路が異なるように VPN ルータの設定を行った。このとき Target から送られるパケットは、右上の VPN ルータに送られるように経路を設定する。そして VPN ルータのマルチルーティング機能により、左の 2 つの VPN ルータ宛にパケットが転送される。このとき、ポート番号の違いにより iSCSI セッションごとに上下の VPN ルータに分かれるように設定した。また通常並列ネットワークストレージにおいて通信の口 (アクセスインタフェース) は 1 つであり、単一のルータが入口となっていることが想定されるが、参考として図 44 に示す通り 2 つの送受信口がある独立複数経路の場合の性能も測定し比較対象とした。

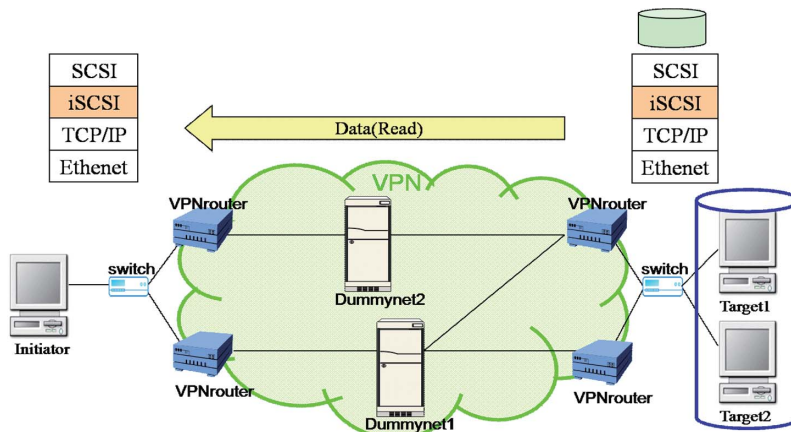


図 41 実験システムの概要

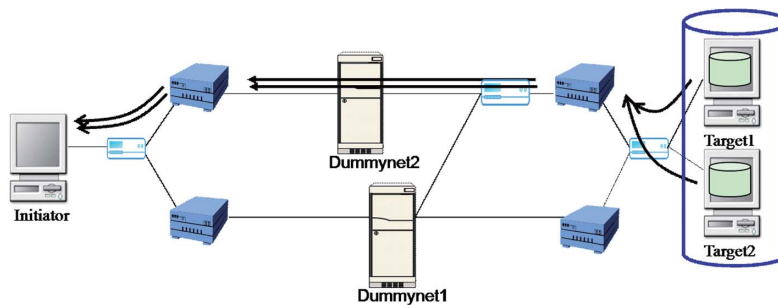


図 42 VPN 単数経路

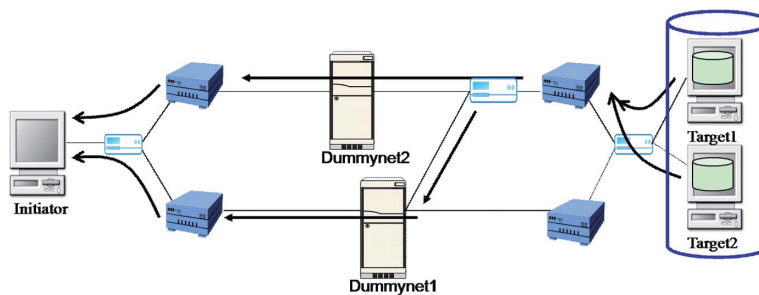


図 43 VPN 複数経路



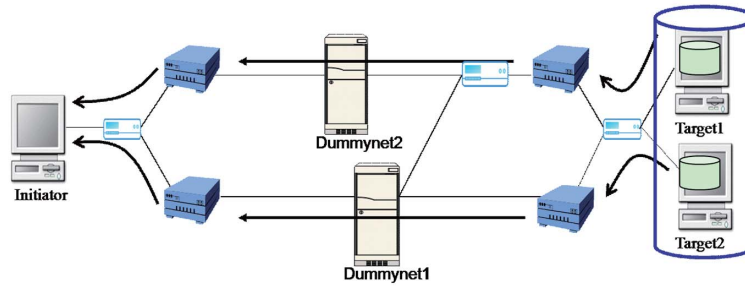


図 44 VPN 独立複数経路

(g) 性能測定結果

iSCSI 並列ストレージアクセスにおいてブロックサイズを変化させた時の各アクセス手法によるスループット比較のグラフを図 45 に示す。この実験において遅延は設定していない。またグラフ中の Target1 は図 41 における Target1 単一のストレージへのアクセスを行った場合の性能を示し、Target2 も同様である。

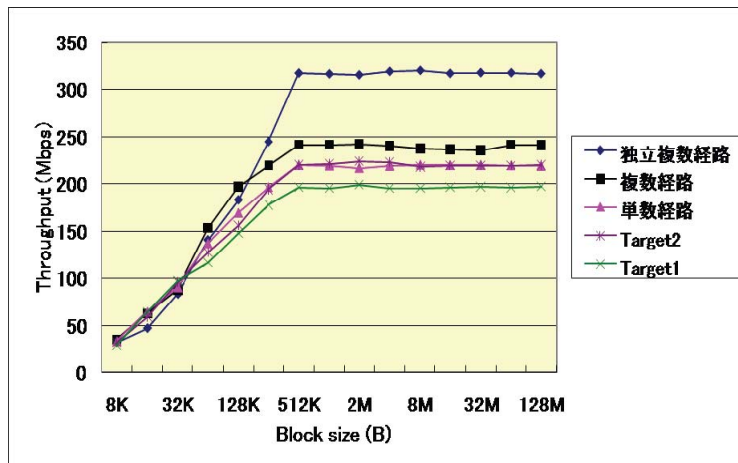


図 45 ブロックサイズとスループット比較

どの場合でもブロックサイズを増加していくとスループットは増加していくが、ブロックサイズ 512KB を過ぎたところからスループットはほぼ一定となる。

次にアクセス方法の違いによる性能変化を見てみる。Target1 や Target2 と単数経路の性能はほぼ同一となっている。このように 2 台の Target を RAID0 としストレージ側では性能が向上しているにもかかわらず、ネットワークなど他の部分がボトルネックとなっていることがわかる。

また単数経路の場合と複数経路の場合を比較すると複数経路の性能が向上しており、本研究で用いた VPN 複数経路接続によるアクセス手法は有効な手段であることがわかる。性能向上の理由としては、VPN 内の経路が 2 倍の容量を持つようになり、さらに Initiator 側の VPN ルータでの処理が分散されたためであろう。一方で、独立複数経路を用いた場合に比べ性能が抑えられてしまっている理由としては、Target 側での VPN ルータにおける暗号化処理等がボトルネックとなっていると考えられる。しかし現在 VPN ルータは急速に進歩しており、VPN ルータの性能を十分に利用できる本アクセス手法は今後より有効な手段になると考えられる。

次に、片道遅延時間を変化させた時の iSCSI 並列ストレージアクセススループット比較のグラフを図 46 に示す。この実験においてブロックサイズは 2MB に設定した。また複数経路において経路ごとの遅延時間は同じ値を設定した。どの場合も片道遅延時間を長くするとスループットは急激に減少した。また、各アクセス方法における変化の様子を見てみると、

遅延時間が短い時には複数経路は単数経路の性能に近い値をとっているが、遅延時間を長くすると独立複数経路の性能に近い値をとるようになっていく。これは遅延時間が短い時は複数経路にしても Target 側の VPN ルータでの処理がボトルネックとなってしまうため、独立複数経路の場合ほど性能は良くなかったが、高遅延環境の場合になると VPN ルータでの処理より経路長による性能劣化の方がボトルネックとなり、独立複数経路の性能と差がなくなってきたと考えられる。このように本実験環境においては VPN ルータの性能がボトルネックとなっていることが確認でき、今後 VPN ルータの進歩により、独立複数経路に近い性能を出すことが可能になるであろう。

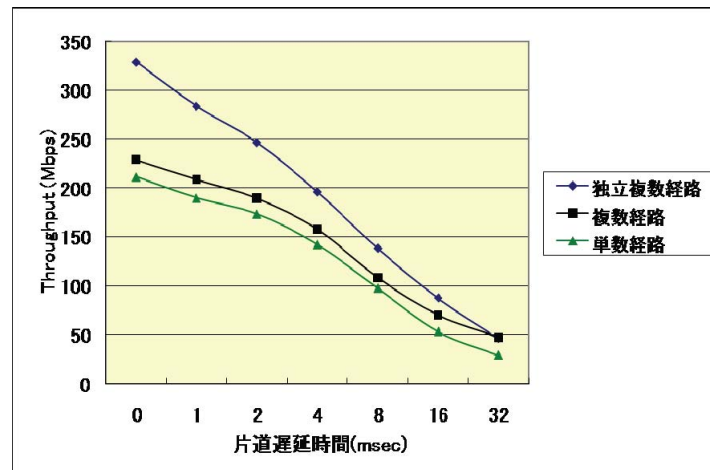


図 46 片道遅延時間とスループット比較

## (2)研究成果の今後期待される効果

本研究では、iSCSI 並列ストレージアクセスにおいて VPN 接続を単数経路、複数経路に変化させた時のスループットの違いと輻輳ウィンドウの振舞いを観察し、比較した。その結果、単一の Target を用いた性能と 2 台の Target を RAID で構築し単数経路でアクセスした性能はほぼ同一となっていた。一方本研究で用いた VPN 複数経路接続時には性能が向上し、複数経路にすることで信頼性だけでなく性能向上も図れることがわかった。

また遅延時間の短い時には複数経路は単数経路の性能に近い値をとっているが、遅延時間を長くすると独立複数経路の性能に近い値をとるようになった。これは高遅延にすることで VPN ルータでの処理より経路のほうがボトルネックとなってきたためであると考えられる。このように、VPN ルータの進歩により本アクセス手法は今後より有効な手段になると考えられる。

本実験では RAID 構成のストレージを用いたが、本研究で提案した手法は他の高機能ストレージにも適用可能である。例えばストレージに自律ディスクを用い、VPN 接続した広域ネットワークで接続してストレージアクセスを行った場合にも、性能や信頼性の向上という面で同様の効果を期待することができる。

さらに VPN 複数経路接続は、クライアントからサーバへのリモートアクセスという形態に限定されない。例えば自律ディスクの各ノードを広域環境に分散させ、その間をネットワークで接続して広域分散型の自律ディスクを構築することが考えられる。その際にも接続ネットワークに VPN 複数経路接続を用いれば、自律ディスクのノード間通信の性能と信頼性を向上させることができると考えられる。

今後先進ストレージシステムが社会で広く用いられるようになると、その性能とディペンダビリティの保証が要求されると考えられる。ストレージそのものの保証については、自律ディスクなどのメカニズムにより実現が期待されるが、その場合接続ネットワークがボトルネックになってしまう可能性があり得る。本研究における提案は、接続ネットワーク部分の性能とディペンダビリティの向上に貢献するものである。

### 3.4 プロトタイプ試作及び実用環境研究(富士通(株) システムアーキテクチャグループ)

#### (1)研究実施内容及び成果

システムアーキテクチャ研究グループは、先進ストレージシステムの機能を実証するための開発プラットフォームとして、自律ディスクプロトタイプのシステムアーキテクチャについて検討及び試作を行った。また実環境での機能、性能検討を目的として自律ディスクソフトウェアへの機能追加と汎用環境へ対応する I/F の追加を行った。これらプロトタイプ試作機及び自律ディスクソフトウェアを各グループへ提供し、先進ストレージシステムの機能検討と実環境での検証を可能とした。

自律ディスクプロトタイプの検討においてはストレージとして要求される機能、性能が多岐に渡るため各グループの研究内容を考慮し、基本となる自律ディスクソフトウェアの構成要素をコンパクトに実装可能なコアモジュールとして用途に応じたストレージデバイス(HDD)と組み合わせる構成とした。このコアモジュールには以下の要件、仕様を設定した。

#### ・デバイス要件

- (1)ストレージデバイス(3.5" HDD)と同程度の大きさで実装可能
- (2)既存の機能/環境が実現容易(IA-32 アーキテクチャ)
- (3)開発/テスト環境が提供可能
- (4)大規模、簡易構成の検証可能

#### ・性能要件

- (1)CPU 1GHz/メモリ 1Gbyte 以上
- (2)消費電力:10W 以下

これらの要件を検討した結果、性能と消費電力を同時に満足する CPU として Transmeta 社 efficeon を、HDD I/F として SATA 方式を採用した。IA-32 アーキテクチャを採用することで linux 等の汎用 OS が使用可能であり、各グループでの開発環境の互換性を図り効率的な研究開発を行うことが可能となった。また SATA 方式はドライブが比較的安価でポータブルマルチプライヤを用いることで多数(本試作機のコアモジュールでは最大 20 台)の HDD を接続可能であるためデータ規模に応じた自由な構成が可能である。

このコアモジュールと 2.5" HDD を用いてコンパクトな筐体によって設置、運用の自由度を高め、オフィス・家庭向け軽量ストレージを想定した S-Type 及び 3.5" HDD 5 台と組み合わせたブレード構成を採用し、5U 筐体に 8 ブレードを搭載し 40 台の HDD 構成を可能とするデータセンタ向け大規模ストレージを想定した B-Type の二つの試作機を開発した。これらの試作機を用いて各研究グループで映像ストレージ、ネットワークストレージ及びデータベースストレージとして実環境での実証実験を可能とした。

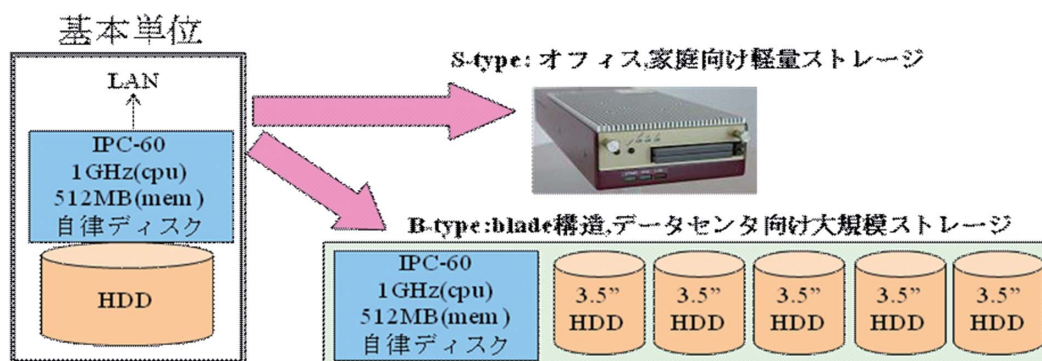


図 47 プロトタイプ基本アーキテクチャ



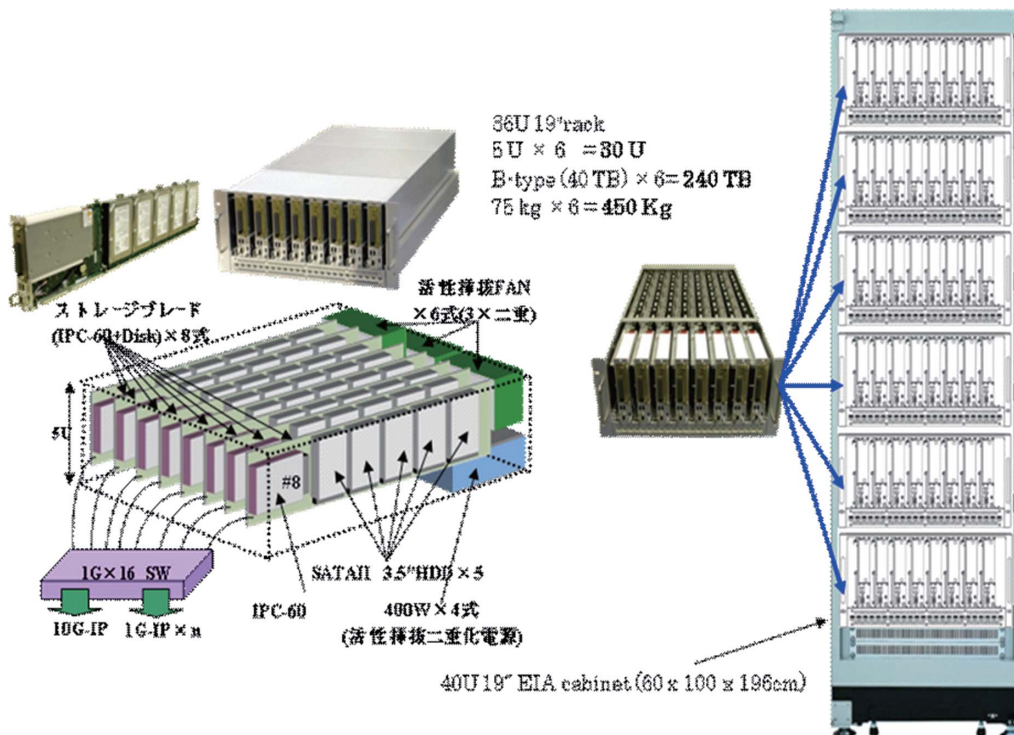


図 48 B-Type 構成例

本グループは先進ストレージ研究統轄・推進グループ及び高度メディア蓄積・管理研究グループと共同で自律ディスクソフトウェアへ大容量データを蓄積するための機能追加と機能強化を行った。メディア蓄積では映像を扱うためリアルタイム性と耐障害性が求められるため、並行制御アルゴリズム改善成果の反映及び仮想 Partition 方式による Fat-Btree 高速復旧による高速化を実現した。また HDD の状態監視機能によって自律的にデータ回避を行う機能の実証実験も行い、自律ディスク機能の有用性を示した。

データアクセスの汎用性を高めるため最も普及している Windows クライアントからも容易にアクセス出来ることを目的として CIFS I/F の実装を行った(図 49)。本インタフェースを介し、利用者から透過的に負荷バランス、容量バランス、ノード追加/削除、故障対策、リカバリといった機能が、十分な性能を保ちながら、実際に実現できていることを確認できる。このため CIFS I/F そのものと CIFS 向けメタデータ処理機能の開発実装を行った。

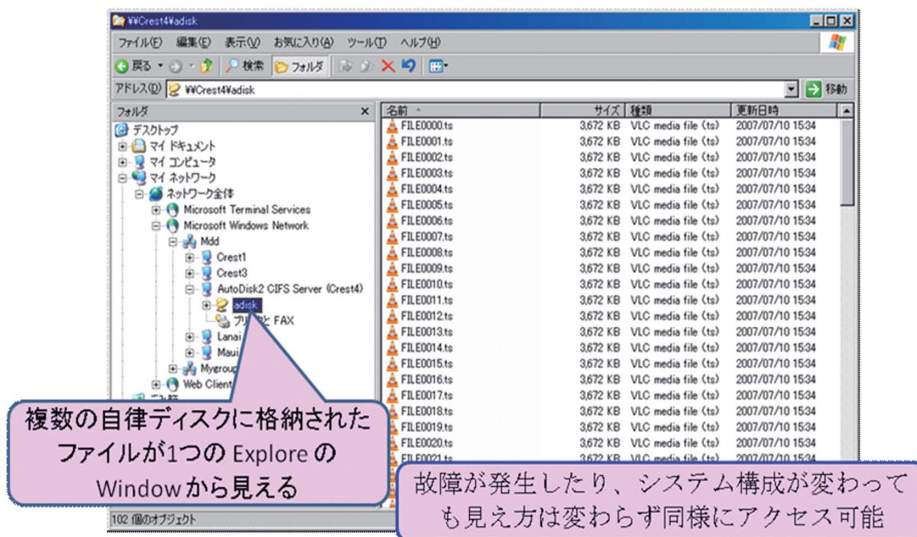


図 49 CIFS I/F を通して Explore からプロトタイプ中のファイルを見たところ

同時に各研究グループからのアクセス性能に対応した改善を行うことでデータベース及びネットワークストレージとしての実証実験を可能とし、汎用用途においても十分な性能が得られることを実証した。利用者から透過的に負荷バランス、容量バランス、ノード追加/削除、故障対策、リカバリといった機能が、十分な性能を保ちながら、実際に実現できていることを確認できるようにした。また、プロトタイプのコラスタ内のデータ量の配分や各自律ディスクの転送性能等を確認するためのモニタリングソフトウェアも開発した(図 50)。

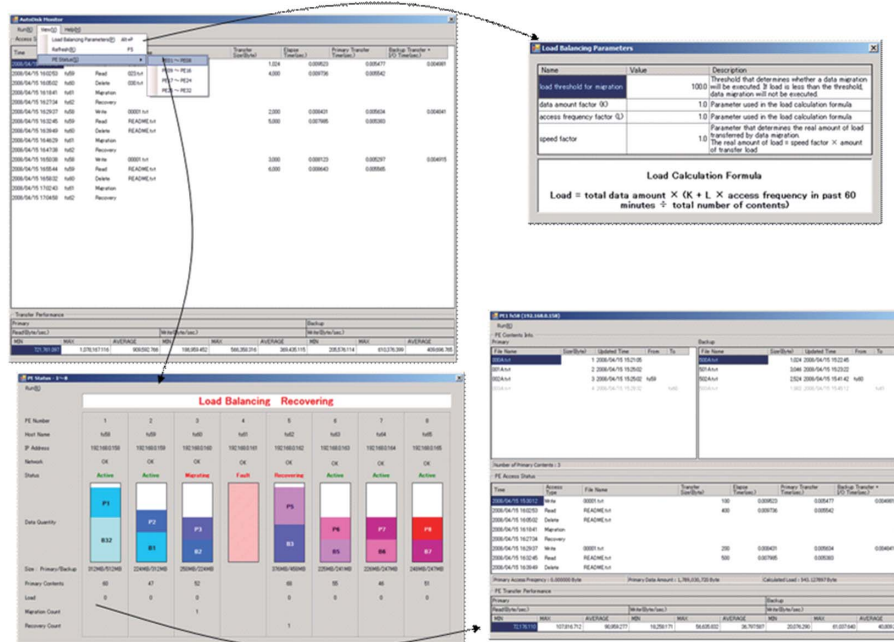


図 50 モニタリング画面の構成と画面遷移

## (2)研究成果の今後期待される効果

本プロトタイプの試作及び各種実環境における実証実験より自律ディスクをベースとした先進ストレージシステムが分野を問わず実用可能であることを確認した。試作に用いたCPUはIA-32の汎用アーキテクチャであり、現在の普及クラスで十分な処理性能である。またストレージI/Fも普及クラスのSATAで構成可能であり将来においても性能/価格も問題なく構成可能と考えられる。これらの成果をベースとしたモジュール型処理エンジン、ストレージデバイス(HDD)を組み合わせによるストレージシステムは管理と性能を両立する先進ストレージシステムとして有用であると考えられる。

### 3.5 関連技術動向調査(日立、ストレージ技術動向調査グループ)

国内外の関連研究に関して調査を行い、提案手法の位置づけの明確化を行った。

#### (a) 大規模自律分散型ストレージに関する関連技術動向に関する調査結果

負荷分散や耐障害性の向上、また運用管理容易化を目的として、データを自律的に分散配置する自律分散型ストレージが注目されている。

UC BerkeleyによるOceanStoreは、P2Pによる広域分散型ストレージである。データがインターネット上に広く分散するOceanStoreの形態では、ファイルやディレクトリのコンシステンスを維持するのが非常に困難となる。したがって、OceanStoreではルーズコンシステンスを採用し、ファイルの更新機能はサポートせず、全て新規置き換えとなる。

CMUによるSelf\* Storage(セルフスターストレージ)は、データセンタ内の大規模ストレージを自律運用するための研究である。標準のNFS、CIFSインタフェースをサポートしているが、ファイルシステムツリーを管理するメタデータサーバが単一ノードのため、耐障害性、負荷分散の面で強化が必要と考える。



Google File System(GFS)は、多数の安価な PC サーバにより分散ファイルシステムを実現している。GFS 内のファイルにアクセスするためには一般的なネットワークファイルシステムは利用できず、専用のクライアントソフトが必要となる。ファイルシステムツリー管理に関しては、Self-\* Storage と同様に単一のマスターサーバがツリー構成を管理するが、ログを保存しておくことによりマスターサーバ障害時にはセカンダリーサーバがツリー構成の管理を引き継ぐ事ができる。これにより耐障害性を高めている。

また、Hadoop プロジェクトは、GFS のオープンソース版の開発を行っている。このうち Hadoop Distributed File System(HDFS)が GFS に相当する。ほとんどの GFS の機能が実現されている。

以上の関連技術に対し「自律ディスク」では、FAT-Btree の採用により分散ストレージにおいても、タイトなコンシステンシを実現し、ファイルの更新機能を実現している。また、Fat-Btree では親ノードが PE 間で同期されているため、ある PE が障害に陥っても、ファイルシステムが部分的に維持できるため縮退運用が可能な方式となっている。これは他の自律分散型ストレージにない特徴である。これらの特徴を生かし、従来の自律分散型ストレージでは適用困難であった、よりミッションクリティカルなシステムへの応用が期待できる。

#### (b) センター間広域分散ストレージに関する関連技術動向に関する調査結果

近年、Amazon や Google などの Web 2.0 系企業のシステムにおいて、従来はデータセンタ内の閉じた環境のみで運用されていた分散ストレージを、複数のデータセンタを跨いでスケールアウトし、仮想的に一つのデータセンタとして運用しようとする動きがある。

Amazon は、複数データセンタを跨いで分散するストレージノードで構築される独自の分散ストレージ(Dynamo)を開発した。

Google でも、従来データセンタ内の閉じた環境で運用されていた分散ストレージ(GFS)を、複数のデータセンタを跨いでスケールアウトし、仮想的に 1 つのデータセンタとして運用しようとする動きがある。

このような動向に対し、「ストレージネットワーク研究グループ」による VPN 複数経路を用いた iSCSI アクセス性能評価の取り組みは、遠隔データセンタ間での効率的なデータアクセスに適用可能であると考えられる。

また「高度メディア蓄積・管理手法研究グループ」の取り組みではネットワークトラフィックを低減可能にする分散 XQuery 問合せ処理を実現しており、広域分散ストレージ上での応用が期待される。

## § 4 研究参加者

①先進ストレージ研究統轄・推進グループ(先進ストレージシステムにおけるデータ管理機能の研究)

	氏名	所属	役職	研究項目	参加時期
○	横田 治夫	東工大	教授	研究チーム全体の統轄	H15.10～H21.3
	藤原 英二	東工大	教授	システムの高信頼化に関する検討	H15.10～H21.3
	奥村 学	東工大	教授	コンテンツ対応の検討	H15.10～H21.3
	山岡 克式	東工大	准教授	分散処理に関する検討	H15.10～H21.3
	小林 亜樹	工学院大学	准教授	分散ストレージ間通信コストの検討	H15.10～H21.3
	小林 隆志	名古屋大学	准教授	データ管理処理に関する検討	H15.10～H21.3

*	梁 文新	東工大	研究員	管理情報記述手法に関する検討	H15.10～H21.3
	那須 紀穂	東工大	事務補佐員	研究チーム内における経理的業務および庶務的業務	H15.10～H21.3
	渡邊 明嗣	東工大	博士課程	アクセスと負荷バランスの両立の検討	H15.10～H21.3
	Neila Ben Lakhel	東工大	博士修了	データ管理分散化に関する検討	H15.10～H19.3
	Ouyang Xiangyong	東工大	博士退学	ストレージシステム管理に関する検討	H15.10～H19.9
	小林 大	東工大	博士修了	通常処理と管理のプライオリティ管理	H15.10～H20.3
	Luo Min	東工大	博士課程	分散データ管理に関する検討	H19.10～H21.3
	花井 知広	東工大	修士修了	通常処理とデータ管理処理のプライオリティ管理	H15.10～H17.3
	山口 宗慶	東工大	修士修了	情報ライフサイクル管理の導入に関する検討	H15.10～H17.3
	吉田 誠	東工大	修士修了	データ管理処理に関する検討	H15.10～H17.3
	岡本 拓明	東工大	修士修了	ストレージ応用に関する検討	H15.10～H18.3
	加藤 英之	東工大	修士修了	データ管理高信頼化に関する検討	H15.10～H18.3
	中野 真那	東工大	修士修了	ストレージ中のデータの版管理の検討	H15.10～H18.3
	越智 悠太	東工大	修士修了	ストレージ応用に関する検討	H16.4～H20.3
	仲野 亘	東工大	修士修了	メタデータの応用に関する検討	H17.4～H19.3
	三木 健士	東工大	修士修了	ストレージ応用に関する検討	H17.4～H19.3
	山元 理絵	東工大	修士修了	メタデータ書式に関する検討	H17.4～H19.3
	吉原 朋宏	東工大	修士修了	分散ストレージの排他制御に関する検討	H17.4～H19.3
	片居木 誠	東工大	修士修了	分散ストレージのスループット改善検討	H17.9～H20.3
	高橋 昭裕	東工大	修士修了	メタデータの格納に関する検討	H18.4～H20.3
	並木 悠太	東工大	修士修了	ルールによるコンテンツ管理の検討	H18.4～H20.3
	渡部 徹太郎	東工大	修士修了	ストレージ応用に関する検討	H18.4～H20.3
	太田 健介	東工大	修士修了	ルールによるコンテンツ管理の検討	H17.9～H18.4

	高山 一樹	東工大	修士課程	分散ストレージのセキュリティに関する検討	H18.9～H21.3
	Do Ngoc Hung	東工大	修士修了	メタデータによる管理に関する検討	H18.9～H20.9
	小田切 健一	東工大	修士課程	ストレージ上検索に関する研究	H20.4～H21.3
	村上 翔一	東工大	学部卒業	メタデータの格納に関する検討	H17.9～H18.3
	濱田 圭二郎	東工大	学部卒業	データ管理分散化の実装に関する検討	H18.9～H19.3
	Le Hanh Hieu	東工大	学部卒業	ストレージ応用に関する研究	H19.4～H20.3
	吉野 悠二	東工大	学部卒業	XML ストレージに関する研究	H19.4～H20.3
	謝 放	東工大	学部卒業	メタデータ利用に関する研究	H18.10～H19.9
	小林 径	東工大	学部	XML ストレージに関する研究	H20.4～H21.3
	呉 怡	東工大	学部	コンテンツ格納に関する研究	H20.4～H21.3
	Matthias Karlsson	東工大	研究生修了	メタデータによる管理に関する検討	H18.9～H19.9

②高度メディア蓄積・管理手法研究グループ(高度メディア蓄積・管理手法に関する研究)

	氏名	所属	役職	研究項目	参加時期
○	宮崎 純	奈良先端大	准教授	高信頼分散ストレージシステムを利用した高度メディアの格納と検索法の開発	H15.10～H21.3
	植村 俊亮	奈良産業大	教授	ユビキタスデータベースアーキテクチャの検討	H15.10～H21.3
	天笠 俊之	筑波大	講師	センサデータ処理手法の開発	H15.10～H21.3
	波多野 賢治	同志社大	准教授	ストリーム処理手法の開発	H15.10～H21.3
	中島 伸介	京都産業大	准教授	ユビキタスデータベース検索手法の検討	H18.4～H21.3
	的野 晃整	奈良先端大	博士修了	ユビキタスデータベースの開発	H15.10～H17.3
	中尾 伸章	奈良先端大	修士修了	センサデータ処理手法の開発	H15.10～H17.3
	羅 勇	奈良先端大	博士退学	高信頼分散ストレージシステムを利用した高度メディアの格納と検索法の開発	H15.10～H17.3
	潘 洪涛	奈良先端大	博士退学	ストリーム処理手法の開発	H15.10～H17.3

	藤原 勤	奈良先端大	博士退学	ユビキタスデータベース 処理手法の開発	H16.4~H19.3
	栗田 裕人	奈良先端大	修士修了	ユビキタスデータの管理 アルゴリズムの開発	H16.4~H18.3
	浜野 泰男	奈良先端大	修士修了	ユビキタスデータの移動 アルゴリズムの開発	H16.4~H18.3
	中窪 仁	奈良先端大	博士退学	ユビキタスデータベース 検索手法の開発	H18.4~H18.6
	奥 健太	奈良先端大	博士課程	ユビキタスデータベース のコンテンツ管理の検討	H18.4~H21.3
	油井 誠	奈良先端大	博士課程	ストリーム問合せプロセス サ開発	H18.4~H21.3
	伊藤 智博	奈良先端大	修士修了	ユビキタスデータベース の検索手法の開発	H18.7~H20.3
	大阪 知克	奈良先端大	修士課程	ユビキタスデータベース のキャッシュ制御の検討	H18.7~H21.3
	中本 レン	奈良先端大	修士修了	ユビキタスデータベース のコンテンツ管理の検討	H18.7~H20.3
	吉井 迪利	奈良先端大	修士課程	XML ストレージの自動チ ューニングの検討	H18.7~H21.3
	松原 裕貴	奈良先端大	修士課程	キャッシュ指向のスト レージチューニングの 検討	H19.7~H21.3
	宮野 耕一	奈良先端大	修士課程	XML ストレージのアクセ スパターン解析	H19.7~H21.3
	城戸健太郎	筑波大	修士修了	XML ストレージの分散 処理	H17.4~H19.3
	Imam Machdi	筑波大	博士課程	XML ストレージの分散 処理	H19.4~H21.3

③ストレージネットワーク研究グループ(ストレージネットワークの性能とディペンダビリティに関する研究)

	氏名	所属	役職	研究項目	参加時期
○	小口 正人	お茶の水女子大学	教授	ストレージネットワークの性能とディペンダビリティ評価	H18.3~H21.3
	神坂 紀久子	お茶の水女子大学	博士課程	ストレージネットワークの信頼性検討	H18.3~H20.3
	武田 裕子	お茶の水女子大学	修士課程	ストレージネットワークのスループット向上手法の開発	H18.3~H19.3
	千島 望	お茶の水女子大学	修士課程	ストレージネットワークのネットワークパラメータ評価	H18.3~H20.3
	松井 愛子	お茶の水女子大学	修士課程	ストレージネットワークのユビキタス環境における基本性能評価	H18.3~H20.3

	原 明日香	お茶の水女子大学	修士課程	ストレージネットワークの信頼性検討	H19.3～H21.3
	新井 絵美	お茶の水女子大学	修士課程	ストレージネットワークのプロトコル評価	H20.4～H21.3
	宇野 美穂子	お茶の水女子大学	修士課程	ストレージネットワークのユビキタス環境における基本性能評価	H20.4～H21.3
	比嘉 玲華	お茶の水女子大学	修士課程	ストレージネットワークのスループット向上手法の開発	H20.4～H21.3

④コンテンツ対応メタデータ研究グループ(コンテンツ対応データ管理の研究)

	氏名	所属	役職	研究項目	参加時期
○	和泉 吉則	NHK 放送技術研究所	主任 研究員	メタデータ検討	H17.5～H19.3
	田口 亮	NHK 放送技術研究所	研究員	コンテンツ対応検討	H15.10～H19.3
	上原 年博	NHK 放送技術研究所	主任 研究員	メタデータ検討	H15.10～H17.5

⑤システムアーキテクチャ研究グループ(システムアーキテクチャの研究)

	氏名	所属	役職	研究項目	参加時期
○	飯田 安津夫	富士通(株)	部長	全体アーキテクチャ検討	H20.10～H21.3
	太田 光彦	富士通(株)	部長付	全体アーキテクチャ検討	H15.10～H20.9
	江尻 革	富士通(株)	研究員	ハードウェア実装	H15.10～H19.9
	戸田 誠二	富士通(株)	研究員	ソフトウェア実装	H16.04～H21.3

⑥ストレージ技術動向調査グループ(最新のストレージシステムの技術動向の研究)

	氏名	所属	役職	研究項目	参加時期
○	中村 隆喜	日立製作所システム開発研究所	研究員	ストレージ技術動向調査	H20.4～H21.3
	山崎 康雄	日立製作所システム開発研究所	主任 研究員	ストレージ技術動向調査	H19.4～H20.3
	藪田 浩二	日立製作所システム開発研究所	主任 研究員	ストレージ技術動向調査	H16.4～H19.3

§ 5 招聘した研究者等

無し。



## §6 成果発表等

(1)原著論文発表 (国内(和文)誌 42件、国際(欧文)誌 4件)

1. 花井知広, 渡邊明嗣, 山口宗慶, 田口 亮, 林 直人, 上原年博, 横田治夫, 「半導体ディスクによる自律ディスククラスタの階層化構成」、DBSJ Letters, Vol. 2, No. 3, pp. 41-44, 2003 年.
2. 小林大, 渡邊明嗣, 山口宗慶, 田口 亮, 上原年博, 横田治夫, 「複製データを併用した効率的なデータマイグレーションの検討」, DBSJ Letters, Vol. 3, No. 2, pp. 65-68, 2004 年.
3. 吉田 誠, 小林隆志, 横田 治夫, 「公開されている論文 DB からのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較」, 情報処理学会論文誌データベース, Vol. 45, No. SIG 7(TOD 22), pp. 24-32, 2004 年.
4. 小林 一仁, 横田 治夫, 「挿入を考慮した XML ラベリング手法の比較」, DBSJ Letters, Vol. 3, No. 1, pp. 85-88, 2004 年.
5. 吉田 誠, 小林隆志, 横田 治夫, 「公開されている論文 DB からのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較」, 情報処理学会論文誌データベース, Vol. 45, No. SIG 7(TOD 22), pp. 24-32, 2004 年.
6. 花井 知広, 渡邊 明嗣, 小林 大, 山口 宗慶, 田口 亮, 林 直人, 上原年博, 横田治夫, 「自律ディスククラスタの階層化構成におけるリクエスト転送先推測による性能改善」 DBSJ Letters, Vol. 3, No. 1, pp. 25-28, 2004 年.
7. 戸田誠二, 石田学, 横田治夫, 「連続メディアデータに対するディスクリタイ手法の評価」, 電子情報通信学会論文誌(D-I), Vol. J87-D-I, No. 7, pp. 769-781, 2004 年.
8. 小林大, 渡邊明嗣, 山口宗慶, 田口 亮, 上原年博, 横田治夫, 「複製データを併用した効率的なデータマイグレーションの検討」, DBSJ Letters, Vol. 3, No. 2, pp. 65-68, 2004 年.
9. 小林 大, 渡邊 明嗣, 上原年博, 横田 治夫, 「アベイラビリティの高い分散ストレージ管理ソフトウェアの更新手法」, 電子情報通信学会, 電子情報通信学会論文誌 (D-I), Vol. J-88-D-I, No. 3, pp. 684-697, 2005 年.
10. 小林隆志, 村木太一, 直井 聡, 横田治夫, 「統合プレゼンテーションコンテンツ蓄積検索システムの試作」, 電子情報通信学会, 電子情報通信学会論文誌 (D) , Vol. J88-D-I, No. 3, pp. 715-726, 2005 年.
11. 吉原朋宏, 小林大, 田口亮, 上原年博, 横田治夫, 「並列 Btree 構造における SMO 発生時の処理性能を改善する並行性制御」, DBSJ Letters, Vol. 4, No. 1, pp. 21-24, 2005 年.
12. 吉田誠, 小林隆志, 横田治夫, 「リサーチマイニング手法を用いた研究の発展経緯確認ツールの実装」, DBSJ Letters, Vol. 4, No. 1, pp. 81-84, 2005 年.
13. 中野真那, 小林大, 渡邊明嗣, 上原年博, 田口亮, 横田治夫, 「バージョン管理用差分情報のアクセス頻度に着目したデータ配置」, DBSJ Letters, Vol. 4, No. 1, pp. 121-124, 2005 年.
14. 小林大, 渡邊明嗣, 田口亮, 上原年博, 横田治夫, 「データ移動コストとキャッシュを考慮した複製へのアクセス分散制御」, DBSJ Letters, Vol. 4, No. 1, pp. 125-128, 2005 年.
15. 加藤英之, 小林隆志, 横田治夫, 「Web サービスを用いたワークフローにおける負荷分散手法のシミュレーションによる評価」, 日本データベース学会 DBSJ Letters, Vol. 4, No. 2, pp. 25-28, 2005 年.
16. 渡邊明嗣, 横田治夫, 「値域分割に基づく分散ストレージにおける効率向上のための複製管理」, DBSJ Letters, Vol. 4, No. 2, pp. 69-72, 2005 年.
17. Wenxin Liang and Haruo Yokota, Subtree-based XML Data Integration Using Leaf-clustering Based Approximate XML Join Algorithms DBSJ Letters, Vol. 4, No. 4, pp. 21-24, 2006.

18. Wenxin Liang and Haruo Yokota, "SLAX: An Improved Leaf-Clustering Based Approximate XML Join Algorithm for Integrating XML Data at Subtree Classes", 情報処理学会論文誌データベース, Vol. 47, No.SIG 8 (TOD 30), pp.47-57, 2006 年.
19. 太田健介, 小林大, 小林隆志, 田口亮, 横田治夫, 「柔軟なコンテンツ管理のためのルール処理への弁別ネットワークの適用」, DBSJ Letters, Vol.5, No.1, pp.1-4, 2006 年.
20. 中野真那, 小林大, 渡邊明嗣, 横田治夫, 「バージョン管理を行う分散ストレージにおける偏り監視範囲分割の影響」, DBSJ Letters, Vol.5, No.1, pp.105-108, 2006 年.
21. 加藤 英之, 小林 隆志, 横田 治夫, 「Web サービスベースのワークフロー管理における信頼性と負荷を考慮したスケジューリングパラメタ調整法」, DBSJ, DBSJ Letters, Vol. 5, No. 1, pp. 161-164, 2006 年.
22. 仲野 亘, 小林 隆志, 勝山 裕, 直井 聡, 横田 治夫, 「講演シーン検索における検索語出現状況に基づくレーザーポインタ情報のフィルタリング」, DBSJ, DBSJ Letters, Vol. 5, No. 2, pp. 1-4, 2006 年.
23. 小林大, 田口亮, 横田治夫, 「ストレージ複製管理のためのアクセス履歴とデータライフサイクル情報利用」, DBSJ Letters, Vol. 5, No. 2, pp. 33-36, 2006 年.
24. 城戸健太郎, 天笠俊之, 北川博之, 「PC クラスタを用いた XML データ並列処理方式の評価」, DBSJ Letters, Vol. 5, No. 2, pp. 85-88, 2006 年.
25. 山元理絵, 小林大, 吉原朋宏, 小林隆志, 横田治夫, 「Web ページ推薦における推薦順位決定のための得点付け手法の比較」, DBSJ Letters, Vol. 5, No. 4, pp. 5-8, 2007 年.
26. 中野真那, 小林大, 渡邊明嗣, 横田治夫, 「COBALT:バージョン管理を行う並列分散ストレージシステムにおけるアクセス負荷と記憶空間利用率の同時均衡化手法」, 電子情報通信学会, 電子情報通信学会和文論文誌 (D), Vol. J90-D, No. 2, pp. 349-358, 2007 年.
27. 岡本拓明, 仲野亘, 小林隆志, 直井聡, 横田治夫, 岩野公司, 古井貞熙, 「音声情報を統合したプレゼンテーションコンテンツ検索」, 電子情報通信学会, 電子情報通信学会和文論文誌, Vol. J90-D, No. 2, pp. 209-222, 2007 年.
28. 山元理絵, 吉原朋宏, 小林大, 小林隆志, 横田治夫, 「アクセスログに基づく Web ページ推薦における LCS の利用とその解析」, 情報処理学会論文誌データベース, Vol. 48, No. SIG 11 (TOD 34), pp. 38-48, 2007 年.
29. 油井誠, 宮崎純, 植村俊亮, 「効率的な XQuery 処理のための DTM に基づく XML ストレージ」, 情報処理学会論文誌: データベース, Vol. 48, No. SIG 11 (TOD34), pp. 128-148, 2007 年.
30. 武田裕子, 小口正人, 「VPN 上の iSCSI 環境における複数経路アクセス適応制御手法の提案と評価」, 日本データベース学会 Letters, Vol. 6, No. 1, pp. 125-128, 2007 年.
31. Tomohiro Yoshihara, Dai Kobayashi, Haruo Yokota, "MARK-OPT: A Coucurrency Control Protocol for Parallel B-Tree Structures to Reduce the COST of SMOs", IEICE Transactions on Information and Systems, Vol.E90-D, No. 8, pp. 1213-1224, 2007.
32. 並木悠太, 神戸康多, 小林大, 横田治夫, 「Fat-Btree をインデックスに用いた PostgreSQL の分散検索」, DBSJ Letters, Vol. 6, No. 2, pp. 61-64, 2007 年.
33. 渡部 徹太郎, 小林 隆志, 横田 治夫, 「ファイル検索に向けたアクセスログからのファイル間関連度の導出」, DBSJ Letters, Vol. 6, No. 2, pp. 65-68, 2007 年.
34. Reyn Nakamoto, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura: "Tag-Based Contextual Collaborative Filtering", IAENG International Journal of Computer Science, Volume 34, Issue 2, pp. 214-219, 2007.
35. 小林大, 横田治夫, 「並列ストレージにおけるデータ再配置による長期的負荷均衡化と短期的応答性能の両立」, 情報処理学会論文誌データベース, Vol. 49, No.

- SIG15 (TOD37), 2008 年.
36. 仲野亘, 小林隆志, 直井聡, 横田治夫, 「講義講演シーン検索におけるレーザーポインタ情報の活用法」, 電子情報通信学会論文誌(D), Vol. J91-D, No. 3, pp. 654-666, 2008 年.
  37. 加藤英之, 小林隆志, 横田治夫, 「OXTHAS: Web サービスベースのワークフロー管理における障害を考慮した負荷分散手法」, 電子情報通信学会和文論文誌(D), Vol. J91D, No. 4, pp. 993-1003, 2008 年.
  38. Yang-Yang Wu, Qing Lei, Duan-Sheng Chen, Haruo Yokota, "A Method of Discovering Relation Information from XML Data", Journal of Software, Journal of Software, Vol. 19, No. 6, pp. 1422-1427, 2008.
  39. 吉野悠二, 梁文新, 横田治夫, 「部分木検索のための走査と結果構築コストを考慮した XML 分割配置と位置情報取得手法」, 日本データベース学会論文誌, Vol. 7, No. 1, pp. 115-120, 2008 年.
  40. 千島望, 山口実靖, 小口正人: 「VPN 複数経路接続を用いた iSCSI 並列ストレージアクセス解析」 Journal of the DBSJ, Vol. 7, No. 1, pp. 127-132, 2008 年.
  41. Imam Machdi, Toshiyuki Amagasa, and Hiroyuki Kitagawa: "Cube-Based Analysis for Maintaining XML Data Partition for Holistic Twig Joins", Journal of the DBSJ, Vol. 7, No. 1, pp. 121-126, 2008 年.
  42. 小林大, 横田治夫, 「負荷変動と応答性能維持を考慮した高可用並列ストレージシステムのための複製利用と更新要求制御」, 情報処理学会, 情報処理学会論文誌, Vol. 49, No. 6, pp. 2054-2069, 2008 年.
  43. 武部浩明, 小澤憲秋, 勝山裕, 横田治夫, 直井聡, 「文字認識技術を利用した講義動画のスライド同定」, 電子情報通信学会, 電子情報通信学会論文誌(D), Vol. J91-D, No. 9, pp. 2280-2292, 2008 年.
  44. 高山一樹, 横田治夫, 「平文を生成しない分散ストレージ上での再暗号化手法」, 日本データベース学会, 日本データベース学会論文誌, Vol. 7, No. 3, pp. 43-48, 2008. 12.
  45. Neila Ben Lakhal, Takashi Kobayashi, Haruo Yokota, "FENECIA: Failure Endurable Nested-transaction based Execution of Composite Web Services with Incorporated State Analysis", The VLDB Journal, Springer Berlin/Heidelberg, Vol. 18, No. 1, pp. 1-56, 2009.1.
  46. 千島望, 山口実靖, 小口正人: 「VPN 複数経路接続時における iSCSI ストレージアクセスの性能評価」, 情報処理学会論文誌採録

(2)学会発表(国際学会発表及び主要な国内学会発表)

- ① 招待講演 (国内会議 1 件、国際会議 4 件)
1. Haruo Yokota, "Efficient Dependable Management for a Large Amount of Stored Data", The 2<sup>nd</sup> Korea-Japan Database Workshop (KJDB2005), 2005.10
  2. Toshiyuki Amagasa, Kentarou Kido, and Hiroyuki Kitagawa, "Querying XML Data using PC Cluster System", 2nd International Workshop on XML Data Management Tools and Techniques (XANTEC'07) in conjunction with DEXA 2007, pp. 5-9, Regensburg, Germany, 2007
  3. Haruo Yokota, "Approaches to Cope with the Local Information Explosion", The 4th Korea-Japan Database Workshop (KJDB2008), 2008.10
  4. 横田治夫, 「データ工学とセキュリティ」, コンピュータセキュリティシンポジウム 2008, 2008.10.
  5. Haruo Yokota, "Management and Search Methods for the Local Information Explosion", National Database Conference (NDBC2008), 2008.10.

② 口頭発表 (国内会議 102 件、国際会議 69 件)

1. Jun Miyazaki, Yohei Abe, and Haruo Yokota, "Availabilities and Costs of Reliable Fat-Btrees", in Proc. of 2004 IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2004), pp.163-172, 2004.3.6.
2. Haruo Yokota, "An Information Storage System for Large-Scale Knowledge Resources" in Proc. of International Symposium on Large-scale Knowledge Resources (LKR2004), Tokyo, pp.87-90, 2004.3.9.
3. Neila Benlakhhal, Takashi Kobayashi, and Haruo Yokota, "THROWS: An Architecture for Highly Available Distributed Execution of Web Services Compositions", in Proc. of the 14th IEEE International Workshop on Research Issues on Data Engineering (RIDE2004): Web Services for E-Commerce and E-Government Applications, Boston, pp.103-110, 2004.3.29.
4. Akitsugu Watanabe, and Haruo Yokota, "AMiGO : An Adaptive Data Migration Strategy to Integrate System Resizing And Load Balancing", 2nd International Conference on Computer Science and its Applications (ICCSA-2004), pp.333-340, 2004.
5. Haruo Yokota, "Research on the Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.41-62, 2004.
6. Akitsugu Watanabe, "Load Balancing for Distributed Storage", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.63-78, 2004.
7. Dai Kobayashi, "Efficient Data Migration in Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.79-94, 2004.
8. Takashi Kobayashi, "Large Scaled Simulation System for Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.95-102, 2004.
9. Katsunori Yamaoka, "Expanding Autonomous Disks to the Internet, Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.131-136, 2004.
10. Aki Kobayashi "Traffic Aspect of Autonomous Disk over the Internet", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.137-142, 2004.
11. Shunsuke Uemura, "Media Oriented Applications for Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.103-104, 2004.
12. Jun Miyazaki, "Data Management for Widely Distributed Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.105-111, 2004
13. Toshiyuki Amagasa, "Towards Processing Large-scale XML Data using Autonomous Disks", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.113-124, 2004
14. Mitsuhiko Ohta, "Development of an Autonomous Disk Prototype", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.181-188, 2004
15. Arata Ejiri, "Development of a Distributed Autonomous Storage System", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.189-204, 2004.
16. Koji Sonoda, "Research on File System for Streaming Media Data", Proceedings of the 1st International Workshop on Advanced Storage System (ADSS 2004), pp.125-130,



- 2004.
17. Akitsugu Watanabe, and Haruo Yokota, "Adaptive Overlapped Declustering: A Highly Available Data-Placement Method Balancing Access Load and Space Utilization", in Proc. of 21st International Conference on Data Engineering (ICDE2005), IEEE, pp.828-839, 2005.4
  18. Neila Ben Lakhal, Takashi Kobayashi, and Haruo Yokota, "Reliability and Performance Estimation for Enriched WS-SAGAS", in Proc. of International Workshop on Challenges in Web Information Retrieval and Integration (WIRI2005), pp.55-64, 2005.4
  19. Mana Nakano, Dai Kobayashi, Akitsugu Watanabe, Toshihiro Uehara, Ryo Taguchi (NHK), and Haruo Yokota, "The Versioning System Balancing Data Amount and Access Frequency on Distributed Storage System" in Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2005), pp. 184-187, 2005.4
  20. Dai Kobayashi, Akitsugu Watanabe, Ryo Taguchi, Toshihiro Uehara (NHK), and Haruo Yokota, "An Efficient Access Forwarding Method Based On Caches On Storage Nodes", in Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2005), pp.188-191, 2005.4
  21. Tsutomu Fujiwara, Jun Miyazaki, and Shunsuke Uemura, "Data Migration for a Widely Distributed Storage System Using Autonomous Disks", Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2005) , pp.196-199, IEEE CS, 2005.4.
  22. Wenxin Liang, and Haruo Yokota, "LAX: An Efficient Approximate XML Join Based on Clustered Leaf Nodes for XML Data Integration", in Proc. of 22nd British National Conferences on Databases (BNCOD22), LNCS 3567, Springer, pp.82-97, 2005.7
  23. Wenxin Liang, and Haruo Yokota, "SLAX: An Improved Leaf-Clustering Based Approximate XML Join Algorithm for XML Data Integration at Subtree Classes", in Proc. of IPSJ DBWeb2005, pp.41-48, 2005.11
  24. Neila Ben Lakhal, Takashi Kobayashi, and Haruo Yokota, "A Failure-Aware Model for Estimating and Analyzing the Efficiency of Web Services Compositions", in Proc. of IEEE 11th Pacific Rim International Symposium on Dependable Computing (PRDC 2005), pp.114-121, 2005.12.
  25. Haruo Yokota, "Overview and Progress of the CREST Advanced Storage System Project", ADSS2005, 2005.12.
  26. Dai Kobayashi and Haruo Yokota, "Logical data migration for online software updates in Autonomous Disks", ADSS2005, 2005.12.
  27. Xiangyong Ouyang and Haruo Yokota, "Neighbor WAL protocol in Autonomous Disk System", ADSS2005, 2005.12.
  28. Tomohiro Yoshihara, Dai Kobayashi, Akitsugu Watanabe, Haruo Yokota, "A Concurrency Control Method Marking SMO Occurrence in a Parallel Btree Structure", ADSS2005, 2005.12.
  29. Mana Nakano, Dai Kobayashi, Akitsugu Watanabe, Toshihiro Uehara, Ryo Taguchi, Haruo Yokota, "COBALT: Combination of Btree and Linked-list Transfer to Balance Access Loads and Amount of Distributed Objects with Version Management", ADSS2005, 2005.12.
  30. Mitsuhiko Ohta, "Feasibility study of an Autonomous Disk Prototype", ADSS2005, 2005.12.
  31. Arata Ejiri, "Self-supervisor, diagnosis, restoration technology of HDD", ADSS2005, 2005.12.
  32. Seiji Toda, "Distributed Storage System Control using Power and Temperature

- Information”, ADSS2005, 2005.12.
33. Shunsuke Uemura, “Progress report: Management of diverse media using network of autonomous disks”, ADSS2005, pp.171–174, 2005.12.
  34. Jun Miyazaki, “Main Memory Storage Supporting Query Processing ”, ADSS2005, pp.175–185, 2005.12.
  35. Toshiyuki Amagasa, Kentaro Kido, Hiroyuki Kitagawa, “Parallel processing of XML data using autonomous disks”, ADSS2005, pp.187–196, 2005.12.
  36. Tsutomu Fujiwara, Jun Miyazaki, Shunsuke Uemura, “Simulation and Performance Evaluation for a Widely Distributed Storage System Using Autonomous Disks”, pp.197–203, ADSS2005, 2005.12.
  37. Koji Sonoda, “A Survey of Distributed file system”, ADSS2005, 2005.12.
  38. Takashi Kobayashi, Haruo Yokota, “Knowledge Store: Large-scale Knowledge Resources Storing System”, ADSS2005, 2005.12.
  39. Seiji Toda, Manabu Ishida, and Haruo Yokota, “Evaluation of Disk-Retry Methods for Continues Media Data”, *Systems and Computers in Japan*, Wiley, Vol.37, Issue.1, pp.47–60, 2006.1.
  40. Wenxin Liang and Haruo Yokota, “A Path-sequence Based Discrimination for Subtree Matching in Approximate XML Joins”, in *Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2006)* , 2006.4
  41. Tomohiro Yoshihara, Dai Kobayashi, Ryo Taguchi and Haruo Yokota, “A Concurrency Control Method for Parallel Btree Structures”, in *Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2006)*, 2006.4
  42. Kensuke Ota, Dai Kobayashi, Takashi Kobayashi, Ryo Taguchi and Haruo Yokota, “Treatment of Rules in Individual Metadata of Flexible Contents Management”, in *Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2006)*, 2006.4
  43. Makoto Kataigi, Dai Kobayashi, Tomohiro Yoshihara, Takashi Kobayashi, Ryo Taguchi and Haruo Yokota, “Evaluation of Placement and Access Assignment for Replicated Object Striping”, in *Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2006)*, 2006.4.
  44. Kentarou Kido, Toshiyuki Amagasa, and Hiroyuki Kitagawa, “Processing XPath Queries in PC Clusters Using XML Data Partitioning”, *The Second International Special Workshop on Databases for Next-Generation Researchers (SWOD 2006) in conjunction with ICDE 2006*, pp. 114–119, 2006.4.
  45. Jun Miyazaki, “A Memory Subsystem with Comparator Arrays for Main Memory Database Operations”, *Proc. of the 21st Annual ACM Symposium on Applied Computing (SAC 2006), Special Track on Database Theory, Technology and Applications (DTTA)*, 2006.4.
  46. Hideyuki Katoh, Takashi Kobayashi and Haruo Yokota, “Activity Scheduling in Web-Service Based Workflow Management for Balancing Load and Handling Failures”, in *Proc of 2006 International Workshop on Future Mobile and Ubiquitous Information Technologies (FMUIT’06)*, pp.245–248, 2006.5,
  47. Neila Benlakhel, Takashi Kobayashi, and Haruo Yokota, “Dependability and Flexibility Centered Approach for Composite Web Services Modeling”, in *Proc. of 14th International Conference on Cooperative Information Systems (CoopIS 2006) in OTM2006, LNCS 4275*, pp.163–182, 2006.11.
  48. Xiangyong Ouyang, Tomohiro Yoshihara, Haruo Yokota, “An Efficient Commit Protocol Exploiting Primary-Backup Placement in a Distributed Storage System”, *Proc. of IEEE 12th Pacific Rim International Symposium on Dependable Computing*

- (PRDC2006), IEEE, pp. 238–247, 2006.12.
49. Wataru Nakano, Takashi Kobayashi, Yutaka Katsuyama, Satoshi Naoi Haruo Yokota, "Treatment of Laser Pointer and Speech Information in Lecture Scene Retrieval", Proc. of IEEE International Symposium on Multimedia (ISM2006), IEEE, pp. 927–932, 2006.12.
  50. Takashi Kobayashi, Wataru Nakano, Haruo Yokota, Koichi Shinoda, Sadaoki Furui, "Presentation Scene Retrieval Exploiting Features in Videos Including Pointing and Speech Information", Proceedings of LKR2007, pp. 95–100, 2007.3.
  51. Reyn Nakamoto, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura, "Contextual Collaborative Filtering Based on Tagging Information", The International MultiConference of Engineers and Computer Scientists 2007 (IMECS 2007), 2007.3.
  52. Dai Kobayashi, Ryo Taguchi, Haruo Yokota, "An Experimental Evaluation of the Adaptive Replica-assisted Migration for Parallel Storage Systems", in Proc. of International Special Workshop on Databases For Next Generation Researchers (SWOD 2007), in conjunction with ICDE 2007, 2007.4
  53. Hiroto Kurita, Kenji Hatano, Jun Miyazaki, Shunsuke Uemura: "Efficient Query Processing for Large XML Data in Distributed Environments", Proc. of IEEE 21st International Conference on Advanced Information Networking and Applications (AINA-07), pp.317–322, 2007.5.
  54. Yangyang Wu, Qing Lei, Wei Luo, Haruo Yokota, "Discovering Relations Among Entities from XML Documents", Springer LNAI, Springer, No. 4571, pp. 899–910, 2007.7.
  55. Dai Kobayashi and Haruo Yokota, "Comparison of Replica Assisting and Speed Adjustment for Service-aware Horizontal Data Migration on Autonomous Disks", in Proc. of ADSS 2007 in conjunction with ICDIM 2007, pp. 539–544, 2007.10.
  56. Kazuki Takayama, Dai Kobayashi and Haruo Yokota, "Consideration of Experimental Evaluation about Encrypted Replica Update Process", in Proc. of ADSS 2007 in conjunction with ICDIM 2007, pp.545–550, 2007.10.
  57. Wenxin Liang, Xiangyong Ouyang and Haruo Yokota, "An XML Subtree Segmentation Method Based on Syntactic Segmentation Rate", in Proc. of ADSS 2007 in conjunction with ICDIM 2007, pp.551–558, 2007.10
  58. Makoto Yui, Jun Miyazaki, Shunsuke Uemura, Hirokazu Kato: "Efficient XML Storage based on DTM for Read-oriented Workloads", Proc of IEEE International Workshop on Advanced Storage Systems (ADSS 2007) in conjunction with IEEE ICDIM 2007, pp.559–564, IEEE CS Press, 2007.10.
  59. Nozomi Chishima, Saneyasu Yamaguchi, and Masato Oguchi: "Analysis of Multiple Target Network Storage Access using Multi-routing VPN Connection," In Proc. International Workshop on Advanced Storage Systems 2007 (ADSS2007) in conjunction with the 2nd IEEE International Conference on Digital Information Management (ICDIM2007), pp.526–531, 2007.10.
  60. Kenji Hatano, Toshiyuki Shimizu, Jun Miyazaki, Yu Suzuki, Hiroko Kinutani, Masatoshi Yoshikawa: "Ranking and Presenting Search Results in an RDB-based XML Search Engine", Pre-Proceedings of INEX 2007 Workshop, pp.156–169, 2007.12.
  61. Tomohiro Yoshihara, Dai Kobayashi, Haruo Yokota, "A Concurrency Control Protocol for Parallel B-tree Structures without Latch-Coupling for Explosively Growing Digital Content", Proc. of the 11th International Conference on Extending Database Technology (EDBT2008), pp. 133–144, 2008.3.
  62. Aiko Matsui, Kikuko Kamisaka, Saneyasu Yamaguchi, Masato Oguchi: "A Framework for Advanced Data Processing in a Multi-hop Network Environment," In Proc. IASTED

- International Conference on Internet and Multimedia Systems and Applications (EuroIMSA 2008), pp.150-155, 2008.3.
63. Makoto Yui, Jun Miyazaki, Shunsuke Uemura, Hirokazu Kato: "XBird/D: Distributed and Parallel XQuery Processing using Remote Proxy", Proc. of the 23rd Annual ACM Symposium on Applied Computing (SAC 2008), Special Track on Database Theory, Technology and Applications (DTTA), pp.1003-1007, 2008.3.
  64. Emi Arai, Yumi Hirano, Tutomu Murase, Masato Oguchi; "Empirical Study on a Session Layer Resource Control over 802.11 DCF Wireless LAN," In Proc. 7th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT2008), pp.151-156, 2008.4.
  65. Wenxin Liang, Haruo Yokota, "Exploiting Path Information for Syntax-based XML Subtree Matching in RDBs", Proc. of International Conference on Web-Age Information Management (WAIM2008), 2008.8.
  66. Do Ngoc Hung, Yutaka Katsuyama, Satoshi Naoi, Haruo Yokota, "Improvement of Telop Recognition Quality by Integrating Web Search Results", Proc. of VLDB2008 Workshops (NTII), pp. 32-35, 2008.8.
  67. Hieu Hanh Le, Thitiporn Lertrudachakul, Tetsutaro Watanabe, Haruo Yokota, "Automatic Digest Generation by Extracting Important Scenes from the Content of Presentations", Proc. of DEXA2008 Workshops (AIEMPro'08), IEEE Computer Society, pp. 590-594, 2008.9.
  68. Wenxin Liang, Takeshi Miki, Haruo Yokota, "Superimposed Code-Based Indexing Method for Extracting MCTs from XML Documents", Database and Expert Systems Applications of LNCS (Proc. of DEXA2008), Springer Berlin / Heidelberg, Vol. 5181, pp. 508-522, 2008.9.
  69. Imam Machdi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "GMX: An XML Data Partitioning Scheme for Holistic Twig Joins", The 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS2008), Linz, Austria, 2008.11.
  70. 小林大, 山口宗慶, 花井知広, 渡邊明嗣, 田口亮, 林直人, 上原年博, 横田治夫, 「仮想ノードを用いた自律ディスクシステム更新の高可用性」、信学技報, 電子情報通信学会, DE2003-109, DC2003-22, pp.39-44, 2003.10.8
  71. 渡邊明嗣, 横田治夫, 「値域分割された分散ストレージにおける効率的なアクセス負荷の記録と管理」、信学技報, 電子情報通信学会, DE2003-116, DC2003-29, pp.31-36, 2003.10.9
  72. 上原年博, "Vision for Future Broadcasting Storage Systems", International Symposium on Ultra-High Density Spinic Storage Systems (ISSS), FFA-02, 仙台, 2004.10.31
  73. 上原年博, 「将来の放送用ストレージシステムの展望」、電情通学技術報告, vol.103, no.555, MR2003-45, pp.1-6, 大阪, 2004.1.16
  74. 小原 俊樹, 渡辺 明嗣, 花井 知広, 山口 宗慶, 小林 大, 上原 年博, 林 直人, 田口 亮, 横田 治夫, 「並列分散ストレージにおけるフラットなデータ管理への階層構造の導入手法」、第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, 4-A-1, 2004.3.5.
  75. 花井 知広, 渡邊 明嗣, 小林 大, 山口 宗慶, 田口 亮, 林 直人, 上原 年博, 横田 治夫, 「自律ディスククラスタの階層化構成における性能改善と評価」、第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, 4-A-2, 2004.3.5.
  76. 小林 一仁, 横田 治夫, 「挿入制限のないコード VLEI を用いた XML ラベリング手法の検討とその評価」、第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, 7-A-1, 2004.3.6.



77. Neila Ben Lakhil, Takashi Kobayashi, Haruo Yokota, “A Simulation System of THROWS Architecture for WS-SAGAS”, 第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, 7-B-4, 2004.3.6.
78. 戸田 誠二, 横田 治夫, 「LCS を用いたアクセスログ解析の並列処理による性能向上」, 第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, 7-B-5, 2004.3.6.
79. 加藤 英之, 小林 隆志, 横田 治夫, 「ワークフローの自動実行における障害対策」, 第 15 回電子情報通信学会データ工学ワークショップ論文集, DEWS2004, I-12-2, 2004.3.6.
80. 羅勇, 天笠俊之, 波多野賢治, 宮崎純, 植村俊亮, “移動オブジェクトのクラスタリング手法に関する一提案”, 情報処理学会研究会報告 FI-74, 2004.3.26.
81. 小林隆志, 鈴木泰山, 横田治夫, 「大規模知識資源蓄積活用システムの構成」, 信学技報, 電子情報通信学会, DE2004-9, pp.49-54, 2004.
82. 山口宗慶, 渡辺明嗣, 小林大, 田口 亮, 林 直人, 上原年博, 横田治夫, 「分散ストレージにおける情報ライフサイクルの効率的な管理」, 夏のデータベースワークショップ DBWS2004 7C, 電子情報通信学会, 信学技報, DE2004-78 (2004-07), pp.177-182, 情報処理学会, 研究報告, 2004-DBS-134 (II) (69), pp.505-512, 2004.
83. 渡辺明嗣, 山口宗慶, 小林大, 田口 亮, 林 直人, 上原年博, 横田治夫, 「分散データ格納環境のための自律的並列偏り除去手法」, 夏のデータベースワークショップ DBWS2004 7C, 電子情報通信学会, 信学技報, DE2004-79 (2004-07), pp.183-188, 情報処理学会, 研究報告, 2004-DBS-134 (II) (70), pp.513-520, 2004.
84. 小林大, 渡邊明嗣, 山口宗慶, 田口亮, 林直人, 上原年博, 横田治夫, 「分散ストレージ上の複製へのアクセス要求配分を取り入れた負荷均衡化手法」, 夏のデータベースワークショップ DBWS2004 7C, 電子情報通信学会, 信学技報, DE2004-80, (2004-07), pp.189-194, 情報処理学会, 研究報告, 2004-DBS-134 (II) (71), pp.521-528, 2004.7
85. 小林大, 渡邊明嗣, 田口亮, 上原年博, 横田治夫, 「負荷分散のためのデータ移動による性能低下を抑制するアクセス回送制御」, 電子情報通信学会, 信学技報, DE2004-111, pp.35-40, 2004.
86. 渡辺明嗣, 小林大, 山口宗慶, 田口 亮, 上原年博 横田治夫, 「偏り除去手続き並列化の分散ストレージにもたらす効果」, 電子情報通信学会, 信学技報, DE2004-112, pp.41-46, 2004.
87. 吉原朋宏, 小林大, 田口亮, 上原年博, 横田治夫, 「並列ディレクトリ構造 Fat-Btree の並行性制御の改善とその評価」, 第 16 回電子情報通信学会データ工学ワークショップ(DEWS2005) 論文集, 2A-o4, 2005.
88. 中野真那, 小林大, 渡邊明嗣, 上原年博, 田口亮, 横田治夫, 「分散環境でのファイル版管理のためのアクセス頻度を考慮したデータ配置法」, 第 16 回電子情報通信学会データ工学ワークショップ(DEWS2005) 論文集, 5B-i5, 2005.
89. 小林大, 渡邊明嗣, 田口亮, 上原年博, 横田治夫, 「負荷変動傾向を考慮したストレージ間データ移動と複製間アクセス回送の協調制御」, 第 16 回電子情報通信学会データ工学ワークショップ(DEWS2005) 論文集, 5B-i7, 2005.
90. 山口宗慶, 渡邊明嗣, 小林大, 田口亮, 上原年博, 横田治夫, 「ルールを含むメタデータによる柔軟なコンテンツ管理」, 第 16 回電子情報通信学会データ工学ワークショップ(DEWS2005) 論文集, 6A-i8, 2005.
91. 長良香子, 小林一仁, 小林大, 横田治夫, 「XML データベースへのラベル付け手法 VLEI の評価」, 第 16 回電子情報通信学会データ工学ワークショップ(DEWS2005) 論文集, 6B-o3, 2005.
92. 田口亮, 上原年博, 藤田欣裕, 山口宗慶, 小林大, 渡邊明嗣, 横田治夫, 「メタデータによるコンテンツの蓄積, 転送管理の検討」, 電子情報通信学会総合大会講演論文

- 集 D-4-11、大阪大学、平成 17 年 3 月 22 日
93. 太田光彦, “自律ディスク試作機の開発”, 磁気記録研究会(電子情報通信学会), 2005/2/10 東京
  94. 加藤英之, 小林隆志, 横田治夫, 「Web サービスを用いたワークフローにおける負荷分散手法」, 夏のデータベースワークショップ DBWS2005, 電子情報通信学会、信学技報、DE2005-46 (2005-7), pp.95-100, 情報処理学会、研究報告 2005-DBS-137 (I) (17), pp.121-127, 2005.7.
  95. 吉原朋宏, 渡邊明嗣, 小林大, 上原年博, 横田治夫, 「並列 Btree 構造における負荷分散処理の並行性制御への影響」, 夏のデータベースワークショップ DBWS2005, 電子情報通信学会、信学技報、DE2005-61 (2005-7), pp.185-190, 情報処理学会、研究報告 2005-DBS-137 (I) (32), pp.237-243, 2005.7.
  96. 渡邊明嗣, 横田治夫, 「Adaptive Overwrapped Declustering: アクセス負荷分散と容量効率向上のための複製管理戦略」, 夏のデータベースワークショップ DBWS2005, 電子情報通信学会、信学技報、DE2005-98 (2005-7), pp.193-198, 情報処理学会、研究報告 2005-DBS-137 (II) (69), pp.515-522, 2005.7.
  97. 藤原勤, 宮崎純, 植村俊亮, 「自律ディスクによる広域分散ストレージの静的な性能解析」, 情報処理学会研究報告/電子情報通信学会技術研究報告, Vol.2005, No.68, 2005-DBS-137(II)-75, pp.561-568/Vol.105, No.172, DE2005-104, pp.227-232, 情報処理学会/電子情報通信学会, 2005.07.
  98. 戸田誠二, 太田, 江尻, 「オーガニックストレージシステムの大規模ハードウェアへの実装」, 電子情報通信学会技術研究報告, Vol.105No.226 P7-12, 2005.8.
  99. 中野真那, 小林大, 渡邊明嗣, 上原年博, 田口亮, 横田治夫, 「アクセス頻度と容量分散を考慮した版管理用データ配置法の実装と評価」, 信学技報, 電子情報通信学会, DE2005-130, DC2005-24, pp.31-36, 2005.10
  100. 太田健介, 小林大, 小林隆志, 田口亮, 横田治夫, 「柔軟なコンテンツ管理に向けた個々のメタデータ中のルール処理の効率化」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 1A-o1, 2006.3.
  101. 中野真那, 小林大, 渡邊明嗣, 横田治夫, 「バージョン管理を前提に負荷・容量均衡化を両立させる分散配置の応答性能への影響」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 4C-o2, 2006.3.
  102. 村上翔一, 小林大, 横田治夫, 「DO-VLEI を用いた XML 格納におけるラベルサイズと問い合わせ性能」, 第 17 回電子情報通信学会データ工学ワークショップ (DEWS2006) 論文集, 7B-o2, 2006.3.
  103. 加藤英之, 小林隆志, 横田治夫, 「Web サービスベースのワークフロー管理における障害を考慮したアクティビティスケジューリング手法」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 7C-o1, 2006.3.
  104. 吉原朋宏, 小林大, 田口亮, 横田治夫, 「並列 Btree 構造 Fat-Btree におけるリクエスト委譲コストを削減する並行性制御手法」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 7C-o3, 2006.3.
  105. 片居木誠, 小林大, 吉原朋宏, 小林隆志, 田口亮, 横田治夫, 「分散ストレージシステム上での複製を含むデータの分割配置とアクセススケジューリング」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 7C-o5, 2006.3.
  106. 油井誠, 宮崎純, 植村俊亮, 「XML コンテンツ統合を支援する XQuery プロセッサ」, 電子情報通信学会 第 17 回データ工学ワークショップ (DEWS2006) 論文集, 1D-d2, 2006.3.
  107. 城戸健太郎, 天笠俊之, 北川博之, “PC クラスタを用いた XML データ並列処理方式の提案”, 電子情報通信学会第 17 回データ工学ワークショップ (DEWS2006)論文集, 6B-oi3, 2006.3.
  108. 栗田裕人, 宮崎純, 波多野賢治, 中島伸介, 植村俊亮, “大規模 XML データに対

- する分散問合せ処理の効率化”, 情報処理学会研究報告, Vol.2006, No.33, 2006-FI-82/2006-DD-54, pp.23-30, 情報処理学会, 2006.3.
109. 武田 裕子, 小口 正人: 「IP ストレージリモートアクセスにおける VPN 利用に関する一検討」, マルチメディア, 分散, 協調とモバイル (DICOMO 2006) シンポジウム 論文集(II), pp.713-716, 2006.7.
  110. 千島 望, 豊田 真智子, 山口 実靖, 小口 正人: 「VPN 環境における iSCSI ストレージアクセス時の TCP パラメータの動作解析」, マルチメディア, 分散, 協調とモバイル (DICOMO 2006) シンポジウム 論文集(II), pp.709-712, 2006.7.
  111. 三木健士, 横田治夫, 「スーパーインポーズドコーディングを用いたXML文書キーワード索引手法」, 夏のデータベースワークショップ DBWS2006, 電子情報通信学会, ISSN 0913-5685 信学技報 vol.106, No.148, DE2006-46(2006-07), pp.145-150, 2006.7.
  112. 小林大, 田口亮, 横田治夫, 「利用履歴とメタデータに基づくストレージ管理のためのデータライフサイクル推定」, 夏のデータベースワークショップ DBWS2006, 電子情報通信学会, ISSN 0913-5685 信学技報 vol.106, No.149, DE2006-59(2006-07), pp.53-58, 2006.7.
  113. 吉原朋宏, 小林大, 田口亮, 横田治夫, 「Fat-Btree における B-link を用いた並行性制御手法」, 夏のデータベースワークショップ DBWS2006, 電子情報通信学会, ISSN 0913-5685 信学技報 vol.106, No.150, DE2006-110(2006-07), pp.109-114, 2006.7.
  114. 油井誠, 宮崎純, 植村俊亮, 「効率的な XQuery 処理のための DTM に基づく XML ストレージ」, 情報処理学会研究報告/電子情報通信学会技術研究報告, Vol.2006, No.77, 2006-DBS-140(I), pp.87-94/Vol.106, No.148, DE2006-34, pp.73-78, 情報処理学会/電子情報通信学会, 2006年07月.
  115. 藤原勤, 宮崎純, 植村俊亮, 「自律ディスクによる広域分散ストレージシステムの通信およびストレージのオーバーヘッドを考慮した性能評価」, 情報処理学会研究報告/電子情報通信学会技術研究報告, Vol.2006, No.78, 2006-DBS-140(II), pp.99-106/Vol.106, No.149, DE2006-63, pp.77-82, 情報処理学会/電子情報通信学会, 2006.7.
  116. 千島 望, 豊田 真智子, 山口 実靖, 小口 正人: 「iSCSI アクセス時の VPN 環境における TCP 輻輳ウィンドウ制御手法の検討」, 夏のデータベースワークショップ (DBWS 2006), 情報処理学会研究報告, 2006-DBS-140(II), pp. 83-90, 2006.7.
  117. 千島 望, 豊田 真智子, 山口 実靖, 小口 正人: 「VPN 接続環境における TCP パラメータと通信性能の相関関係評価」, 第 5 回情報科学技術フォーラム (FIT 2006), L-042, 2006.9.
  118. 松井 愛子, 豊田 真智子, 神坂 紀久子, 小口 正人: 「無線 LAN 通信時における TCP パラメータの振舞とパケット転送タイミングの解析」, 第 5 回情報科学技術フォーラム (FIT 2006), L-077, 2006.9.
  119. 小林大, 田口亮, 横田治夫, 「並列ストレージにおけるサービス性能を保った負荷均衡化の影響」, 電子情報通信学会, 信学技報, Vol. 106, No. DE-291, pp. 19-24, 2006.10.
  120. 江尻 革, 太田 光彦, 吉原 朋宏, 小林 大, 歐陽 湘勇, 田口亮, 横田 治夫, 「自律分散型ストレージシステムにおける性能および可用性改善の試み」, 電子情報通信学会, 信学技報, Vol. 106, No. DE-291, pp. 25-30, 2006.10.
  121. 山元理絵, 小林大, 吉原朋宏, 小林隆志, 横田治夫, 「アクセスログに基づく Web ページ推薦における LCS の利用とその解析」, 情報処理学会, Proc. of IPSJ DBWeb2006, pp. 43-50, 2006.11.
  122. 武田 裕子, 小口 正人: 「VPN を利用した IP ストレージへの複数経路アクセス制御手法の提案」, 電子情報通信学会コンピュータシステム研究会 (CPSY), CPSY2006-39,

- pp. 7-12, 2006.12.
123. 千島 望, 山口 実靖, 小口 正人: 「iSCSI ストレージにおける VPN 複数経路アクセス方式の検討」, 電子情報通信学会コンピュータシステム研究会 (CPSY), CPSY2006-38, pp. 1-6, 2006.12.
  124. 松井 愛子, 神坂 紀久子, 山口 実靖, 小口 正人: 「無線LANマルチホップ環境下における階層プロトコル上のパケットトレース」, 電子情報通信学会 ITS 研究会, ITS2006-39, pp. 25-30, 2006.12.
  125. 高山一樹, 小林大, 横田治夫, 「複製を利用したストレージ中での暗号化データの権限失効処理」, 電子情報通信学会, DEWS2007 論文集, pp. B9-8, 2007.3.
  126. 吉原 朋宏, 小林大, 田口亮, 横田治夫, 「並列 B-tree 構造における負荷分散処理時の並行性制御の評価」, 電子情報通信学会, DEWS2007 論文集, pp. L5-2, 2007.3.
  127. 山元理絵, 吉原朋宏, 小林大, 小林隆志, 横田治夫, 「大規模商用サイトログを用いた Web ページ推薦手法 WRAPL の評価と考察」, 電子情報通信学会, DEWS2007 論文集, pp. L4-1, 2007.3.
  128. 小林大, 田口亮, 横田治夫, 「並列ストレージにおけるサービス性能を保った複製利用負荷均衡化に対する更新リクエストの影響」, 電子情報通信学会, DEWS2007 論文集, pp. L2-1, 2007.3.
  129. 仲野 亘, 小林隆志, 直井聡, 横田治夫, 古井貞熙, 「講義講演シーン検索手法におけるレーザーポインタ情報と音声情報の粒度を考慮した統合」, 電子情報通信学会, DEWS2007 論文集, pp. E1-3, 2007.3.
  130. 三木健士, 横田治夫, 「検索キーワード を含む最小 XML 部分文書抽出のための索引手法」, 電子情報通信学会, DEWS2007 論文集, pp. C1-4, 2007.3.
  131. 油井誠, 宮崎純, 植村俊亮, 「DTM に基づく XML データベースのための逆経路索引」, 第 18 回データ工学ワークショップ (DEWS2007), 電子情報通信学会, 2007.3.
  132. 城戸健太郎, 天笠俊之, 北川博之, 「PC クラスタを用いた XML データ並列処理方式における性能改善と評価」, 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), 2007.3.
  133. Reyn Nakamoto, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura, “Tag-Based Contextual Collaborative Filtering”, 18th IEICE Data Engineering Workshop (DEWS2007), 2007.3.
  134. 武田 裕子, 小口 正人 : 「iSCSI over VPN 環境における複数経路アクセス適応制御手法の提案と評価」, DEWS2007, 2007.3.
  135. 千島 望, 山口 実靖, 小口 正人 : 「iSCSI ストレージにおける VPN 複数経路アクセス時の性能と TCP パラメータ解析」, DEWS2007, 2007.3.
  136. 松井 愛子, 神坂 紀久子, 山口 実靖, 小口 正人: 「無線 LAN ツーホップ環境の中継ノードにおけるモバイル DB 転送時の動作解析」, DEWS2007, 広島, 2007.3.
  137. 吉井迪利, 宮崎純, 植村俊亮, 「自律分析による XPath の出現頻度抽出とデータ分割」, 電子情報通信学会 2007 総合大会論文集, 電子情報通信学会, 2007.3.
  138. 横田治夫, 「機能追加によるストレージの高付加価値化のアプローチ」, 電子情報通信学会, 電子情報通信学会技術報告書, No. MR2006-79, pp. 17-22, 2007.3.
  139. 千島望, 山口実靖, 小口正人: 「VPN 複数経路接続時における iSCSI ストレージアクセスの特性解析」, マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム, 3H-4, pp.631-638, 2007.7.
  140. 松井愛子, 神坂紀久子, 山口実靖, 小口正人: 「MANET におけるモバイル DB データ転送時の中継ノード高機能化」, マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム, 3A-3, pp.451-459, 2007.7.
  141. 原明日香, 神坂紀久子, 小口正人: 「並列相関係数抽出実行時の IP-SAN 統合型 PC クラスタの特性評価」, マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム, 5G-1, pp.1055-1058, 2007.7.



142. 高橋昭裕, 梁 文新, 横田治夫, 「挿入によって他ラベル変更を起こさない可変長 XML ラベルの容量評価」, 電子情報通信学会, 電子情報通信学会 信学技報, Vol. 107, No. 131, pp. 97-102, 2007.7.
143. 渡部 徹太郎, 小林隆志, 横田治夫, 「ファイル検索におけるアクセスログから抽出した関連度の利用」, 電子情報通信学会, 電子情報通信学会 信学技報, Vol. 107, No. 131, pp. 503-508, 2007.7.
144. 並木悠太, 神戸康多, 小林大, 横田治夫, 「並列 B-Tree 構造 Fat-Btree を用いた PostgreSQL の分散検索の試み」, 電子情報通信学会技術研究報告, Vol. 107, No. 131, pp. 473-478, 2007.7.
145. 片居木誠, 小林大, 横田治夫, 「性能の異なるディスクが混在する環境でのデータ配置及びディスク配列に関する考察」, 電子情報通信学会, 電子情報通信学会 信学技報, Vol. 107, No. 131, pp. 533-538, 2007.7.
146. 油井誠, 宮崎純, 植村俊亮, 加藤博一: “Remote Proxy を利用した並列分散 XML 問合せ処理手法の提案”, 電子情報通信学会技術研究報告, Vol.107, No.131, DE2007-22-114 / 情報処理学会研究報告, Vol.2007, No.65, 2007-DBS-143, pp.217-222, 電子情報通信学会/情報処理学会, 2007.7.
147. 並木悠太, 神戸康多, 小林大, 横田治夫, 「Fat-Btree を用いた PostgreSQL 分散化におけるページ分割手法の検討」, 電子情報通信学会技術研究報告, Vol. 107, No. 254, pp. 1-6, 2007.10.
148. 渡部徹太郎, 小林隆志, 横田治夫, 「キーワード非含有ファイルを検索可能とするファイル間関連度を用いた検索手法の評価」, DEWS2008 論文集, 2008.3.
149. 吉野悠二, 梁文新, 横田治夫, 「部分文書処理コストを考慮した XML データの分割配置とアクセス手法」, DEWS2008 論文集, 2008.3.
150. 高橋昭裕, 梁文新, 横田治夫, 「要素挿入に強い XML ラベルの構造情報抽出手法の提案と評価」, DEWS2008 論文集, 2008.3.
151. 並木悠太, 神戸康多, 横田治夫, 「Fat-Btree を用いた PostgreSQL 分散化におけるページ分割手法の評価」, DEWS2008 論文集, 2008.3.
152. 片居木誠, 横田治夫, 「仕様が異なるディスク混在ストレージの帯域・容量同時有効利用に向けた領域割当てとデータ配置の評価」, 電子情報通信学会, DEWS2008 論文集, 2008.3.
153. レー ヒェウハン, ティティポーン ルートラットデーチャーケン, 渡部 徹太郎, 横田 治夫, 「講義講演ビデオからダイジェスト自動作成のための重要シーン抽出手法の評価」, DEWS2008 論文集, 2008.3.
154. 油井誠, 宮崎純, 植村俊亮, 加藤博一: “遅延評価を利用した並列分散 XQuery 問合せ処理”, 電子情報通信学会第 18 回データ工学ワークショップ(DEWS2008), 2008.3.
155. Imam Machdi, Toshiyuki Amagasa and Hiroyuki Kitagawa, “A Multidimensional Data Structure for Maintaining XML Data Partitions”, 電子情報通信学会第 18 回データ工学ワークショップ(DEWS2008), 2008.3.
156. 千島望, 山口実靖, 小口正人: 「VPN 複数経路接続を用いた iSCSI 並列ストレージアクセス性能評価」電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008), D4-5, 2008.3.
157. 松井愛子, 神坂紀久子, 山口実靖, 小口正人: 「マルチホップネットワーク上の分散モバイル DB に適したミドルウェア実装と性能評価」電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008), D6-6, 2008.3.
158. 原明日香, 神坂紀久子, 山口実靖, 小口正人: 「IP-SAN 統合型 PC クラスタにおける並列相関関係抽出実行時のシステム特性解析」電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008), D1-3, 2008.3.
159. レー ヒェウハン, ティティポーン, ルートラットデーチャーケン, 渡部徹太

- 郎, 横田治夫, 「講義講演ビデオの重要シーン抽出によるダイジェスト自動作成」, 情報処理学会, 情報処理学会第70回全国大会講演論文集, 2008.3.
160. 渡部徹太郎, 小林隆志, 横田治夫, 「キーワード非含有ファイルを検索可能とするファイル間関連速度を用いた検索手法」, 情報処理学会第70回全国大会講演論文集, 2008.3.
  161. Imam Machidi, Toshiyuki Amagasa, Hiroyuki Kitagawa, "XML Data Partitioning for Parallel Holistic Twig Join Processing", 第70回情報処理学会全国大会, 筑波大学, 2008.3.
  162. 千島望, 山口実靖, 小口正人:「VPN 複数経路接続時における並列ストレージアクセスの動作解析」情報処理学会第70回全国大会, 2ZK-3, つくば, 2008.3.
  163. 松井愛子, 神坂紀久子, 山口実靖, 小口正人:「MANET におけるモバイル DB データ高度処理のためのフレームワークの提案およびミドルウェア実装」情報処理学会第70回全国大会, 2ZK-6, つくば, 2008.3.
  164. 原明日香, 神坂紀久子, 小口正人:「並列データマイニング実行時の IP-SAN 統合型 PC クラスタの動作解析」情報処理学会第70回全国大会, 2ZK-4, つくば, 2008.3.
  165. ドウンゴ フウン, 勝山裕, 直井聡, 横田治夫, 「Web サーチを活用した TV テロップ認識率向上手法」, 電子情報通信学会, 信学技報, Vol. 109, No. 93, pp. 163-168, 2008.6.
  166. 油井誠, 宮崎純, 植村俊亮, 加藤博一: "計算リソースとデータ量にスケラブルな XML データベースシステムの実現", 情報処理学会研究報告, Vol. 2008 No. 56, 135, 2008.6.
  167. 新井 絵美, 平野 由美, 村瀬 勉, 小口 正人:「無線 LAN 環境におけるトランスポート層プロトコルの相違による帯域公平性の検討」マルチメディア, 分散, 協調とモバイル (DICOMO2008)シンポジウム, 2C-1, pp.304-316, 2008.7.
  168. 宇野 美穂子, 小口 正人:「MANET における汎用 OS を用いたセキュリティメカニズムの応答性評価」マルチメディア, 分散, 協調とモバイル (DICOMO2008)シンポジウム, 7E-4, pp.1644-1650, 2008.7.
  169. 原 明日香, 神坂 紀久子, 山口 実靖, 小口 正人:「並列データマイニング実行時の IP-SAN 統合型 PC クラスタのネットワーク特性解析」マルチメディア, 分散, 協調とモバイル (DICOMO2008)シンポジウム, 3H-2, pp.705-714, 2008.7.
  170. 比嘉 玲華, 神坂 紀久子, 山口 実靖, 小口 正人:「パケットロス発生における iSCSI 遠隔ストレージアクセスに関する評価」マルチメディア, 分散, 協調とモバイル (DICOMO2008)シンポジウム, 3D-4, pp.609-615, 2008.7.
  171. 高山一樹, 横田治夫, 「平文を生成しない分散ストレージ上での再暗号化手法の提案」, iDB フォーラム 2008, 2008.9.

③ ポスター発表 (国内会議 8 件、国際会議 5 件)

1. Kazuhito Kobayashi, Wenxin Liang, Dai Kobayashi, Akitsugu Watanabe, and Haruo Yokota, "VLEI code: An Efficient Labeling Method for Handling XML Documents in an RDB", in Proc. of 21st International Conference on Data Engineering (ICDE2005), IEEE, pp.386-387, 2005.4
2. Akihiro Takahashi, Wenxin Liang, Haruo Yokota, "Storage Consumption of Variable-length XML Labels Uninfluenced by Insertions", in Proc. of the Second IEEE International Conference on Digital Information Management, IEEE, pp. 571-573, 2007.10.
3. Makoto Kataigi, Dai Kobayashi, Haruo Yokota, "Consideration of disk segmentation and disk sequencing for efficient use of disk resource on heterogeneous disk environment", in Proc. of the Second IEEE International Conference on Digital Information Management, IEEE, pp. 577-579, 2007.10.

4. Tetsutaro Watanabe, Takashi Kobayashi, Haruo Yokota: A Method for Searching Keyword-lacking Files Based on Interfile Relationships, 16th Intl Conf. Cooperative Information Systems (CoopIS2008). 2008.11. (accepted)
5. 藤原 勤, 宮崎 純, 植村 俊亮, 「自律ディスクによる広域分散ストレージシステムの通信およびストレージのオーバヘッドを考慮した性能評価」, 情報処理学会研究報告/電子情報通信学会技術研究報告, Vol.2006, No.78, 2006-DBS-140(II), pp.99-106/Vol.106, No.149, DE2006-63, pp.77-82, 情報処理学会/電子情報通信学会, 2006.7.
6. Wenxin Liang, and Haruo Yokota, "A Path-sequence Based Method for Solving the One-to-multiple Matching Problem in Leaf-Clustering Based Approximate XML Join Algorithms", 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 4A-i10, 2006.3.
7. 藤原 勤, 宮崎 純, 植村 俊亮, 「自律ディスクを用いた広域分散ストレージシステムのシミュレーションによる性能評価」, 電子情報通信学会 第 17 回データ工学ワークショップ (DEWS2006) 論文集, 4C-i7, 2006.3.
8. 渡部 徹太郎, 横田 治夫, 「アクセスログから抽出したファイル間関連度を利用した検索システム」, データベースと Web 情報システムに関するシンポジウム(DBWeb2007), 2007.11.
9. 油井誠, 宮崎純, 植村俊亮, 加藤博一: "call-by-need 呼出しを利用した分散 XML 問合せ", 情報処理学会 DBWeb2007, 2007.11.
10. 渡部徹太郎, 小林隆志, 横田治夫, 「キーワード非含有ファイルを検索可能とするファイル間関連度を用いた検索手法の評価」, 第 19 回データ工学ワークショップ, DEWS2008, 2008.3.
11. Wenxin Liang and Haruo Yokota. Syntax-based XML Subtree Segmentation, Matching and Integration. In Proc. of iDB Forum 2008, IEICE Technical Report, Vol. 108, No. 211, pp.7-8 (ISSN 0913-5685), 2008.9.
12. 小田切健一, 渡辺陽介, 横田治夫, 「アクセス履歴に基づくファイル間関連度を用いたデスクトップ情報管理ツールの開発」, 電子情報通信学会, 電子情報通信学会信学技報, No. 329, pp. 49, 2008.12.

### (3)特許出願

#### ①国内出願 (15 件)

1. “コンテンツ蓄積システム、その方法およびそのプログラム”、横田治夫、渡邊明嗣、花井知広、山口宗慶、田口亮、林直人、上原年博、東京工業大学、日本放送協会、平成 15 年 11 月 6 日、特願 2003-376918
2. “ソフトウェア更新方法およびソフトウェア更新プログラム”、横田治夫、渡邊明嗣、花井知広、小林大、山口宗慶、上原年博、田口亮、林直人、東京工業大学、日本放送協会、平成 16 年 2 月 6 日 特願 2004-31042
3. “自律ストレージ装置、自律ストレージシステム、ネットワーク負荷分散プログラム及びネットワーク負荷分散方法”、横田治夫、渡邊明嗣、田口亮、林直人、上原年博、国立大学法人東京工業大学、日本放送協会、平成 16 年 6 月 29 日、特願 2004-190823
4. “ストレージ装置、並びに、コンテンツ蓄積管理方法及びコンテンツ蓄積管理プログラム”、横田治夫、山口宗慶、渡邊明嗣、小林 大、上原年博、田口 亮、林 直人、国立大学法人東京工業大学、日本放送協会、平成 16 年 11 月 10 日、特願 2004-325865
5. “自律ストレージ装置、自律ストレージシステム、負荷分散プログラム及び負荷分散方法”、横田治夫、小林 大、渡邊明嗣、山口宗慶、田口 亮、上原年博、林 直人、国立大学法人東京工業大学、日本放送協会、平成 16 年 12 月 3 日、

特願 2004-351458

6. “ポインタの位置情報を利用するプレゼンテーションコンテンツ検索システムおよびその方法、ならびにコンピュータ読み取り可能な記憶媒体”、横田治夫、小林隆志、仲野亘、越智悠太、直井聡、国立大学法人東京工業大学、平成 17 年 2 月 18 日、特許出願 2005-43011
7. “ディレクトリ更新方法およびディレクトリ更新プログラム、ならびに木構造型データ記憶装置”、横田治夫、吉原朋宏、小林大、上原年博、田口亮、国立大学法人東京工業大学、日本放送協会、平成 17 年 2 月 18 日、特願 2005-43023
8. “データの検索方法”、横田治夫、長良香子、小林大、小林一仁、国立大学法人東京工業大学、平成 17 年 2 月 18 日、特願 2005-43207
9. “自律ストレージ装置、コンテンツ記憶装置及びコンテンツ分散プログラム”、横田治夫、上原年博、田口亮、国立大学法人東京工業大学、日本放送協会、平成 17 年 5 月 23 日、特願 2005-148905
10. “自律ストレージ装置、自律ストレージシステム、負荷分散プログラム及び負荷分散方法”、横田治夫、小林大、渡邊明嗣、山口宗慶、上原年博、田口亮、林直人、東京工業大学、日本放送協会、平成 17 年 7 月 12 日、特願 2005-203524
11. “データ記憶装置およびバージョン管理プログラム”、横田治夫、中野真那、小林大、上原年博、田口亮、国立大学法人東京工業大学、日本放送協会、平成 17 年 7 月 13 日、特願 2005-203896
12. “複数 XML 文書の類似度検出手法および類似度検出システム、ならびに複数の XML 文書の統合手法”、横田治夫、梁文新、東京工業大学、平成 17 年 8 月 16 日、特願 2005-236039
13. “コンテンツ多重蓄積読出し、多重伝送方式”、横田治夫、上原年博、田口亮、南浩樹、国立大学法人東京工業大学、日本放送協会、平成 17 年 8 月 29 日、特願 2005-241792
14. 分散型ストレージシステムのデータ多重化方式、欧阳 湘勇・横田 治夫・吉原 朋宏・小林 大・江尻 革・太田 光彦、富士通(株)、平成 18 年 10 月 6 日、特願 2006-275820
15. 分散型ネットワークストレージシステムにおけるデータ暗号化データ格納方法、横田治夫、高山一樹、小林大、東京工業大学、平成 19 年 2 月 14 日、特願 2007-033716

②海外出願 (0 件)

#### (4)受賞等

##### ①受賞

- ・ 第 2 回日本データベース学会論文賞  
受賞日: 2005.7.14  
受賞論文: 天笠俊之, 植村俊亮:”リージョンディレクトリを用いた関係データベースによる大規模 XML データ処理”, 日本データベース学会 Letters Vol.3, No.2, pp.33-36, 2004.
- ・ DBWeb2005 学生発表奨励賞受賞  
受賞論文: Wenxin Liang, and Haruo Yokota, ”SLAX: An Improved Leaf-Clustering Based Approximate XML Join Algorithm for XML Data Integration at Subtree Classes”, in Proc. of IPSJ DBWeb2005, pp.41-48, 2005.11
- ・ DBWS2005 学生奨励賞受賞  
受賞論文: 渡邊明嗣, 横田治夫, 「Adaptive Overwrapped Declustering: アクセス負荷分散と容量効率向上のための複製管理戦略」, 夏のデータベースワークショップ DBWS2005, 電子情報通信学会、信学技報、DE2005-98 (2005-7), pp.193-198, 情報処理学会、研究報告 2005-DBS-137 (II) (69), pp.515-522, 2005.7
- ・ DBWS2005 学生奨励賞受賞



- 受賞論文: 岡本拓明, 小林隆志, 直井聡, 横田治夫, 古井貞熙, 「講義講演シーン検索における音声データの利用」, 夏のデータベースワークショップ DBWS2005, 電子情報通信学会、信学技報、DE2005-107 (2005-7), pp.1-6, 情報処理学会、研究報告 2005-DBS-137 (II) (78), pp.585-591, 2005.7.
- DEWS2006 優秀論文賞  
受賞論文:加藤英之, 小林隆志, 横田治夫, 「Web サービスベースのワークフロー管理における障害を考慮したアクティビティスケジューリング手法」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 7C-o1, 2006.3.
  - DEWS2006 優秀論文賞  
受賞論文:高砂幸代, 小林亜樹, 山岡克式, 酒井善則, 曾根原登, “Web サーバ間での部分 Web グラフ同期方式の提案”, DEWS2006, 7C-o2, 2006.3.
  - DEWS2006 優秀論文賞  
受賞論文:吉原朋宏, 小林大, 田口亮, 横田治夫, 「並列 Btree 構造 Fat-Btree におけるリクエスト委譲コストを削減する並行性制御手法」, 第 17 回電子情報通信学会データ工学ワークショップ(DEWS2006) 論文集, 7C-o3, 2006.3.
  - データ工学ワークショップ DEWS2006 優秀プレゼンテーション賞  
受賞論文:城戸健太郎, 天笠俊之, 北川博之, 「PC クラスタを用いた XML データ並列処理方式の提案」, 第 17 回電子情報通信学会データ工学ワークショップ (DEWS2006), 2006.3
  - DBWeb2006 学生研究発表奨励賞  
受賞論文: 山元理絵, 小林大, 吉原朋宏, 小林隆志, 横田治夫, 「アクセスログに基づく Web ページ推薦における LCS の利用とその解析」, 情報処理学会, Proc. of IPSJ DBWeb2006, pp. 43-50, 2006.11
  - 情報処理学会データベースシステム研究会学生研究発表奨励賞  
受賞論文:油井 誠, 宮崎 純, 植村 俊亮, 「効率的な XQuery 処理のための DTM に基づく XML ストレージ」
  - IPSJ Digital Courier 船井若手奨励賞  
受賞論文: Wenxin Liang and Haruo Yokota, “SLAX: An Improved Leaf-Clustering Based Approximate XML Join Algorithm for Integrating XML Data at Subtree Classes,” IPSJ Digital Courier, 2007.3.
  - DBWS2007 学生研究発表奨励賞  
受賞論文:渡部 徹太郎, 小林隆志, 横田治夫, 「ファイル検索におけるアクセスログから抽出した関連度の利用」, 電子情報通信学会, 電子情報通信学会 信学技報, Vol. 107, No. 131, pp. 503-508, 2007.7.
  - DBWS2007 学生研究発表奨励賞  
受賞論文:並木悠太, 神戸康多, 小林大, 横田治夫, 「並列 B-Tree 構造 Fat-Btree を用いた PostgreSQL の分散検索の試み」, 電子情報通信学会技術研究報告, Vol. 107, No. 131, pp. 473-478, 2007.7.
  - 平成 19 年度 電子情報通信学会学術奨励賞  
受賞論文:吉井迪利, 宮崎純, 植村俊亮: “自律分析による XPath の出現頻度抽出とデータ分割”, 2007 年電子情報通信学会総合大会論文集, 電子情報通信学会, 2007.3.
  - Merit Paper Award  
受賞論文:Reyn Nakamoto, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura: Contextual Collaborative Filtering Based on Tagging Information”, Proc. of The International MultiConference of Engineers and Computer Scientists 2007 (IMECS 2007), pp.964-969, 2007.
  - DICOMO2007 優秀論文賞  
受賞論文:千島望, 山口実靖, 小口正人:「VPN 複数経路接続時における iSCSI ストレージアクセスの特性解析」, マルチメディア, 分散, 協調とモバイル (DICOMO2007)

シンポジウム, 3H-4, pp.631-638, 2007年7月.

- DEWS2008 優秀論文賞/優秀インタラクティブ発表賞  
受賞論文:渡部徹太郎, 小林隆志, 横田治夫, 「キーワード非含有ファイルを検索可能とするファイル間関連度を用いた検索手法の評価」, DEWS2008 論文集, 2008.3.
- DB ワークショップ 2008 優秀若手研究者賞  
油井誠, 2008.6.
- Excellent Presentation Award  
受賞論文:Wenxin Liang and Haruo Yokota. Syntax-based XML Subtree Segmentation, Matching and Integration. In Proc. of iDB Forum 2008, 2008.9.

### ②新聞報道等

- 毎日新聞社 Web サイト MYCOM PC WEB の記事にて, NHK 技研公開 2004 レポートで自律ディスクが紹介された. (2004.5)
- 日経エレクトロニクス 2004年07月05日号(No.877)の特集「エターナル・ストレージ」(p118～)で自律ディスクが紹介された. (2004.7)
- DB マガジン 2005/02月号のコラム「日本のデータベース研究最前線」(p162～)で自律ディスクが紹介された. (2005.2)
- 毎日新聞社 Web サイト MYCOM PC WEB の記事にて, 自律ディスクの試作機 IP-Storage540 が紹介された (2005.7.12)
- 日経バイト No. 263 特集「生物指向コンピューティング」のコーナー, pp.31にて自律ディスクが紹介された(2005.4.4).

### ③その他

- 2007年11月電子情報通信学会フェロー称号授与 横田治夫  
「データベースシステムとディペンダブルシステムに関する先駆的研究」
- 2009年3月情報処理学会フェロー称号授与 横田治夫  
「高信頼で効率的な情報蓄積とその高度活用に関する先進的研究」

### (5)その他特記事項

- NHK 技研公開 2004 で自律ストレージシステムを展示
- 2005年6月7日～10日 N+I2005(Interop 2005 Tokyo:幕張)にて試作プロトタイプ機(B-type)を展示 (IP-processor コンソーシアムブースにて富士通株式会社が展示)
- 2008年3月12日～13日情報処理学会全国大会併設イベント「わくわくIT」におけるデモンストレーション

## §7 研究期間中の主な活動

年月日	名称	場所	参加人数	概要
平成15年10月16日	研究の立ち上げおよび研究方針の意識合わせのためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	15	研究代表者からの研究方針の説明と各参加者の研究内容の確認
平成16年3月26日	平成15年度の研究成果および平成16年度の研究方針の確認のためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	18	平成15年度の研究成果の取りまとめと平成16年度の研究方針の説明

平成16年7月5日	研究の進捗状況確認とワークショップの準備のためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	16名	各グループ毎に研究の進捗状況を報告するとともに、12月に開催予定のワークショップのワーキンググループを構成
平成16年8月23日	ヒアリングの報告とワークショップ準備のためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	12名	7月28日に行われたヒアリングの報告と12月のワークショップの内容検討
平成16年12月6日～7日	研究進捗状況の確認とトップレベルの専門家からのコメントの収集のためのワークショップ (ADSS2004)	米国カリフォルニア州サンフランシスコ近郊日立アメリカビル (ブリスペイン)	18名	トップレベルの研究者4名による招待講演とグループメンバー12名による研究発表
平成17年6月1日	研究の進捗状況確認とワークショップの準備のためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	11名	各グループ毎の研究進捗状況報告と、12月のADSSワークショップに関する検討
平成17年6月29日	ヒアリングの報告とワークショップ準備のためのミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	10名	7月12日に行われたヒアリングの報告と12月のADSSワークショップの内容検討
平成17年12月16日～17日	研究進捗状況の確認とトップレベルの専門家からのコメントの収集のためのワークショップ (ADSS2005)	米国カリフォルニア州サンフランシスコ市内ホリデーインゴールデングート	32名	トップレベルの研究者5名による招待講演とグループメンバー15名による研究発表
平成18年3月17日	ヒアリングの報告と研究の進捗状況確認	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	12名	1月13日に行われたヒアリングの報告と各グループの研究進捗状況報告
平成18年5月29日	研究の進捗状況確認とシンポジウムに向けた検討のためのチームミーティング	東京工業大学大岡山キャンパス西8E号棟10階1004会議室	12名	6月13日に行われたヒアリングの報告と各グループ毎の研究進捗状況報告およびシンポジウムでのデモの検討
平成18年9月14日	サイトビジットとシンポジウムの検討のための	東京工業大学大岡山キャンパス西	10名	10月6日に行われるサイトビジットへの対応と11月10日のシンポジウムにお

	ミーティング	8 E号棟7階 コラボレー ションルー ム		けるデモの検討
平成19年4月 24日	チームミーテ ィング	東京工業大 学大岡山キ ャンパス西 8 E号棟10階 1004会議室	14名	19年度の方針と各グループ 毎の研究進捗状況報告
平成19年7月 10日	チームミーテ ィング	東京工業大 学大岡山キ ャンパス西 8 E号棟10階 1004会議室	11名	中間ヒアリングと国際ワー クショップ準備
平成19年8月 29日	チームミーテ ィング	東京工業大 学大岡山キ ャンパス西 8 E号棟10階 1004会議室	11名	国際ワークショップ準備
平成19年8月 29日	国際ワークショ ップ ADSS2007	フランス・リ ヨ ン INSA-Lyon	30名以上 (途中で の出入り があり、正 確な人数 は把握不 可能)	IEEE ICDIM 2007 の併設ワ ークショップとして開催、 投稿数15件、口頭発表 7件、 ポスター 3件
平成20年6月 16日	チームミーテ ィング	東京工業大 学大岡山キ ャンパス西 8 E号棟10階 1004会議室	10名	20年度の方針と各グループ 毎の研究進捗状況報告
平成20年8月 4日	チームミーテ ィング	東京工業大 学大岡山キ ャンパス西 8 E号棟7階 コラボレー ションルー ム	10名	最終ヒアリングの報告、研 究進捗状況確認とシンポジ ウムに向けての検討
平成20年8月 4日	チームミーテ ィング	奈良猿沢荘 「おおみや」 会議室	9名	シンポジウムに向けてデモ ンストレーションの検討

#### 国際ワークショップについて

当研究チーム主催の国際ワークショップである ADSS (ADvanced Storage System workshop) を 2004 年、2005 年、2007 年に開催し、チーム内の研究成果の発表の場とするとともに、研究成果に対する海外のトップレベルの専門家からのコメントの収集と、海外の研究動向の調査を目的とし、チームの研究の位置づけを明確にしてきた。

2004 年 12 月 6 日～7 日に米国カリフォルニア州サンフランシスコ近郊のプリズペインに



ある日立アメリカにおいて開催した ADSS2004 は、チーム主催の最初のワークショップのトライアルとして、チームメンバーと招待講演者のみのクローズド形式とし、ワークショップの後で関連する研究所の見学も行った。先進ストレージシステムの最先端の研究者である、CMU の Prof. Greg Ganger、HP Lab. の Dr. Mustafa Uysal、IBM Almaden Research Center の Dr. Winfried Wilcke、東大生産技術研究所の喜連川優教授による招待講演 4 件を含む 17 件の研究発表を行い、その後 IBM Almaden Research Center と HGST Lab. の見学を行って、当該分野のトップレベルの研究者と交流を図るとともに、情報を収集して研究の位置づけを明確にし、大成功を収めた。



ADSS 2004 の様子

2005 年 12 月 16 日～17 日に米国サンフランシスコ市内で開催した ADSS2005 は、発表自体は同様にクローズド形式ではあるが、関連する国際会議 (FAST2005) と同じ会場で国際会議に続く形で開催した。国際会議に引き続いて参加する専門家も多く、参加者の半数以上がメンバー外で盛況であった。CMU の Greg Ganger 教授、New Hampshire 大の Robert Russell 教授、HP の Arif Merchant 博士、御茶ノ水女子大の小口正人助教授、東大の喜連川優教授による招待講演 5 件を含む 22 件の発表があり、熱心な議論が行われた。



ADSS2005 の様子

2007 年 8 月 29 日にフランスのリヨンで開催した ADSS2007 は、ADSS2004、ADSS2005 とは異なり一般に論文募集を行う完全にオープン形式とし、関連する国際会議である IEEE ICDIM2007 の併設ワークショップとした。各国からの 15 件の投稿に対して査読を行い、口頭発表 7 件、ポスター発表 3 件という構成で開催した。参加者数については、出入りがあったので正確な数は把握できなかったが、ICDIM の参加者を含む 30 名以上参加者を得て、盛況に開催された。



ADSS2007 の様子

## §8 結び

新しい情報社会に求められる先進ストレージを実現する上で必要となる各技術要素に対して様々な手法を提案し、多数のノードを用いた実験等を通してそれぞれの提案手法が従来手法に比べて大きく勝っていることについて実証でき、良い成果を出すことができたと判断している。それらに関する論文は、当該分野のトップレベルで日本からの採択数が非常に少ない国際会議である ICDE2005 (採択率 13%) や、同様にレベルが高く難関と定評のある BNCOD2005 (採択率 15%)、EDBT2008 (採択率 17%)、DEXA2008 (採択率 19%)、CoopIS2006 (採択率 28%)、PRDC2004 (採択率 31%) 等の国際会議に採択されており、研究レベルが国際的に評価されたと判断している。また、国内においても様々な会議で多数口頭発表を行い、成果の公表を行ってきた。さらに、VLDB Journal、IEICE Transactions といった国際的に認知されている英文論文誌や国内の電子情報通信学会和文論文誌、情報処理学会ジャーナル等にも多数採録された。これらの口頭発表、論文に対して論文賞、奨励賞等様々な賞を受賞することができた。これにより若手研究者の育成もはかられたと認識している。

また、単に論文等で公表するだけでなく、特許を申請するとともに、実際にプロトタイプとなるハードウェアとソフトウェアを開発し、その効果をデモンストレーション等で目に見える形で示すことができたことは良かった。特に、実際の利用を前提にプロトタイプ用に CIFS インタフェースを開発し、Windows から利用できる環境を整え、これまでのストレージシステムでは不可能であった、利用者から透過的に負荷バランス、容量バランス、ノード追加／削除、故障対策、リカバリといった機能が、十分な性能を保ちながら、実際に実現できていることをその場で確認できるようにしたことは、インパクトが大きかったと思われる。次の写真は、2008年3月12日～13日に情報処理学会全国大会の併設イベントである「わくわく IT」においてデモンストレーションを行ったときの様子である。



わくわく IT におけるデモンストレーションの様子

提案するアプローチの先進ストレージシステムの実用化に向け、完成度を高めてきたと認識しているが、まだ実際に製品化するためにはさらなる検討が必要である。そのため、JST の産学共同シーズイノベーション化事業の顕在化ステージに向けて検討を進めていきたいと考えている。