iSCSI 遠隔ストレージアクセス時の性能評価

比嘉　玲華　(指導教員：小口 正人)

1　はじめに

近年、インターネットなどの発達により、企業でも家庭でも、個人の所有する情報量が爆発的に増えてきた。それに伴って問題となってくるのが、データを保全、管理する作業である。データを格納するストレージ装置の増設や万一に備えたデータのバックアップ、データの移管など、面倒な作業が幾つも生じてくる。さらに、データが複数の機器に分散するという問題が起こる可能性もある。

こうした問題に抜本的な解決策を与えるプロトコルとして登場したのが、iSCSI である [1]。しかし、iSCSI は、複雑な階層構成で処理されており、バースト的なデータの転送も多いことから、通常の通信と比較して、特に、高遅延環境においては性能が著しく劣化してしまう。特に、下位基盤の TCP/IP 層が提供できる限界性能を超えることはできず、最大限の性能が発揮できるよう TCP パラメータなどを制御することが求められる [2]。一般に需要が多い遠隔ストレージへのデータバックアップを考えた場合、データの読み出し量よりも書き込み量の方が圧倒的に多く、また遠隔ストレージ側には標準的な環境のみを使用し、カスタマイズが不可能である場合も多い。

そこで本研究では、iSCSI のパラメータを最適に設定し、Initiator 側をカスタマイズして TCP 輻輳ウィンドウの振舞とスループットを観察する。特に、高遅延環境における iSCSI のシーケンシャルライトアクセスの性能を高めるための手法を考案、検討する。

2　研究背景

2.1　Linux TCP 実装

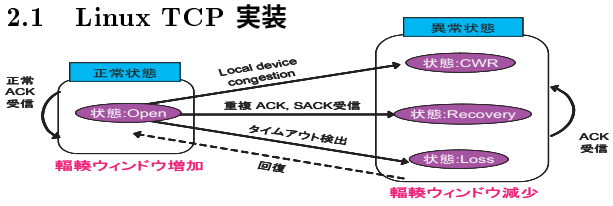

図 1: LinuxTCP の状態遷移

TCP では，通信の制御にウィンドウサイズという概念を用いている。ウィンドウサイズとは、ホストが ACK なしに一度に送信できるデータのサイズで、TCP ヘッダに含まれる。また、このウィンドウサイズは、データの送信側では輻輳ウィンドウ、受信側では広告ウィンドウと呼ばれ、このどちらか小さい方の値がウィンドウサイズとして用いられる。広告ウィンドウは現在の受信ウィンドウの空き容量を示しており，ACK で送信側に送られる．一方、輻輳ウィンドウは送信側の制御パラメータで、ネットワークの混雑を回避するため送信側が OS のカーネル内において自主的に制限する値である。輻輳制御ではこの輻輳ウィンドウが利用されている。

本実験で用いた Linux OS においては、通信時の状態が正常であれば ACK 受信ごとに輻輳ウィンドウは増加するが、エラーが検出されると異常と判断され、輻輳ウィンドウは低下する (図 1)。輻輳ウィンドウが低下する原因としては、送信側デバイスドライバのバッファが溢れることによる Local Congestion エラーを検出した場合（CWR)、重複 ACK 又は SACK を受信した場合（Recovery)，タイムアウトを検出した場合（Loss）の 3 つが挙げられる．また、Linux の TCP 実装では、通信中に一度設定された輻輳ウィンドウは、そのウィンドウの値を使い切らない限りは変化しないという特徴を持ち、この時スループットはほぼ一定の値で安定することが確認されている。

3　研究概要

3.1　実験手順

本研究において、Initiator と Target 間は GigabitEthernet で接続し、TCP/IP 接続を確立した。使用したシステムを表 1 に示す。また、Initiator 側の TCP ソースコードにモニタ用の関数を挿入し、ユーザ空間からもアクセス可能なカーネルメモリ空間に記録する仕組みを作成した。これにより、Initiator から Target へのライトアクセス時の輻輳ウィンドウなど TCP パラメータの値が観察可能になる。遅延装置を使い高遅延環境を作り出し、デフォルトの iSCSI とパラメータ設定を変更した iSCSI を起動して測定を行った。

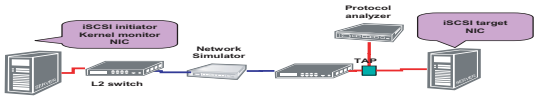

図 2: 実装システム

OS	Red Hat Enterprise Linux 2.618-8.e.15
CPU	Quad Core Intel Xeon 1.6GHZ
Main Memory	2GB
NIC	Intel PRO/1000PT Server Adaptor on PCI Express
HDD	73GB SAS×2(RAID0)
RAID Controller	SAS5/iR
iSCSI	Initiator : open-iscsi-2.0-865
	Target : iSCSI Enterprise Target(IET)-0.4.15
Network Analyzer	ClearSight Network Recorder
Network Simulator	ANUE

表 1: 実験環境

3.2　iSCSI パラメータ設定

本研究において、iSCSI のパラメータ設定をライトアクセス時における最適な状態になるように調整した。Target 側で、Unsolicited なライトアクセス通信を行えるように設定し、Unsolicited なライトの最大長を 64KB から 256KB へ、Target が受信する PDU の最大セグメント長を 8KB から 128KB へ変更した。

3.3　bonnie++

ハードディスクベンチマークツールとしては bonnie++1.03 を用いた [3]。これはデータベースのような大規模なファイル操作のスループットを測定できる。また比較的小さなファイルの作成、読込み、削除のスループットも測定可能である。本研究で

は、Sequential Write(連続書き込み) のスループットを測定した。

4 実験結果

4.1 ローカルディスク、iSCSI アクセスにおけるスループット

遅延装置を使って、片道遅延時間 0,1,2,4,8,16ms の遅延環境を作り、デフォルトの iSCSI と最適パラメータ設定を行った iSCSI のスループットを測定した。また比較として、ローカルディスク (SAS) アクセスの性能も測定した。この結果を図 3 に示す。ローカルディスクに高速な SAS ディスクをハードウェア RAID0 構成で用いているため、ローカルアクセスが極めて性能が良いことが確認できた。これと比較すると iSCSI の性能は低くなる。iSCSI を用いた場合も低遅延環境においては比較的良好なスループットが出ているが、高遅延環境においては、遅延時間と反比例するようにスループットが低下していた。また、最適パラメータ設定の iSCSI のスループットはデフォルトの iSCSI のスループットよりも高い値になっているが、高遅延環境になるにつれて差がなくなっている。

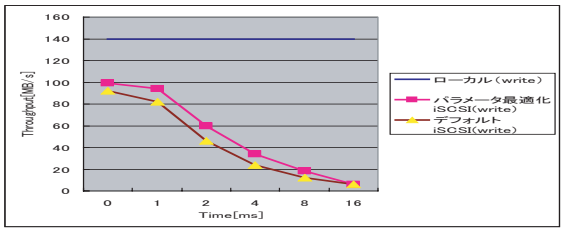

図 3: ストレージアクセスのスループット

4.2 輻輳ウィンドウモニタ

高遅延環境においてスループットが急激に低下する背景として、以下のようなことが考えられる。iSCSI は SCSI コマンドを、TCP/IP パケット内にカプセル化しており、SCSI over iSCSI over TCP/IP over Ethernet という複雑な構成となっている。iSCSI を用いる通信は下位レイヤである TCP/IP の提供するスループットを超えることは不可能であり、TCP の設定や振舞が性能に大きな影響を与えていると考えられる。

そこで、輻輳ウィンドウの値をモニタし、振舞を調べた。write システムコールにより Direct I/O を行うプロセスを 20 並列で起動し、ターゲットへのライトアクセスを実行した（図 4、図 5）。片道遅延時間 8ms の遅延環境において測定した。パラメータを最適化した iSCSI では、デフォルトの iSCSI に比べて、輻輳ウィンドウの成長は比較的早く、またその最大値は約 300 から約 400 へと大きくなることが確認できた。

4.3 プロトコルアナライザ

高遅延環境において性能が著しく劣化する原因を解明するため、本研究ではネットワーク上のパケットを調べていく。プロトコルアナライザを設置し、iSCSI アクセス時のパケットキャプチャを行える環境を整えた。図 6 にラダー表示を示す。パケット間

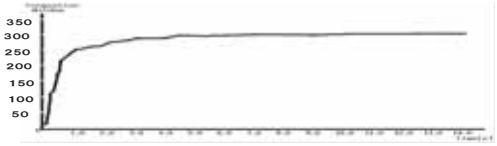

図 4: デフォルト iSCSI 輻輳ウィンドウ

図 5: パラメータ最適化 iSCSI 輻輳ウィンドウ

隔が他のパケット間に比べて大きい箇所を見つけ出し、性能低下の原因を解明していく。

図 6: プロトコルアナライザ　ラダー表示

5 まとめと今後の課題

iSCSI のパラメータを最適値に設定することで、シーケンシャルライトアクセスにおけるスループットが向上すること、および輻輳ウィンドウの値が大きくなることが本研究で確認できた。しかし、高遅延環境になるにつれてスループットが大幅に低下し、差もあまり見られなくなってしまう。その原因としては、大きなブロックサイズで write システムコールを発行しても、SCSI 層において小さなブロックサイズに分割されてしまうことや、輻輳ウィンドウ制御の ACK 待ち状態が起きていることなどが考えられる。これについてはネットワーク上のパケットを調べれば解明できるため、プロトコルアナライザを設置し、iSCSI アクセス時のパケットキャプチャを行える環境を整えた。今後はパケットを解析していくことで高遅延環境で性能が落ちる原因を具体的に調べていく。

参考文献

[1] 喜連川優，ストレージネットワーキング，オーム社出版局

[2] 豊田真智子，山口実靖，小口正人：”高遅延ネットワーク環境における iSCSI リードアクセス時の TCP 輻輳ウィンドウ制御手法の性能評価” SACSIS2005，pp.443-450，2005 年 5 月

[3] Bonnie++ http://www.textuality.com/bonnie/intro.html_5.c

[4] 比嘉玲華、神坂紀久子、山口実靖、小口正人:”iSCSI 遠隔ストレージアクセス時の TCP の振舞に関する一検討”, 情報処理学会第 70 回全国大会,2ZK-2,2008 年 3 月発表予定