

# iSCSI 遠隔ストレージアクセスの複数レイヤにまたがる最適化

比嘉 玲華<sup>†</sup> 松原 幸助<sup>††</sup> 岡廻 隆生<sup>††</sup> 山口 実靖<sup>‡</sup> 小口 正人<sup>†</sup>

<sup>†</sup>お茶の水女子大学 <sup>††</sup>ソフトバンクテレコム株式会社 <sup>‡</sup>工学院大学

## 1 はじめに

コンピュータシステムにおける処理データ量の増大に伴い、効率的にストレージを管理したいという要望が高まっている。iSCSI を用いることにより広域環境における IP-SAN を低コストで構築でき、遠隔地のデータセンタなどにデータをバックアップすることが容易となるため、ストレージのアウトソーシングといったサービスへの利用が期待されている。iSCSI とは IP-SAN の代表的なプロトコルであり、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。しかし現状で iSCSI は、複数階層で構成されることなどにより高遅延環境になるに従い性能劣化が著しい。従って iSCSI を用いたストレージアクセスにおいては、iSCSI 層だけではなく複数の層にまたがる制御を施すことによる性能向上が期待される。

## 2 複数レイヤにまたがる最適化

本研究では、SCSI/iSCSI 層においては iSCSI パラメータの最適化、TCP/IP 層においては TCP 輻輳制御アルゴリズムを変化させた場合のスループットの測定比較、Ethernet 層では NIC のパラメータを変更して最適化を行った。

本実験において使用した OS は、Red Hat Enterprise Linux 2.6.18-8.e.15、CPU は Quad Core Intel Xeon 1.6GHz である。iSCSI は Initiator に open-iscsi-2.0-865、Target に iSCSI Enterprise Target(IET)-0.4.15 を使用した。

実験環境を図 1 に示す。

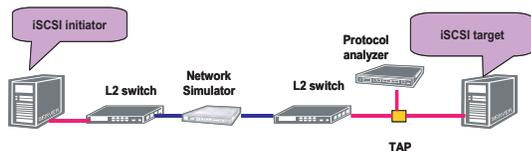


図 1: 実装システム概要

### 2.1 iSCSI パラメータ最適化

本実験においては iSCSI のパラメータをライトアクセス時において最も高い性能が出るように最適化した。また、比較としてソケット通信、ローカルディスクアクセスの性能も測定した。実験結果を図 2 に示す。ローカルディスクとして高性能な SAS ディスクをハードウェア RAID0 構成で用いていることより、ローカルアクセス性能は極めて良い。また最適化により iSCSI アクセスが RTT32ms の場合に約 4 倍の性能向上が得られた。

### 2.2 異なる TCP 輻輳制御アルゴリズムを用いたスループット測定

iSCSI 通信における TCP 輻輳制御アルゴリズムの変更の有効性を評価するために、まず基礎実験としてソケット通信における TCP 輻輳制御アルゴリズムの変更の性

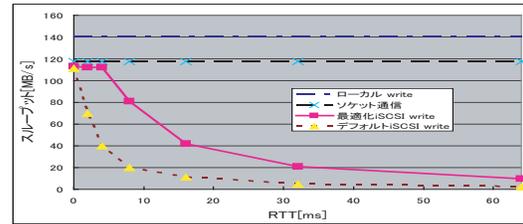


図 2: デフォルト iSCSI, 最適化 iSCSI, ソケット通信のスループット比較

能評価を行い、次に iSCSI 通信における性能評価を行った。実験結果より、ソケット通信時においてはアルゴリズムごとの違いを確認できたが、iSCSI 通信においては見られなかった。その理由としては、TCP の輻輳ウィンドウの違いによる性能向上分が iSCSI のブロックアクセスのシーケンスに吸収されて消えてしまっているのではないかと考えられる。

### 2.3 NIC パラメータの最適化

iSCSI 階層プロトコルの更なる最適化を行うべく Ethernet の NIC パラメータの最適化を行った。その結果 RTT32ms においては約 5%、RTT64ms においては約 8%の性能向上が確認された。

## 3 iSCSI ブロックアクセスのモデル化とその検証

次に iSCSI ブロックアクセスのモデル化を行い、性能決定要因を解析した。理論値と実測値を比較し、これに基づき iSCSI 遠隔ストレージアクセスの性能評価モデルを構築した。その際に、sg\_dd コマンドという iSCSI アクセスにおいて SCSI レベルで指定したブロックサイズによるアクセスが可能となるコマンドを用いた。

その結果、高遅延環境における iSCSI アクセスが理論値よりも低下する理由は、RTT にらず一定であるはずのデータ転送時間が、実際には RTT に比例する値になっていたためであることがわかった。その結果に基づいてデータ転送時間を RTT の一次関数で表しモデル化した値と実測値はほぼ等しく、このモデル化は有効であることが確認できた。

## 4 今後の課題

今後は iSCSI 通信において TCP 輻輳制御アルゴリズムの変更の差が生じないことの原因を輻輳ウィンドウの数値を解析することで解明したいと考えている。また、何故データ転送時間が RTT に比例するのかという原因解析を行うことで iSCSI ストレージアクセスの性能を高める手法を考案、検討していく。更に、現在はソフトウェア iSCSI を Target 側に用いて実験を行っているが、標準的なハードウェア iSCSI Target を用いた比較実験も行っていきたい。