社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ARC－157　(25)
2004－HPC－ 97　(25)
2004／3／3

IPsecを利用したiSCSIストレージアクセス時のTCPパケット転送の解析

神坂 紀久子 †　　山口 実靖 ‡　　小口 正人 †

近年，ストレージ統合技術である SAN (Storage Area Network) として，Ethernet と TCP/IP を用いて構築する iSCSI プロトコルに期待が高まっている．iSCSI プロトコルを利用した IP ネットワークを介するリモートストレージアクセスには，セキュリティ対策が重要な課題である．その手段として，IP プロトコルレベルで暗号化を行う IPsec が利用できる．しかし，IPsec の暗号化処理によってホストの CPU 負荷を増大させ，スループットを低下させる可能性があるため，TCP 層の振る舞いを詳しく調べる必要がある．本稿では，iSCSI プロトコルと IPsec を利用したシーケンシャルリードアクセスの性能評価を行い，TCP パケット転送の詳細を可視化し，解析する．

Analysis of TCP Packet Transfer on iSCSI Storage Access using IPsec

Kikuko Kamisaka†　　Saneyasu Yamaguchi‡　　Masato Oguchi†

Nowadays, the iSCSI protocol based on TCP/IP and Ethernet is becoming more important as a realization of consolidation techniques using storage area network (SAN). Securety is critical issue for the iSCSI protocol, in which remote storage is accessed over the IP network. Although IPsec encryption on IP layer can be applied in the iSCSI protocol, using IPsec increases CPU load and decreases throughput of the network. In this paper, we evaluated sequential read access performance when iSCSI protocol and IPsec are used. The behavior of TCP layer is analyzed by visualizing TCP packet transfer.

1　はじめに

ブロードバンドネットワークの普及や計算機の性能向上により，取り扱うデータ量が急激に増大し，大容量のデータを格納するために必要となる管理コストが増えてきた．そのため，現在ではデータを効率的に格納するストレージ統合技術，SAN (Storage Area Network) として，Fibre Channel とよばれる専用ネットワークを用いた FC-SAN が広く普及している．しかし，Fibre Channel はコストが高いことなどから，Ethernet と TCP/IP を用いて構築する IP-SAN の技術として，2003 年に IETF に承認された iSCSI プロトコルに期待が高まっている．

iSCSI プロトコルでは一般的に不安定で信頼性の低い場合がある IP ネットワークを介してストレージ

† お茶の水女子大学大学院 人間文化研究科
Graduate School of Humanities and Sciences, Ochanomizu University
‡ 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo

アクセスを行うため，セキュリティを強化してデータを保護することが重要となる．そのセキュリティ対策として，IP プロトコルレベルで暗号化を行う IPsec と併用して用いることが多い．しかし，IPsec 使用により，暗号化処理がホストの CPU 負荷を増大させ，スループットを低下させる可能性がある．IP-SAN に関しては [1][2][3] のような研究が発表されており，これらはシステム全体の性能評価などを行っているが，階層構造を持つ IP-SAN の性能向上のためには，重要な階層である TCP 層の振る舞いを詳細に調べる必要がある．

そこで本稿では，iSCSI プロトコルと IPsec を利用したストレージアクセスの性能評価を行い，TCP 層に着目して TCP パケット転送の詳細を可視化し，解析する．

2 研究背景

2.1 SAN

従来は，ストレージとして，サーバに直接 1 対 1 接続されたストレージデバイスを利用する DAS (Direct Attached Storage) が使われてきた．しかし，接続距離制限などの拡張性の問題やストレージ管理に膨大なコストがかかることから，複数のサーバと複数のストレージ間を接続できる，高速な専用のネットワークとして SAN(Storage Area Network) が登場してきた．

SAN では，異種間サーバでのデータの共用，一元管理や長距離接続による非常災害対策も可能であるという利点がある．そのため現在では，Fibre Channel とよばれる専用回線を用いた FC-SAN が広く普及している．

2.2 iSCSI

最近では，FC-SAN に代わる次世代 SAN の技術として，IP-SAN が登場してきた．IP-SAN は専用回線ではなく，IP ネットワークを利用してストレージアクセスを行う．IP-SAN の有力な候補である iSCSI プロトコルが 2003 年に IETF に承認されたことから，さらに期待が高まっている [4][5]．

iSCSI では，サーバの役割をするイニシエータとストレージであるターゲットの間で，SCSI プロトコルを TCP/IP プロトコルにカプセル化する．そのため，通常の SCSI アクセスと同様の操作で遠隔ストレージアクセスを実現できる．Fibre Channel と比べ，接続距離の限界がなく，既存の IP ネットワークとのシームレスな統合が行えるため，相互運用性が高い．また，運用技術に精通した技術者が多く，ハードウェアなどのコストも抑えられる上，Ethernet の性能向上の速さからも，既存の FC-SAN に取って代わる技術として今後の需要が期待できる．

2.3 IPsec

IP-SAN である iSCSI プロトコルはインターネットを利用するため，セキュリティ対策として IP ネットワークにおいて広く利用される IPsec を利用することが可能である．

IPsec は IP パケットの認証と暗号化を行う規格であり，ホストから送信されるあらゆる通信を IP 層レベルで透過的に暗号化できるため，アプリケーションの種類によらず安全な通信が可能である．暗号化には共有鍵暗号方式，アルゴリズムとしては対称ブロック暗号である DES (Data Encryption Standard), 3DES を利用している．

iSCSI ネットワークでは，一般にギガビットイーサネット上で大容量のデータ転送を行うため，通常の小さい帯域ではあまり問題にならない TCP プロトコル処理が，ホストの CPU 負荷に影響する場合がある．その上で，安全性を高めるために IPsec を利用すると，暗号化や復号化処理がホストの CPU 負荷に大きく影響を及ぼす可能性がある．ストレージ技術にとって，パフォーマンスは大きな課題といえるが，一方，安全性を高めることも必要不可欠である．パフォーマンスとセキュリティにはトレードオフの関係があり，iSCSI ストレージアクセスにおいては，双方に適切な対策を考えるために，iSCSI ネットワークで IPsec を用いた場合の詳細な解析を行うことが重要である．

2.4 プロトコルスタック

iSCSI，IPsec は図 1 に示す階層構造になっており，iSCSI は SCSI 層と TCP 層の間に位置する．また IPsec は，一般的に TCP 層と IP 層の間に位置する．

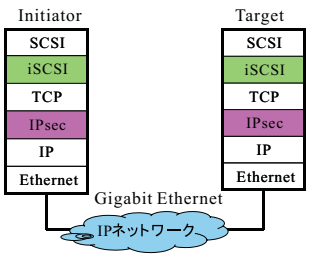


図 1：プロトコルスタック

アプリケーションから送られるカプセル化された SCSI パケットは，TCP 層に送られ，IPsec と IP 層で暗号化される．その後，通常の IP パケットとして IP ネットワークに送り出される．相手側のホ

ストは，ネットワークから IP パケットを受け取り，これらの階層の逆をたどって SCSI パケットとして上位層に送られる．

3 性能評価実験

3.1 実験概要

iSCSI ネットワークにおいて，IPsec を用いてセキュリティを強化した場合のリモートストレージアクセスの性能を測定した．我々の実験では，iSCSI イニシエータと iSCSI ターゲットを Gigabit Ethernet で接続し，ターゲットの raw デバイスに対して，ストレージアプリケーションにおいて多く用いられるシーケンシャルリードアクセスの性能測定を行った．表 1 に我々の性能評価実験で使用したイニシエータとターゲット計算機の実験環境を示す．iSCSI イニシエータとターゲットの実装は，ニューハンプシャー大学 InterOperability Lab[6] が提供している reference implementation を用いた．IPsec の実装には Linux において広く利用されている FreeS/WAN[7] を用いた (表 2)．

表 1：実験環境：使用計算機

OS	initiator:Linux 2.4.18-3 target:Linux 2.4.18-3
CPU	Intel Xeon 2.4GHz
Main Memory	512MB DDR SDRAM
HDD	36GB SCSI HD
NIC	Intel PRO/1000XT Server Adapter on PCI-X (64bit, 100MHz)

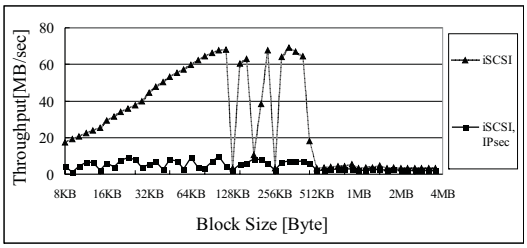
表 2：実験環境：使用実装

iSCSI	UNH IOL reference implementation ver. 3 on iSCSI Draft 18
IPsec	FreeS/WAN ver. 2.01

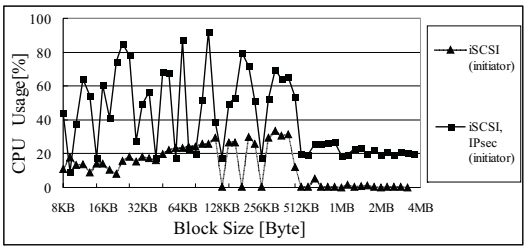
3.2 考察

図 2 に IPsec を利用した場合と利用しない場合の iSCSI ストレージアクセスにおけるスループット及び CPU 使用率の測定結果を示す．IPsec を利用しない場合には，ブロックサイズが大きくなるにしたがって性能が向上するのに対し，IPsec を利用した場合には，ブロックサイズに関わらず著しく性能が低く，CPU の負荷が高くなっていることがわかった．

また，我々の iSCSI ストレージアクセスの性能評価実験では，IPsec を利用しないケースにおいて，ブロックサイズ約 128KB，256KB 付近，約 512KB 以上でスループットが特徴的に低くなっており，CPU の使用量についても同様のブロックサイズ付近で CPU 負荷が低下していることがわかった．この性能劣化は，我々が使用した iSCSI 実装によってパケットが分割され，微小パケットが生じてしまうことが原因によるものである．この性能劣化についての詳細は次の章でパケットの振る舞いとともに説明する．



(a)：スループット測定結果



(b)：CPU 使用率測定結果

図 2：iSCSI, IPSec 利用時のシーケンシャルリードアクセスの性能

4 TCP パケット転送の解析

性能評価実験結果から，IPsec 利用時の性能劣化と本稿の実験に特徴的な iSCSI ストレージアクセスの性能劣化の現象を分析するために，TCP 層でのパケット転送の振る舞いを解析した [8]．

4.1 Read シーケンス

SCSI プロトコルの Read シーケンスは以下のようになっており [9]，本稿の実験では，100 個の Read システムコールを発行している．

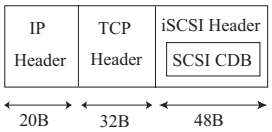
(1) イニシエータからターゲットへ，“SCSI Read Command” である iSCSI PDU (Protocol Data Unit) を送信する．

(2) 送られた “SCSI Read Command” に対し，ターゲットからイニシエータへ，“SCSI Data-in” の iSCSI PDU を送信する．

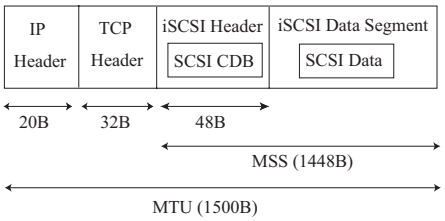
(3) 転送されるデータが 1 つのパケットのおさまらなかったときには，データを分割して送信する．

(4) ターゲットからイニシエータへ，“SCSI Re-

sponse” である iSCSI PDU を送信する．



(a)：SCSI Read Command，SCSI Response



(b)：SCSI Data-in

図 3：iSCSI PDU の構造

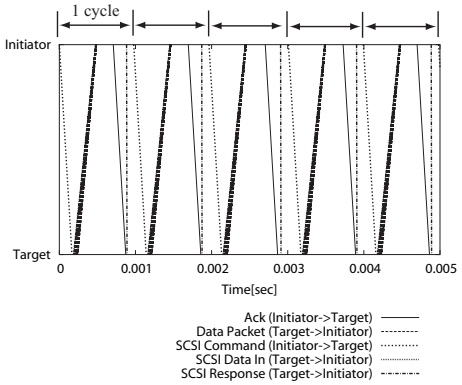
図 3 は，本稿の実験で IPsec を利用しない場合の iSCSI PDU 構造の例である．IP ヘッダは 20 バイト，TCP ヘッダは 32 バイトであり，“SCSI Read Command” と “SCSI Response” は，iSCSI ヘッダのみの構成となっている．“SCSI Data-in” では，SCSI CDB (Command Descriptor Block) の入った iSCSI ヘッダと iSCSI データセグメントで構成されており，SCSI データの転送が 1 つにおさまらなかった場合は，MSS (Maximum Segment Size) 毎に分割されて，イニシエータに送信される．

4.2 暗号化処理による性能劣化

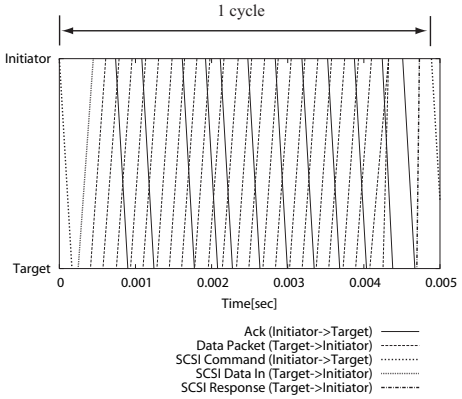
図 4 は，iSCSI ネットワークにおけるシーケンシャルリードアクセス時の，TCP パケット転送を可視化したものであり，ブロックサイズは 32KB である．図 4 では，イニシエータから Read Command が発行され，ターゲットからのすべてのデータが送信しおわるまでを 1 サイクルとして示している．以下の式は，1 サイクルに発行された SCSI Read Command で要求している送信データの大きさである．512Byte は SCSI CDB のチャンクサイズである．

- $64 \times 512\text{Byte} = 32\text{KB}$

図 4 (a) の IPsec を用いない場合には，ブロックサイズ 32KB 分のデータを転送し終わるまでの 1 サイクルが約 0.001 秒であり，ほぼ同時刻にイニシエータにまとめてデータを送信している．一方，IPsec を利用する場合には，図 4 (b) のように 1 パケットずつイニシエータにデータを送信しており，1 サイクルに約 0.005 秒かかっている．IPsec 層における各パケットの暗号化処理に時間がかかることよって，パケットの送受信間隔がまばらになり，これが性能低下の原因となっていることがわかった．



(a)：IPsec を利用しない場合



(b)：IPsec を利用した場合

図 4：ブロックサイズ 32KB での iSCSI ネットワークにおけるシーケンシャルリードアクセスのパケット転送

4.3 Nagle アルゴリズムと遅延確認応答による性能劣化

図 5 は，3 章の性能評価実験結果について，ブロックサイズ 1MB での特徴的な性能劣化が現れているときの TCP パケット転送を可視化したものであり，図 6 は図 5 (a) の部分的な拡大図である．

以下の式は，すべてのデータを送信しおわるまでの 1 サイクルにおいて，イニシエータからターゲットへの CDB が入った SCSI Read Command で要

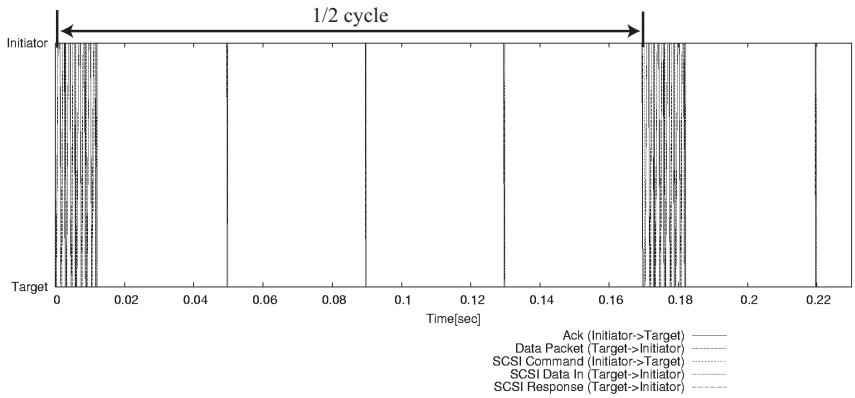
求している送信データの大きさである．

- 255 × 512Byte = 127.5KB　が 4 組
- 1 × 512Byte = 0.5KB　が 4 組
- 255 × 512Byte = 127.5KB　が 4 組
- 1 × 512Byte = 0.5KB　が 4 組

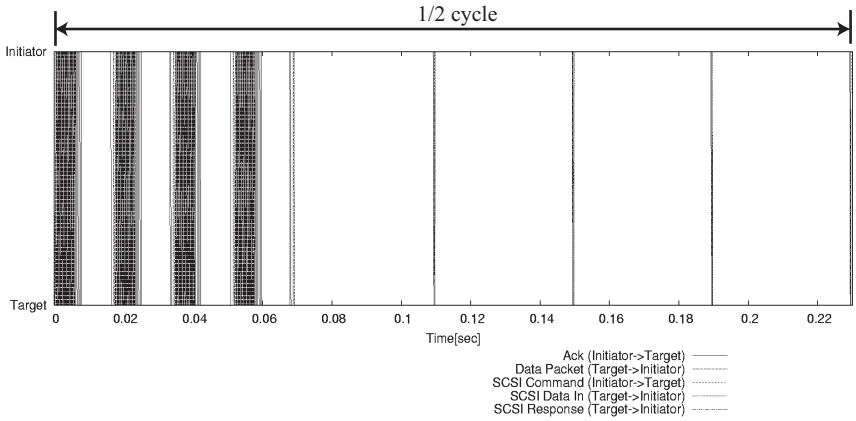
このような SCSI Read Command で要求している 0.5KB の微小 iSCSI PDU が発行される原因は，本稿で使用した iSCSI の実装に依存するものである．図 5，図 6 では，0.5KB の SCSI Data-in の微小 iSCSI PDU が，ターゲットからイニシエータへ送信されることにより，ターゲットは，確認応答 (以下，Ack) が返るまで次のデータパケットを送らないという現象が起きていることがわかった．これは，Nagle アルゴリズムとよばれるものであり，データ部分の小さい微小パケットを混雑の多いネットワーク上で転送する場合に，輻輳状態にしてしまわないためのアルゴリズムである．具体的には，TCP コネクションは確認応答されていない 1 つの非常に小さな未処理のセグメントだけを持つことができるというもので，Ack が受け取られるまで，それ以外の小さなデータなどのセグメントは送られない．

このアルゴリズムは今回使用した Linux 実装では有効になっており，そのため微小パケットを送った後，Ack が返るまでは次のデータパケットを送らず，それが遅延確認応答の原因となっている．

遅延確認応答とは，通常 TCP はただ 1 つのパケットを受け取っても，直ちに Ack を返すことはなく，Ack を遅らせることで，次のパケットの Ack，もしくはその Ack と同じ方向に向かうデータと組み合わ



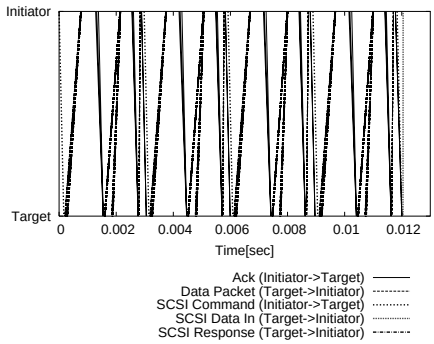
(a)：IPsec を利用しない場合



(b)：IPsec を利用した場合

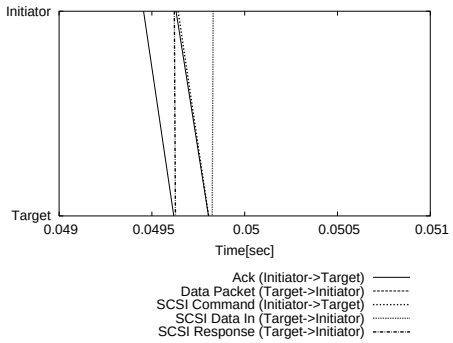
図 5：ブロックサイズ 1MB での iSCSI ネットワークにおけるシーケンシャルリードアクセスのパケット転送

せて送り，効率化を図ろうとするものである．次のパケットなどが到着しない場合にはタイムアウトとなり Ack が送られるが，その時間は実装により異なる．ブロックサイズが 32KB の場合は，微小パケットは送られておらず，このような現象はおきていない．それに対し，ブロックサイズが 1MB の場合には，0.5KB という微小パケットをターゲットが受け取っても，すぐには Ack を返さず，タイムアウトが起きるまで待つことにより，著しくスループットが落ちるという現象が起きていることがわかった．



(a)：パケットが連続して送信されている箇所

(0〜0.012[sec] 間)



(b)：微小パケットが送信されている箇所

(0.049〜0.051[sec] 間)

図 6：ブロックサイズ 1MB での iSCSI ネットワークにおけるシーケンシャルリードアクセスのパケット転送拡大図

5　まとめ

iSCSI ネットワークにおいては，セキュリティ問題が大きな課題となるが，パフォーマンスも重要である．本研究では，iSCSI プロトコルと暗号化を行う IPsec を利用した場合のシーケンシャルリードアクセスの性能を測定し，その振る舞いを解析した結果，TCP 層で起きている性能劣化に関わるいくつかの要因が明らかになった．今後は，IPsec 使用時の性能劣化への対処を検討する．

謝辞

本研究は一部，文部科学省科学研究費特定領域研究課題番号 13224014 によるものである．

参考文献

[1] W. T. Ng, B. Hillyer, E. Shriver, E. Gabber, and B. Ozden, “Obtaining High Performance for Storage Outsourcing,” *Proc. FAST 2002, USENIX Conference on File and Storage Technologies*, pp. 145-158, Jan. 2002.

[2] P. Sarkar and K. Voruganti, “IP Storage: The Challenge Ahead”, *Proc. Tenth NASA Goddard Conference on Mass Storage Systems and Technologies*, pp. 35-42, Apr. 2002.

[3] P. Sarkar, S. Uttamchandani, and K. Voruganti, “Storage over IP: When Does Hardware Support help?”, *Proc. FAST 2003, USENIX Conference on File and Storage Technologies*, pp. 231-244, Jan. 2002.

[4] IETF IP Storage (ips) Charter, http://www.ietf.org/html.charters/ips-charter.html

[5] Storage Networking Industry Association, http://www.snia.org/

[6] InterOperability Lab in the University of New Hampshire, http://www.iol.uhn.edu/consortiums/iscsi/

[7] FreeS/WAN Project, http://www.freeswan.org/

[8] 山口 実靖，小口 正人，喜連川 優: 「iSCSI 解析システムの構築と高遅延環境におけるシーケンシャルアクセスの性能向上に関する考察」，電子情報通信学会論文誌, D–I, 2004 年 2 月号掲載予定

[9] iSCSI Draft, http://www.ietf.org/internet-drafts/draft-ietf-ips-iscsi-20.txt