

自律ディスクを用いた VPN 接続広域分散ネットワークストレージの評価

浅田 菜那[†] 小口 正人[†]

[†] お茶の水女子大学

〒 112-8610 東京都文京区大塚 2-1-1

E-mail: [†]nana-asa@ogl.is.ocha.ac.jp, ^{††}oguchi@computer.org

あらまし RAID は高機能ストレージとして知られている。これに対し、分散ディレクトリやチェインド・デクラスタリングなどの技術を用い、負荷分散、故障対策、障害回復などにおいて RAID より高機能化された自律ディスクの研究が行われている。本研究においては、VPN を利用することにより自律ディスクを広域分散ネットワークに適用することを検討する。広域分散環境において自律ディスクの振舞を解析することにより、自律ディスクの特性を把握し、より良いネットワークストレージを実現するための方式を検討する。

キーワード 自律ディスク, ストレージ管理, ネットワークストレージ, RAID, VPN, 広域ネットワーク

An Evaluation of Autonomous Disks in VPN-Connected Distributed Wide Area Network

Nana ASADA[†] and Masato OGUCHI[†]

[†] Ochanomizu University Otsuka 2-1-1, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [†]nana-asa@ogl.is.ocha.ac.jp, ^{††}oguchi@computer.org

Abstract RAID is well-known as high functional storage. On the other hand, as regards load balance control, measures of failure and recovering, Autonomous disks which are functioned more highly than RAID are developed. In this research, we have examined that autonomous disks are applied to distributed wide area network by using VPN. In addition, in a wide area network, we have analyzed the feature of autonomous disks and examined the method to realize a better network storage system.

Key words autonomous disks, storage management, network storage, RAID, VPN, wide area network

1. はじめに

近年のインターネット技術の発展により、人々が個々に管理、所有し利用する情報量は爆発的に増加し続けている。これにともなって、それらの情報を格納し保存するストレージの増設やシステム管理コストの増加も問題となっている。さらに大量の情報を使いこなす為には情報を効率よく管理し、必要な場面で素早くかつ適切に取り出すことができるようなシステムが必要である。

RAID (Redundant Arrays of Inexpensive (もしくは Independent) Disks) は、1987 年カリフォルニア大学バークリー校の David A.Patterson、Garth Gibson、Randy Katz の 3 人によって提唱された。今日、RAID はミラーリングやパリティを用いた高機能ストレージのひとつとしてよく知られ、様々な場面で使われている。RAID は、複数台のハードディスクを組み合わせること

で冗長性の高い仮想的な一台のハードディスクとして運用する技術である。ディスクアレイの代表的な実装形態で、主に信頼性の向上をねらって用いられることが多い。

これに対し、負荷分散、故障対策、障害回復などにおいて、RAID よりさらに高機能化された自律ディスクの研究が行われている [1]。これはストレージの管理をストレージ上で自律的に行うシステムである。

本研究においては、VPN を利用することにより自律ディスクを広域分散ネットワークに適用させ、その状況における自律ディスクの振舞を解析する。これにより広域分散環境における自律ディスクの特性を把握し、より優れたネットワークストレージを実現するための手法を検討する。

本稿は以下のように構成される。まず第 2 章と第 3 章では研究背景としてそれぞれ自律ディスクと広域ネットワークについて記述する。次に第 4 章で本研究における実験システムや実験

方法について述べる。さらに第5章で自律ディスクの性能評価結果を示し、第6章で考察を述べ、最後に第7章でまとめと今後の課題を述べる。

2. 自律ディスク

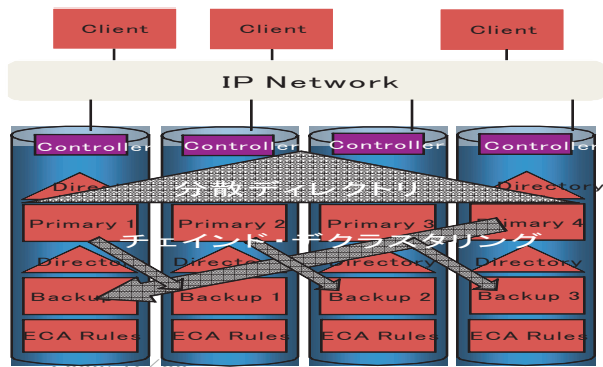


図1 自律ディスク構成図

図1は自律ディスクの構成例である。自律ディスクは、高機能化したストレージ装置をネットワークで結合してストレージクラスタを構成し、クラスタ内の各ストレージ装置間で必要に応じて通信し合い、クラスタ全体の管理を行うシステムである。具体的には、これまでサーバ上で管理者が行ってきた分散データ配置制御、容量分散、負荷分散制御、トランザクション制御、サービス品質制御、セキュリティ制御などを自律的におこなうことで、管理者の労力を大幅に削減するとともに、ストレージ装置の付加価値を高めることができる。

自律ディスクにより構築したストレージクラスタでは集中管理を行わず、どのストレージ装置にアクセスしても目的とするデータに高速に到着できるような分散ディレクトリを使用し、高いスケーラビリティを備えている。

冗長データ配置法としてはチェインド・デクラスタリングが採用されている。これは、プライマリ領域とバックアップ領域を隣接するストレージ装置にチェーンして配置するしくみである。仮に一台が失われた場合でも残りのノードが自動的にチェインド・デクラスタリングを組みなおし動作する。これらの仕組みにより、自律ディスクはRAIDより高い信頼性を持ち、障害発生時や回復時にも安定したデータ提供を行うことができる。

3. VPN 広域分散ネットワーク

VPNの構成例を図2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ることができることである。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自律ディスクの振舞を解析するため、VPNルータを使用する。

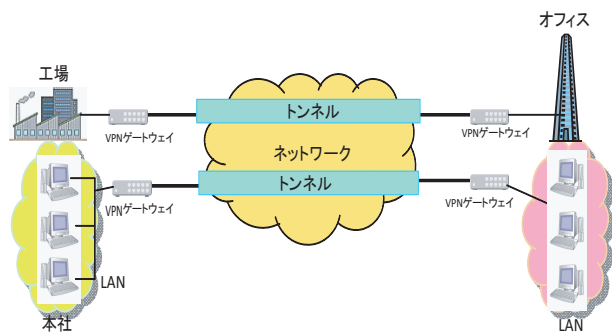


図2 VPN使用例

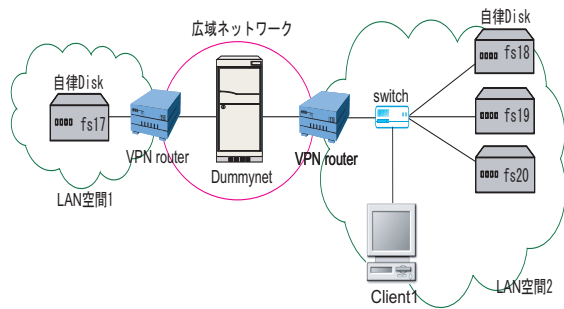


図3 自律ディスクが4台そろった構成

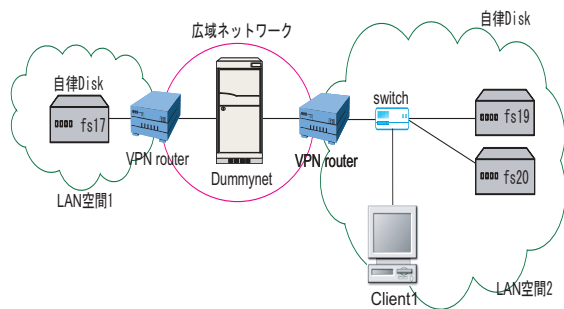


図4 fs18を除いた構成

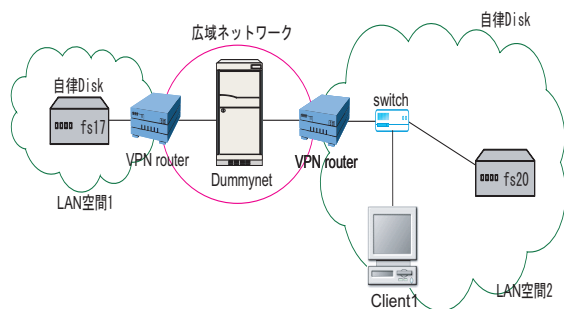


図5 fs18,fs19を除いた構成

4. 実験概要

本研究では、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1台の自律ディスクfs17を、もう一方のLAN空間には3台の自律ディスクfs18-20を接続させた。計4台の自律ディスクがfs17→fs18,fs18→fs19,fs19→fs20,fs20→fs17という形でチェインドデクラスタリングにより接続されている。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。また fs18-20 をつなぐスイッチングハブに性能評価を行うクライアント端末を接続した。自律ディスクとクライアント間には SMB (Server Message Block) プロトコルで接続されており、fs20 に自律ディスク向けに改変した Samba サーバを立ち上げてクライアントからのアクセスを受け付ける。自律ディスクおよびクライアント端末のスペックを下記に示す。また VPN ルータには Fujitsu Si-R180 を用いた [3]。クライアント端末上では、Windows からのアクセス性能評価として IOMeter を、Linux からのアクセス性能評価として自作ツールの HDDbench を使用した。IOMeter においては Sequential Read (連続読み込み) と Sequential Write (連続書き込み) のそれぞれについてスループット、レスポンスタイム、CPU 使用率を、HDDbench においては Sequential Write のスループットとレスポンスタイムを測定した。

実験の手法としては、まず遅延装置を使用して、0-64ms の遅延環境を作る。次に下記のように実験環境を 3 パターン構築し、それぞれにおいて測定した。

(1) fs17-20 が 4 台そろった構成 (図 3)

(2) fs18 を除いた構成 (図 4)

(3) fs18, fs19 を除いた構成 (図 5)

はじめに 4 台を立ち上げた後、図 4 の構成では fs18 を、図 5 の構成では fs18 と fs19 を物理的に切り離す。このようにすることで、1 つまたは 2 つのノードが突然失われた場合でも残りのノードが速やかにチェインド・デクラスタリングを構成しなおし正常に作動するかを調べることができる。

以上の実験を、クライアント端末を 1 台に設定した構成と 2 台に設定した構成それぞれについて測定する。

自律ディスク

CPU : Transmeta 社製 1GHz

OS : Linux CentOS 5.0(kernel 2.6.18-8)

HDD : 2.5 インチ 100GB

Windows クライアント端末

OS : Windows XP

CPU : Intel(R) Pentium(R) M processor 1400MHz

Main Memory : 256MB RAM

Hard Disk : 14.8GB

Linux クライアント端末

CPU : Fedora CORE

kernel : 2.6.10-iscsi

OS : Intel PentiumIII 640MHz

Main Memory : 640MB

Hard Disk : 10GB

5. 自律ディスク性能測定

5.1 クライアント端末 1 台の構成

図 3-5 の実験環境において Windows から測定し比較した連続読み込みにおけるスループットのグラフを図 6、レスポンス

タイムのグラフを図 7、Linux と Windows において測定し OS による比較をした連続書き込みにおけるスループットのグラフを図 8、レスポンスタイムのグラフを図 9、Windows から測定した CPU 使用率のグラフを図 10 に示す。

Windows をクライアント端末に設定した状態での、連続読み込みのスループットのグラフでは、3 本のグラフは、遅延時間が大きくなるにつれて減少していくが、それらはあまり重なり合わず、多少不安定な結果となった (図 6)。また同実験環境における、連続読み込みのレスポンスタイムも 3 本のグラフは異なる動きをみせた (図 7)。これはクライアント端末からのアクセスのタイミングにより、自律ディスクの状態が多少異なっているためと考えられる。

端末一台を設定し連続書き込みにおけるスループットの OS による比較をしたところ、OS によって上下に大きく分かれるものの、それぞれの OS について 3 本のグラフは重なり合い、同程度の性能であった (図 8)。この結果より自律ディスクにおける、Windows を端末に設置した場合の連続書き込みは、連続読み込みの場合より安定した結果を出力することがわかった。また Linux のグラフは Windows のグラフよりも常に上に存在していることから、自律ディスクは Linux をクライアント端末に設定した場合においてより良い性能になる。

一方、レスポンスタイムの比較をしたところ、スループットのような OS による差は出なかった。6 本のグラフは全て同程度の数値、形状となった (図 9)。

CPU 使用率についてはどのグラフも片道遅延時間が増すにつれて減少している。これは遅延時間が増すと通信の待ち時間が長くなり、自律ディスクへのアクセス処理が CPU を占有している時間の割合が少なくなるためだと考えられる。

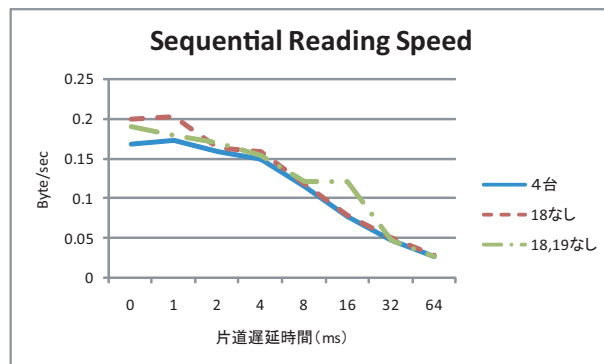


図 6 連続読み込みのスループット【Windows/端末 1 台】

5.2 クライアント端末 2 台の構成

次にクライアント端末 1 と端末 2 を図 11 のように LAN 空間 2 に設置し同時に測定することで、さらに負荷をかけた場合の自律ディスクの性能を調べた。

Windows をクライアント端末として設置した測定では、自律ディスク 4 台が揃った構成、fs18 を除いた構成の 2 種類で測定し、Linux をクライアント端末として設置した測定では、端末 1 台の実験と同様に 3 種類の構成で測定した。なお Windows においても Linux においても、2 台の端末からはほぼ同程度の結

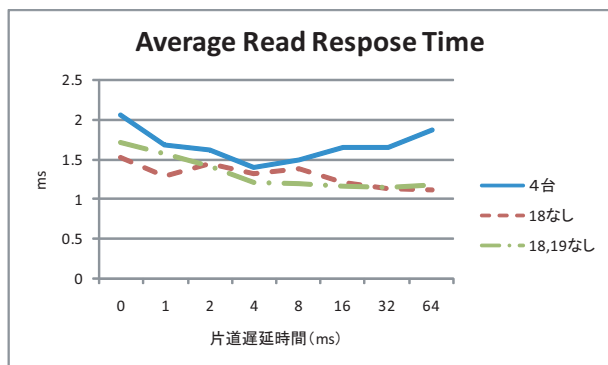


図7 連続読み込みのレスポンスタイム【Windows/端末 1 台】

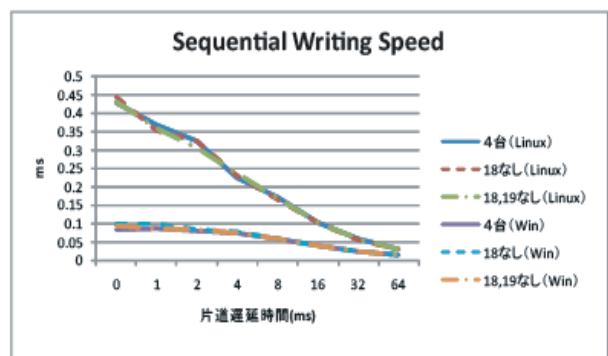


図8 連続書き込みのスループット【端末 1 台ずつ】

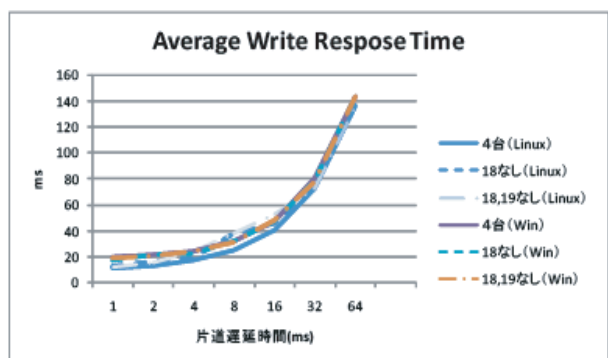


図9 連続書き込みのレスポンスタイム【端末 1 台ずつ】

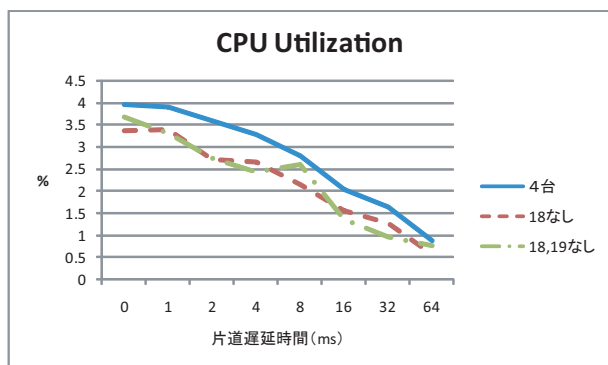


図10 CPU 使用率【Windows/1 台】

果が得られたため、端末 1 における結果だけを掲載する。

図 11 の実験環境において Windows から測定し比較した連続読み込みにおけるスループットのグラフを図 12、レスポンス

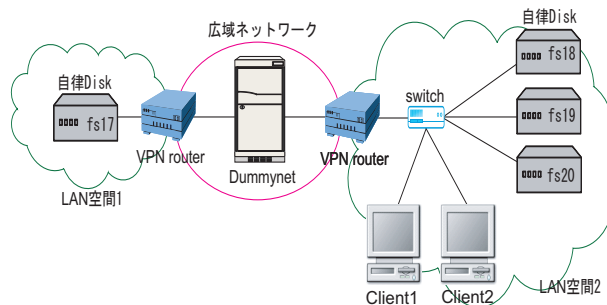


図 11 クライアント端末を 2 台設置した構成

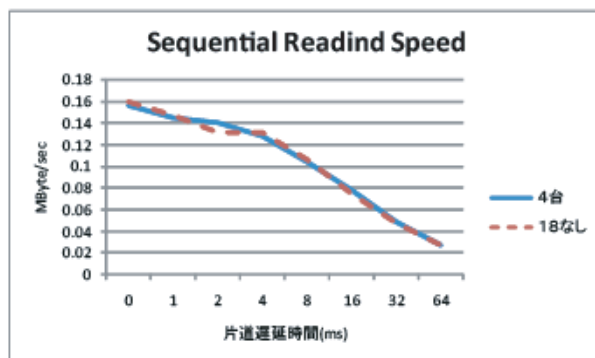


図 12 連続読み込みのスループット【Windows/2 台】

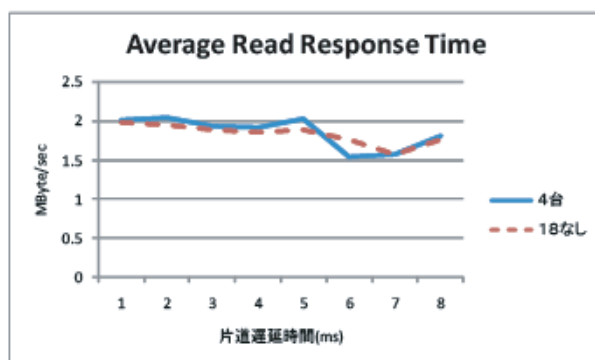


図 13 連続読み込みのレスポンスタイム【Windows/2 台】

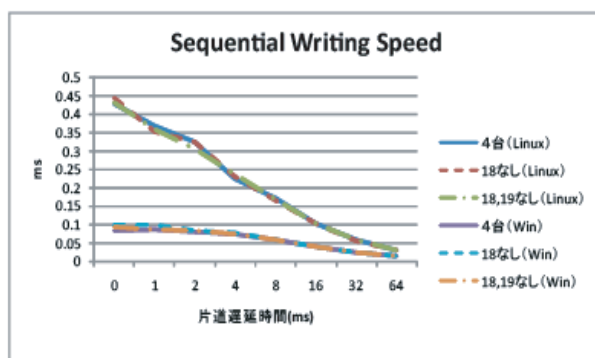


図 14 連続書き込みのスループット【端末 2 台ずつ】

タイムのグラフを図 13、Linux と Windows において測定し OS による比較をした連続書き込みにおけるスループットのグラフを図 14、レスポンスタイムのグラフを図 15、Windows から測定した CPU 使用率のグラフを図 16 に示す。

Windows をクライアント端末として 2 台設置し、連続読み込

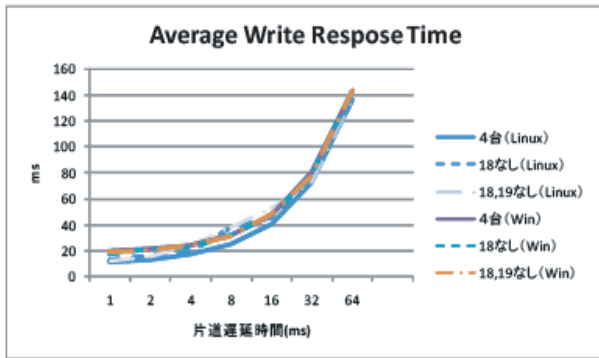


図 15 連続書き込みのレスポンスタイム【端末 2 台ずつ】

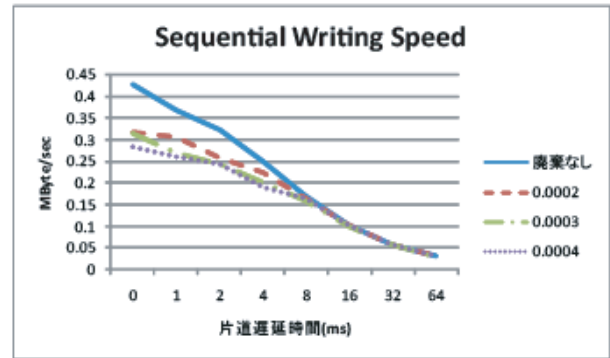


図 17 スループットの比較【Linux/1 台】

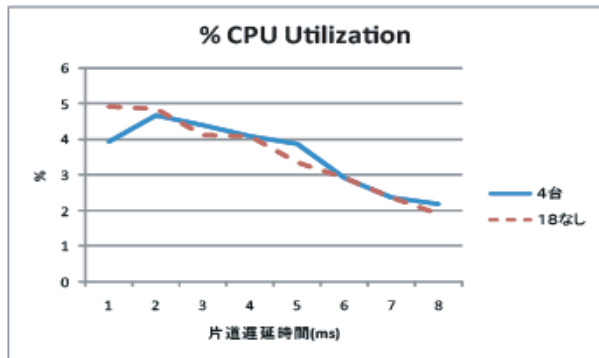


図 16 CPU 使用率【Windows/2 台】

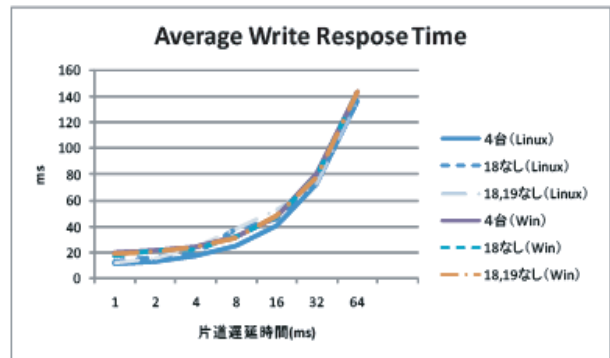


図 18 レスポンスタイムの比較【Linux/1 台】

みを行った結果のスループットはやはり片道遅延時間が大きくなるにつれて下がった (図 12)。またレスポンスタイムの 2 本のグラフは重なり合わず、多少不安定な結果となった (図 13)。

連続書き込みにおけるスループットの OS による比較では、端末 1 台の構成における結果と同様、Linux のグラフは Windows のグラフよりも常に上に位置し OS による差がでるが (図 14)、レスポンスタイムにおいてはグラフは全て同じ様な形状になり OS によって性能が変わることはなかった (図 15)。

また Windows における CPU 使用率においても片道遅延時間が上がるにつれて下がっていく結果となった (図 16)。

5.3 パケット廃棄が起こるネットワーク上の評価

ネットワークの状態が様々に変わった場合のシステムへの影響の評価を調べるため、一定の率でパケット廃棄が起こるような現実的な広域分散ネットワーク上で自律ディスクの評価を行った。Linux を 1 台クライアント端末として設置し、自律ディスクが 4 台揃った構成における連続書き込みのスループットとレスポンスタイムを測定した。なおパケット廃棄率 0.02 % の場合、0.03 % の場合、0.04 % 場合において、廃棄率を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か解析した。スループットの比較の結果を図 17、レスポンスタイムの比較の結果を図 18 に示す。

図 17 より、片道遅延時間が短い時には、パケット廃棄率の違いがスループットにある程度大きな影響を及ぼしていることがわかった。しかし片道遅延時間が長くなるにつれてスループットの差は縮まり、8ms 以降はパケット廃棄率によらず、どの状態においても同程度の数値となった。これはパケット廃棄によ

るオーバーヘッドの差違が、遅延時間の増大による性能低下により隠蔽されているからではないかと考えられる。

一方レスポンスタイムにおいては 6 本のグラフがほぼ同じ様な形状をしたことから、どの遅延環境においてもスループット程影響を受けなかったことが判明した。

6. 考 察

自律ディスクは広域分散ネットワーク環境でも正常に動作することがわかった。4 台のうちの 1 台または 2 台がダウンしても残りのノードでチェインド・デクラスタリングを構成しなおし、機能することを確かめることが出来た。また、3 つの異なる構成におけるスループットやレスポンスタイムの結果が大きく変わるということではなかった。このことより、本実験で自律ディスクのチェーンが直接つながっている fs17,20 がそろっている場合において fs18,19 はそれほど大きく関わらないのではないかと考えることが出来る。理由としてはデフォルトでクライアント端末は fs20 宛に SMB でアクセスし、そのデータが自律ディスクの仕組みによって fs17 にバックアップとして書き込まれているためであろう。自律ディスクは片道遅延時間が大きくなるにつれてスループットが低下し、レスポンスタイムが長くなっているため、性能は劣化していることが確認できた。

クライアント端末を 1 台設置した場合でも 2 台設置した場合でも、スループットの比較により Linux を端末に設定したほうが自律ディスクの性能は高くなった。

またパケット廃棄率を設定した場合、片道時間 8ms まではパケット廃棄率が高いほどスループットが低下し性能が劣化して

いるが、8ms を超えるとパケット廃棄率にはあまり関係しないことがわかった。

7. まとめと今後の課題

本研究では VPN 広域分散環境で自律ディスクの振舞を解析、評価した。その結果、このような環境でも動作し、ノードが欠けてしまった場合でも残りのノードで構成を組みなおし正常に機能することが確認できた。今後はより詳細な解析、評価を行うとともに、例えば 4 台とも同じ LAN 空間につなぎ、VPN でつないだ反対側の LAN 空間からクライアントでアクセスするなど、今までとは異なるネットワーク環境において評価する。

謝 辞

本研究は一部、独立行政法人科学技術振興機構戦略的創造研究推進事業 CREST によるものである。

文 献

- [1] 横田 治夫: 機能追加によるストレージの高付加価値化のアプローチ, MR2006-79, 2007 年 3 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] 富士通 IP アクセスルータ Si-R シリーズ Si-R180, <http://fenics.fujitsu.com/products/sir/sir180/>