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

千島　望[†]　豊田　真智子[† ‡]　山口　実靖[⋆]　小口 正人[†]

† お茶の水女子大学　⋆ 東京大学 生産技術研究所

Analysis of TCP Parameters on of iSCSI Storage Access through VPN

Nozomi Chishima[†]　Machiko Toyoda[† ‡]　Saneyasu Yamaguchi[⋆]　Masato Oguchi[†]

†Ochanomizu University　⋆Institute of Industrial Science, The University of Tokyo

1　はじめに

近年，インターネット技術の進展などにより，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．そこでストレージネッワークが登場し，その代表的なものとして FC-SAN(Fibre Channel - Storage Area Network) が広く用いられるようになった．SAN とは，サーバとストレージを物理的に切り離し，ストレージ同士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 [1][2]．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．

現状において，SAN はサーバサイト内のみでしか使用されていない．そこで，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する．iSCSI は複雑な階層構成のプロトコルスタックで処理されており，バースト的なデータ転送も多いことから，通常のソケット通信と比較して，特に高遅延環境においては性能の劣化が著しい [3]．また，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．

そこで本研究では，iSCSI ストレージアクセスにおいて，VPN を利用した時に，TCP パラメータである輻輳ウィンドウはどのような振舞をするのか解析する．さらに，VPN 利用時における輻輳ウィンドウ制御によりストレージアクセス性能にどのような影響が現れるか評価する．

‡ 現在：NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories
⋆ 現在：工学院大学
Kogakuin University

2　VPN

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．

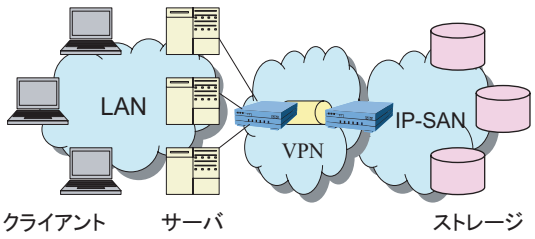

図 1: VPN 利用モデル

iSCSI を用いて遠隔バックアップなどを行うには，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，広域ネットワーク内の VPN 越しにアクセスを行うという方法が考えられる (図 1)．この場合，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有り得る．iSCSI は通常ギガビットクラス以上の太いネットワーク上で用いられるが，途中に細い回線が挟まることにより，トラフィックとして大いに性質の異なるものになると考えられる．従って iSCSI が最大限の性能が発揮できるように TCP パラメータなどを制御することが求められる．

3　Linux TCP 実装

TCP では，通信能力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，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いう値が決定され，このどちらか小さい方がウィンド

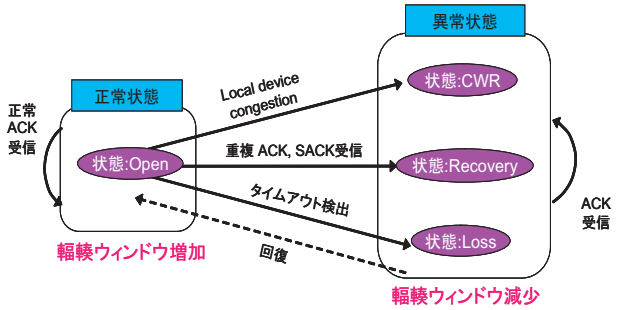

図 2: LinuxTCP の状態遷移

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 LinuxOS においては，通信時の状態が正常であれば ACK の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 2)．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．

4 既存研究

我々は，これまでに iSCSI ストレージアクセスにおいて，輻輳ウィンドウ値を動的にコントロールする手法を提案した [4]．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た．

また，iSCSI を用いたアプリケーション実行性能と TCP パラメータの相関関係の評価も行った [5]．その結果，広告ウィンドウの値を制限することで，輻輳ウィンドウの値も制限でき，それによって実行性能にも影響が出ることが確認された．

5 実験システム

本研究では，iSCSI ストレージアクセスにおいて，VPN を利用した時の TCP パラメータである輻輳ウィンドウと，その輻輳ウィンドウ制御によるストレージアクセスの性能を評価するために，図 3 に示す実験環境を構築した．

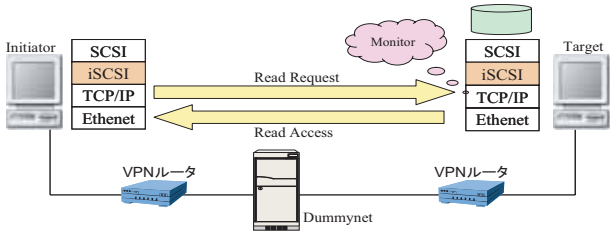

図 3: 実験システムの概要

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 Target の間に VPN ルータを 2 台挟み，さらに，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 [6]．Initiator と Target には，OS が Linux2.4.18-3，CPU が Intel Xeon2.4GHz，MainMemory が 512MB DDR SDRAM，NIC が Intel Pro/1000XT Server Adapter on PCI-X (64bit,100MHz)，iSCSI は UNH IOL reference implementation ver.3 on iSCSI Draft 18 を用いた [7]．この実験環境において，TCP 輻輳ウィンドウの影響を見るため，モニタ関数を挿入しカーネルを再コンパイルした．そして，VPN 接続環境において，1 対 1 の iSCSI シーケンシャルリードアクセス時のデータを集め，その性質を調べた．

本実験ではストレージアクセスのみの性能を評価するため，Initiator の raw デバイスを使用することにより，キャッシュの影響を排除した．また，iSCSI ストレージアクセスにおけるネットワーク性能に焦点を当てて評価を行うため，Target は UNH 実装が提供するメモリモードで動作させ，ディスクアクセスを伴わないようにした．

6 VPN 利用による影響

6.1 輻輳ウィンドウへの影響

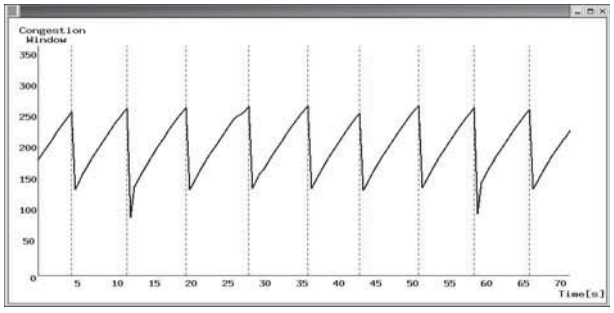

図 4: 輻輳ウィンドウの変化（VPN なし）

図 4，5 は TCP 輻輳ウィンドウをモニタした際の時間変化の様子である．図 4 は VPN ルータを挟まず，iSCSI シーケンシャルリードアクセスの通信を行った時の輻輳ウィンドウをモニタした様子である．また，図 4 に示された細かい縦の破線は Local device congestion(CWR) エラーが起こったことを表しており，これは送信側のデバイスドライバのバッファが溢れることによるエラーである．輻輳ウィンドウは約 250 パケットまで増加した後，CWR エラーが検出され輻輳ウィンドウが急激に減少している．

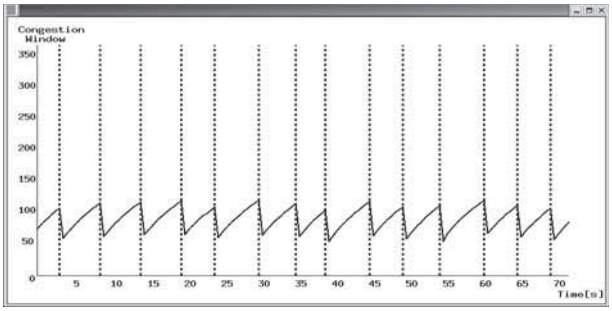

図 5: 輻輳ウィンドウの変化（VPN あり）

図 5 は VPN ルータを 2 台挟んだ時の輻輳ウィンドウをモニタした様子である．また，図 5 に示された太い縦の破線は重複 ACK，SACK を受信したことによるエラーであり，これはパケットロスによるものである．輻輳ウィンドウは約 120 パケットまで増加した後，エラーが検出されている．

このように，VPN ルータを挟むことによって，輻輳ウィンドウの上限値は低下し，また，エラーの種類も変化した．

VPN を挟まない時は，途中のネットワークにより通信が制限されることなく，輻輳ウィンドウが高い値まで増加している．そして送信側のデバイスドライバのバッファが限界に達すると CWR エラーが起こり，輻輳ウィンドウが急激に減少する，という動作を繰り返して鋸型のグラフになっている．これに対し，VPN ルータが挟まれた時は，送信側のバッファより先に通信経路途中のルータが限界に達するため，輻輳ウィンドウが十分に高い値になる前にパケットロスで低下し，これを繰り返して低い位置での鋸型のグラフとなっている．

6.2 スループットへの影響

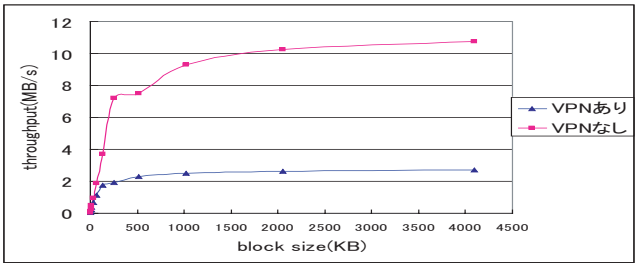

図 6: ブロックサイズの変更

図 6 は iSCSI ストレージアクセスにおいてブロックサイズを変えていった時のスループットの値である．この実験において，片道遅延時間は 16ms に設定した．グラフ上側が VPN 接続を用いない直接接続通信の場合，下側が VPN 接続の場合である．グラフからもわかるように，VPN 接続の場合ブロックサイズが 500KB を過ぎた時からスループットにあまり変化はなく，ほぼ 2.5MB/s で落ち着いた．直接接続通信の場合はブロックサイズが 2000KB を過ぎた時からスループットにあまり変化はなく，スループットは 10.5MB/s 位で落ち着いた．また，VPN 接続にすることで，スループットは著しく低下することが確認された．

図 7 は iSCSI ストレージアクセスにおいて片道遅延時間を変えた時のスループットの値である．この実験において，ブロックサイズは 1024KB に設定した．図 8 は VPN 接続の場合のみの結果を拡大したものである．これらのグラフより，遅延時間を長くするとスループットが著しく低下することがわかる．また，直接接続通信の場合には遅延時間が短い時スループットは急激に減少しており，VPN 接続環境においては急激な減少は見られない．さらに，片道遅延時間が 64ms の場合，直接接続通信の場合では約 97% も性能が低下しているのに対し，VPN 接続の場合は約 86% 低下している．これは，VPN 接続の場合はもともとのスループットが低いためと考えられる．いずれにしても，遅延の起こる環境では，性能が著しく低下することがわかる．

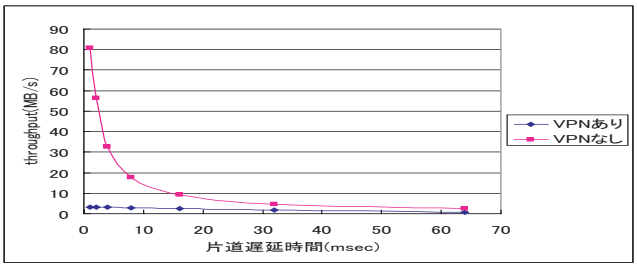

図 7: 片道遅延時間の変更

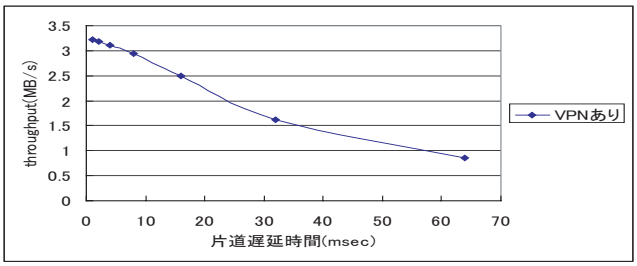

図 8: 片道遅延時間の変更 2

7 輻輳ウィンドウコントロール手法の適用

7.1 実験概要

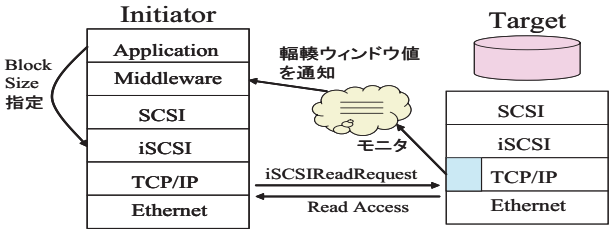

図 9: コントロール手法適用概要

我々は，これまで 1 対 1 ソケット通信時に，iSCSI ストレージアクセスにおいて，輻輳ウィンドウ値を動的にコントロールする手法を提案した [4]．これを，VPN 接続環境において適用する．図 9 は，このコントロール手法の概要図で，iSCSI シーケンシャルリードアクセス時に Target の輻輳ウィンドウをモ

ニタし，CWR エラーが起きた時，Initiator に輻輳ウィンドウ値を通知する．通知を受けた Initiator はミドルウェアでブロックサイズを決定し，アプリケーションがブロックサイズを再指定する．その値を受け Initiator から Target にシーケンシャルリードコマンドを送信し，ストレージアクセスを行う．Target は Initiator に向けて要求されたブロックサイズのデータ転送を実行する．この処理を繰り返すことで，輻輳ウィンドウは CWR エラーが起こらない限界値で一定に保たれる．

7.2 実験結果

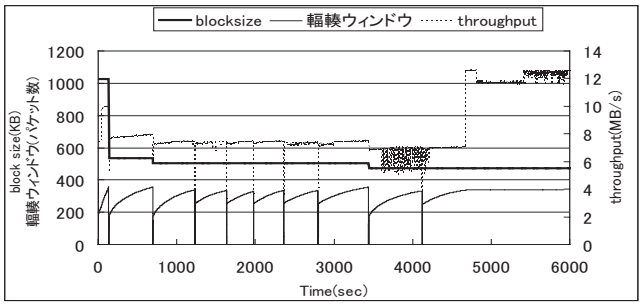

図 10: コントロール手法適用（VPN なし）

図 10 に，この手法を用いた場合の実験結果として，片道遅延時間 16ms の環境における輻輳ウィンドウ，ブロックサイズ，スループットの時間変化を示す．輻輳ウィンドウが一定値となった後のスループットは，鋸型の変化をする時に比べ大幅に向上している．

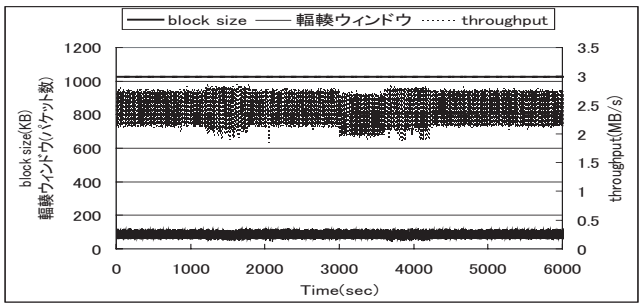

図 11: コントロール手法適用（VPN あり）

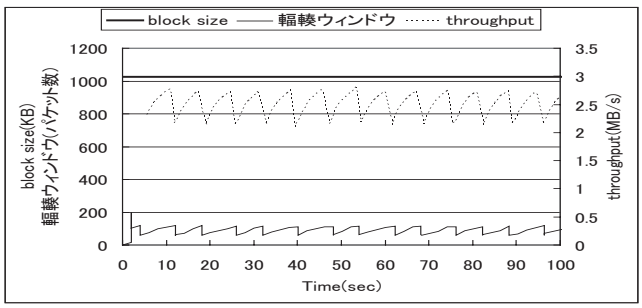

図 12: コントロール手法適用 2（VPN あり）

次に，VPN 接続環境において同様のコントロール手法を用いた場合の結果として，図 11 に片道遅延時間 16ms の環境における輻輳ウィンドウ，ブロックサイズ，スループットの時間変化を示し，図 12 にこの場合の 100s までの結果を拡大したものを示す．この時，ブロックサイズは変化せず，輻輳ウィンドウは一定値にならずに鋸型の変化を繰り返す．それに伴いスループットも鋸型の変化を繰り返す．また，全体のスループットも 2.5MB/s とコントロール手法適用前と比べてほぼ変化していない．VPN 接続環境では先程述べた通り，CWR エラーは起こらず，代わりにパケットロスによる輻輳ウィンドウ低下が頻繁に起こっている．そのため CWR エラーの検出を元にブロックサイズを変更させる従来の輻輳ウィンドウコントロール手法は VPN 接続環境には対応しきれていない．従ってパケットロスによる輻輳ウィンドウ低下にも対応できるよう，従来手法を改良することが必要である．

8 まとめと今後の課題

本研究では，VPN 接続環境における iSCSI ストレージアクセス時に輻輳ウィンドウはどのような振舞をするか，また，TCP パラメータである輻輳ウィンドウのコントロール手法を適用した場合，輻輳ウィンドウ，実行性能にはどのような変化が起こるか評価した．輻輳ウィンドウは，VPN 接続環境ではパケットロス，VPN を用いない直接接続環境では CWR エラーと全く異なるエラーが検出された．従来のコントロール手法は，CWR エラーの検出を元に制御を行うため，VPN 接続環境には対応できていないことがわかった．

今後は，CWR エラー以外の重複 ACK/SACK の受信，タイムアウトの検出の場合にも輻輳ウィンドウをコントロールするように実装を改良し，その時の性能評価を行う．さらに，1 対 1 通信のみではなく，1 対多，多対多の場合の性能評価も行いたい．

参考文献

[1] iSCSI Specification, http://www.ietf.org/rfc/rfc3720.txt?number=3270

[2] SCSI Specification, http://www.danbbs.dk/d̃ino/SCSI

[3] 山口実靖，小口正人，喜連川優：”高遅延広帯域ネットワーク環境下における iSCSI プロトコルを用いたシーケンシャルストレージアクセスの性能評価ならびにその性能向上手法に関する考察”，DEWS2003，4-B-02，March 2003

[4] 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”, SACSIS 2005，pp.443-450，2005 年 5 月

[5] 千島 望，豊田 真智子，山口 実靖，小口 正人:”iSCSI における TCP パラメータとアプリケーション実行性能の相関関係評価” 第 68 回情報処理学会全国大会，pp.131-132，2006 年 3 月

[6] L.Rizzo:”dummynet”, http://info.iet.unipi.it/l̃uigi/ip_dummynet/

[7] InterOperability Lab, Univ,of New Hampshire, http://www.iol.unh.edu/consortiums/iscsi/