

VPN 遠隔接続環境における iSCSI 遠隔 ストレージアクセスの各層最適化についての評価

浅田 菜那^{†1} 比嘉 玲華^{†1} 小口 正人^{†1}

コンピュータシステムにおける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。またストレージの管理コスト低減などの目的で SAN の導入が進んでいる。現状では、SAN は主にローカル環境のみで用いられている。そこで、ネットワークを利用した IP-SAN として iSCSI が注目されている。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サ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可能になるためである。本研究では、VPN を利用することにより iSCSI を広域ネットワークに適用させ、高遅延環境における iSCSI ストレージアクセスの特性、解析を評価しスループット向上の方法について検討する。

An Examination of Each Layer Optimization of Performance iSCSI Remote Storage Access using VPN connection

NANA ASADA,^{†1} REIKA HIGA^{†1}
and MASATO OGUCHI^{†1}

As the volume of data computer systems process increases, it is important that storage is managed efficiently. The introduction of SAN progresses for the purpose of the storage management cost reduction. SAN is mostly used in the server site currently. iSCSI is expected as IP-SAN that uses IP network. Because iSCSI can configure the wide area IP-SAN with low cost and perform easy data backup to remote place, for example data center, iSCSI is expected to be used as Storage Outsourcing service.

In this research, we have first applied iSCSI used in a local environment in the WAN using VPN router. Second, we evaluated performance in the case of iSCSI storage access in WAN and examined the method to improve throughput.

1. はじめに

近年、インターネット技術の進展などにより、個人のユーザや企業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 とは、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、また距離に制約がある。これに対し、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の発展により、ギガビットクラスの回線実現が期待され、iSCSI の有効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使用されている。そこで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 で接続することが期待されている。

そのような背景をふまえて本研究では、拠点間接続等に用いられる 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。VPN 環境ではルータにおけるパケット暗号化処理等により、通常のネットワークとはトラフィックの性質が異なったものになると考えられる。そこで本研究では、広域ネットワークを遅延装置で模擬したネットワーク上に VPN を張った環境において iSCSI 遠隔ストレージアクセスを実行した際の性能について検討を行う。

2. 研究背景

2.1 iSCSI

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

^{†1} お茶の水女子大学

ケットでカプセル化する規格で、iSCSIによりSANをIP機器だけで構成することが可能となる。またiSCSIは図1のように、SCSI over TCP/IP over Ethernetという複雑な階層構造のプロトコルスタックとなる。そのオーバーヘッドなどが影響し、iSCSIによる通信は特に高遅延環境においては大幅に性能が劣化することがわかっている [1]。そこでそのような環境におけるスループットの低下を改善することが求められている。

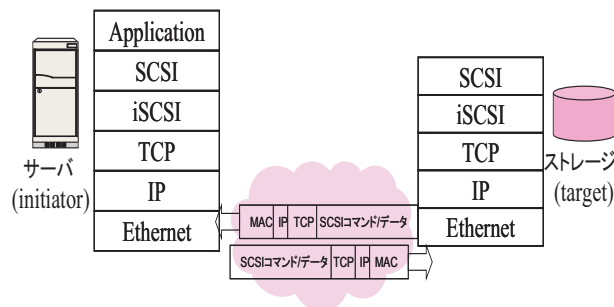


図1 iSCSIのプロトコルスタック

2.2 iSCSI リモートストレージアクセス時における性能低下問題

iSCSIは接続距離に制約を受けないことや、安価に構築できるといった様々なメリットを活かして、高遅延環境での適用が望まれる。しかし現状においてiSCSIは、図2のように複雑な階層構成のプロトコルスタックで処理しており、パ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ことが知られている。

実際に本論文で用いた実験環境においても後述するVPN遠隔接続環境とVPNを使用しない遠隔接続環境の2種類の高遅延環境を構築しiSCSIの性能測定を行った。その結果を示した図によると、ソケット通信時においては高遅延環境下においてもほぼ性能を保っているのに対して、iSCSI通信においては、高遅延環境になるにつれて性能が劇的に劣化していったという結果が得られた。

やはり、本実験環境においても、高遅延環境下における急激な性能劣化という結果となった。リモートバックアッププロトコルとしてiSCSIを使用するためには、この問題を解決しなければならない。

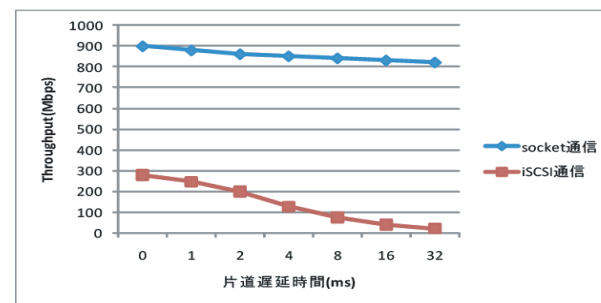


図2 スループット比較

2.3 VPN (Virtual Private Network)

VPNの構成例を図3に示す。VPNは公衆回線をあたかも専用回線のように利用できる仕組みである。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用い、拠点間を仮想的に閉じたネットワークで接続する。今日では、企業内ネットワークの拠点間接続などに使用されている。VPNの大きなメリットは専用回線を導入するよりもコストを抑えられることである。一方、公衆網のデメリットとして機密性が低いことがあげられるが、VPNでは暗号化や認証などを用いて、実質的な専用網を実現する。本研究では広域ネットワーク環境におけるiSCSIの振舞を解析するため、VPNルータを使用する。

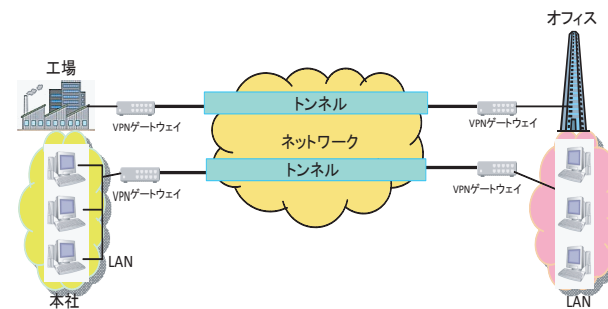


図3 VPN使用例

3. 既存研究

我々は、これまでに 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ときのスループットの違いを観察するという実験を行った [2].

この実験によって、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ことがわかっている。

この既存研究と本研究との最も大きく異なる点は実験環境のソフトウェア環境である。この既存研究において Initiator と Target には、OS は Linux2.4.18-3, CPU は Intel Xeon 2.4GHz, Main Memory は 512MB を使用しており、iSCSI は UNH IOL reference implementation ver.3 on iSCSI Draft 18, VPN ルータは Fujitsu Si-750 を用いた [4]. 既存研究は iSCSI の性質を調べることにに関して成果を挙げたが、使用した iSCSI ドライバは試験的な実装であり、その性能は高いものではなかった。それに対し、本研究では Initiator と Target の OS は Linux2.6.18-8.e15, iSCSI は Initiator に Open-iSCSI, ターゲットに iSCSI Enterprise Target を使用しており、ローカル環境では十分に高い性能を発揮できる iSCSI 実装である [5].

また、我々は iSCSI 遠隔ストレージアクセス時のパケット送信に関する考察についても研究を行ってきた。この実験において、iSCSI パラメータ、輻輳制御アルゴリズム変更、NIC パラメータの複数の層にまたがる最適化とデータ転送部分における解析、ソケットバッファ解析という3つの実験を行った結果、RTT20ms においてデフォルト時よりも約7倍スループットがあがり、大幅に性能を向上させることが出来た。ただしこの実験において、Initiator と Target の間には遅延装置しか挟んでおらず、遠隔アクセスの往復遅延時間が長くなると性能にどのような影響が見られるかという評価およびその際の性能向上のみに焦点が当てられてきた。しかし現実の広域ネットワーク環境においては遠隔アクセスの両端が長いケーブルのみで接続されている訳ではなく、間の接続に VPN などを用いることが一般的である。そこで、本研究では遅延装置だけでなく VPN を使用することによる、より実用性を考慮した手法の提案を行う。

4. 実験概要

本章では、本実験で使用した測定ツール、実験環境および実験手順を示す。

Initiator と Target の OS は Linux2.6.18-8.e15, CPU は Intel Xeon 2.4GHz, Main Memory は 512MB を使用しており、iSCSI は Initiator に Open-iSCSI, ターゲットに iSCSI

Enterprise Target を用いた。また VPN ルータには Fujitsu Si-750 を用いた。これは 3DES 暗号化速度が最大 500Mbps である。

4.1 測定方法

本実験では、sg_dd コマンドを用い、シーケンシャルライトアクセスのスループットを算出した。また、リモートバックアップを考えた場合、データの書き込み量と、読み込み量とを比較すると、圧倒的に書き込み量のほうが多い。以上のことを踏まえ、本研究においては、シーケンシャルライトアクセスのスループットの焦点を当てて実験を進めていく。

4.2 システム構成

本実験では VPN 遠隔接続環境 (図 4) と VPN を使用しない遠隔接続環境 (図 5) の2つのシステム構成を作り、それぞれにおいて連続書き込み、連続読み込みを測定した。

4.2.1 VPN 遠隔接続環境

はじめに図 3 に示す実験環境を構築した。VPN ルータを 2 台用いて一方の LAN 空間には Initiator を、もう一方の LAN 空間には Target を接続させた。2 台の VPN ルータの間には、広域ネットワークを想定して人工的な遅延装置である Dummynet を挿入した。Dummynet には FreeBSD 4.9-RELEASE を用いた。

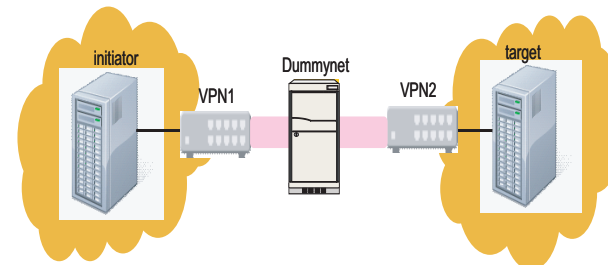


図 4 VPN 遠隔接続環境

4.2.2 VPN を使用しない遠隔接続環境

次に図 4 のように VPN ルータを外し、同じアドレス空間のネットワークで Initiator と Target を接続させた。図 3 の場合と同様に Initiator と Target の間に Dummynet を挟み、測定した。

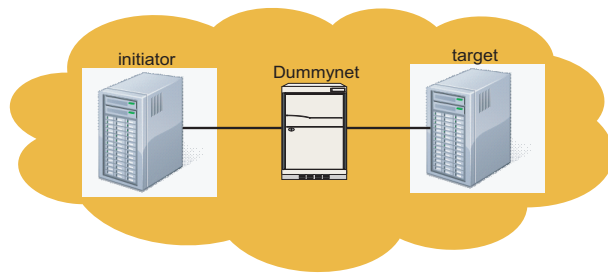


図 5 VPN を使用しない遠隔接続環境

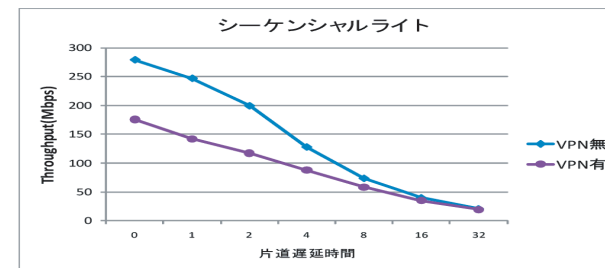


図 6 シーケンシャルライトアクセスのスループット

5. 2 種類の環境における測定結果の比較と考察

図は上記の2種類の環境におけるシーケンシャルライトアクセスのスループットの結果である。遅延時間が大きくなるにつれスループットは低下しているため、性能が落ちていることがわかる。VPN ルータを挟んだ場合と挟まない場合を比較したところ、全体的にルータを設定しない環境におけるスループットが高くなった。しかし片道遅延時間が大きくなるにつれ、ルータを設定する場合としない場合の差は縮まっている。片道時間 32ms における測定結果は連続書き込みと連続読み込みのどちらにおいてもほぼ同程度の数値となり、それぞれ2本のグラフにも差が見られない。これは遅延時間が短いときは Target 側の VPN ルータでの処理がボトルネックとなっていたため、VPN 経路接続をしない環境における結果の方がスループットが高いが、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、VPN 接続環境でない場合における結果との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低遅延の場合には VPN ルータの性能向上が期待され、高遅延になったら iSCSI の性能向上が期待される結果となった。

6. iSCSI 層における最適化

6.1 iSCSI パラメータ最適化

iSCSI はパラメータを最適化することによりスループットが向上することが確認されている。本実験においては、iSCSI のパラメータ設定をライトアクセス時における最適状態になるように調整した。変更内容は表の通りである。

Writeに関するパラメータ		default	1024KB
Target	Initiator	YES	NO
	ImmediateData	NO	YES
	FirstBurstLength	65536	1048576
	MaxBurstLength	262144	1048576
	MaxRecvDataSegmentLength	8192	1048576
Initiator	node.conn[0].iscsiMaxRecvDataSegmentLength	131072	1048576
	node.session.iscsiFirstBurstLength	262144	1048576

表 1 iSCSI パラメータ設定

6.2 iSCSI パラメータ最適化考察

iSCSI パラメータの最適化によって、VPN を使用しない遠隔接続環境ではデフォルト時と比較して片道遅延時間 32ms で約 3 倍の性能向上という結果を得ることが出来た。また VPN 遠隔接続環境においてもわずかではあるが全体を通して性能が向上する結果となった。しかしまだなお高遅延環境下での劇的な性能劣化という問題は解決されていない。

ソケット通信の場合は、高遅延環境においても高いスループットを保っていることが確認されている。したがって iSCSI 通信の場合も、高遅延環境において性能の劣化を極力防ぎ、スループットを保つことが目標となる。

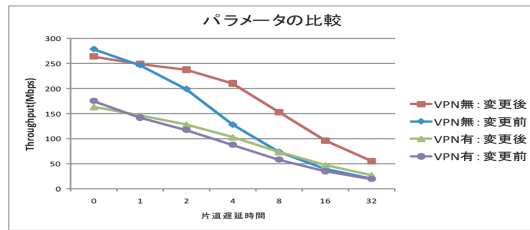


図 7 パラメータ最適化

7. パケット廃棄が起こるネットワーク上の評価

一定の率でパケット廃棄が起こる、現実的なネットワーク上においても測定した。設定しなかった場合と比較しどの程度影響を受け性能が低下するの解析する。なお、パケット廃棄率は 0.01% の場合と、0.02% の場合とを設定し、それぞれ比較した。その結果を図 8 (VPN 遠隔接続環境におけるスループットの比較)、図 9 (VPN を使用しない遠隔接続環境におけるスループットの比較) に示す。

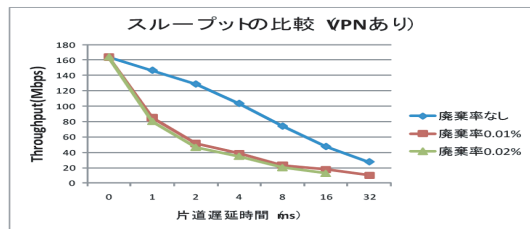


図 8 VPN 遠隔接続環境におけるスループットの比較

パケット廃棄率を設定しない場合において、高遅延環境では急激なスループットの低下がみられることは既に述べたとおりであるが、パケット廃棄率を設定した場合は、低遅延環境においてもスループットが急激に低下する。よってこの低下の改善をすることが、今後望まれると考えられる。

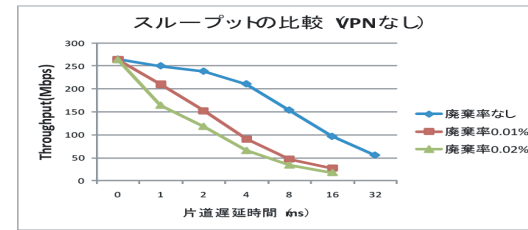


図 9 VPN を使用しない遠隔接続環境におけるスループットの比較

8. TCP 層における最適化

既存研究より、iSCSI 通信中とソケット通信中とは、明らかにソケットバッファの振舞が異なることが確認されている。両者のカーネル中の処理がどのように行われているのかを調べたところ、通過する場所は異なるものの、どちらも timeout が生じる。ソケット通信においては、図 10 における 1415 行目の schedule() を通過し、iSCSI 通信においては 1439 行目の schedule() を通過する際にタイムアウト待ちに入り RTT 後の ACK で再開されており、iSCSI 通信においては不必要なタイムアウト待ちが頻繁に生じていることが分かっている。

両者ともにタイムアウト待ちに入っているのにも関わらず、ソケット通信においては性能は維持されたままで、iSCSI 通信においては性能低下を引き起こしている。その原因としては、ソケット通信の場合はソケットバッファのキューに十分なデータが補遺されているからパケット送信は途切れしないということが考えられる。

8.1 ソケットバッファ最適化

以下の図 11 に、iSCSI 通信中において不必要なタイムアウト待ちに陥るコードの特定までの結果を示す。

この結果から、sk_stream_memory_free 関数が条件分岐となっていることが分かる。sk_stream_memory_free 関数を以下に示す。この関数は sk_wmem_queued と sk_sndbuf の大きさを比較して値を返している。iSCSI 通信中の sk_sndbuf の値を調べたところ、ソケット通信中では 4MB まで成長するのに対して、1MB までしか成長していない。このことから、iSCSI 設定ファイルのどこかで、sk_sndbuf の値が設定されていると想定される。

既存研究においては、sk_sndbuf が大きい場合にタイムアウト待ちに陥ることが解析の結

```

1400 fastcall signed long schedschedule_timeout(signed long timeout)
1401 {
1402     struct timer_list timer;
1403     unsigned long expire;
1404
1405     switch (timeout)
1406     {
1407     case MAX_SCHEDULE_TIMEOUT:
1415     schedule();
1416     goto out;
1417     default:
1425     if (timeout < 0)
1426     {
1427     printk(KERN_ERR "schedule_timeout: wrong timeout
1428     "value %lx from %fn", timeout,
1429     __builtin_return_address(0));
1430     current->state = TASK_RUNNING;
1431     goto out;
1432     }
1433     }
1435     expire = timeout + jiffies;
1436
1437     setup_timer(&timer, process_timeout(unsigned long)current,
1438     expire);
1439     schedule();
1440     del_timer_sync(&timer);
1441
1442     timeout = expire - jiffies;
1443
1444     out:

```

図 10 kernel ソースコード

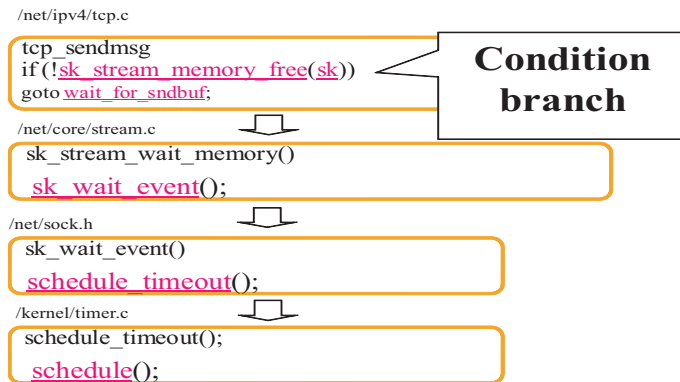


図 11 trace code

果分かったため、sk_sndbuf の値を最適化するソフトウェアモジュールを挿入している。その結果、不必要なタイムアウト待ちが格段に減った。

本研究においては、現在ソフトウェアモジュールを挿入し、測定中である。既存研究においては、ソケットバッファ最適化をしたことにより片道遅延時間 20ms において約 7 倍の性能向上が達成されている。このことから、本研究においても高遅延環境における大幅な性能向上を予想することが出来る。

linux+v2.6.18.5/include/net/sock.h#L446

```

static inline int sk_stream_memory_free(struct sock *sk)
{
    return sk->sk_wmem_queued < sk->sk_sndbuf;
}

```

図 12 sk_stream_memory_free

9. 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のスループットの測定を行った。また VPN 接続をしない環境においても同様に測定し、それらの結果にどの程度差があるか調べた。スループットは片道遅延時間に反比例するが、遅延時間が長くなると 2 つの異なる環境におけるスループットの差はほとんど無くなった。これより VPN 接続環境における iSCSI のリモートアクセスは VPN 接続しない場合に比べ、遠隔アクセス時の性能低下は少なく、遠隔アクセス時には iSCSI の通信プロトコル処理による性能低下が支配的であることが明らかになった。

また iSCSI パラメータを最適化し、わずかではあるが高遅延環境における性能向上を確認することが出来た。今後はこの最適地をパラメータに設定し実験を行っていく。

今後の課題としてはまず、ソケットバッファ最適化をすることによる性能向上を確認する。また今回はソフトウェアとしての iSCSI ターゲットを使用した。今度は代わりにハードウェアターゲットを置き、iSCSI 性能がどのように変化するかを調べていく。

参考文献

- 1) 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, 電子情報通信学会論文誌 Vol.J87-D-I, No.2, pp.216-231, 2004 年 2 月
- 2) 千島 望, 山口 実靖, 小口 正人: VPN 複数経路接続を用いた iSCSI ストレージアクセス性能評価, DICOMO2007, 3H-4, 2007 年

- 3) 比嘉 玲華, 岡廻 隆生, 山口 実靖, 小口 正人: Linux カーネルのログ解析によるパースト的 iSCSI 遠隔ストレージアクセス時のパケット送信に関する考察 DICOMO2009
 - 4) 富士通 IP アクセスルータ GeoStream Si-R シリーズ GeoStream Si-R570,
<http://fenics.fujitsu.com/products/sir/sir570/index.html>
 - 5) 比嘉 玲華, 松原 幸助, 岡廻 隆生, 山口 実靖, 小口 正人: パケット解析と輻輳ウィンドウ解析による遠隔 iSCSI アクセスの断続的パケット送出に関する考察, DEIM2009, E1-1, 2009 年
-