

# 業務開始時刻に基づくディスク読み書き量の制限による バックアップとユーザのファイル転送時間の両立

高橋 風太<sup>1</sup> 飯島 貴政<sup>2</sup> 串田 高幸<sup>1</sup>

**概要:** 映像制作会社ではユーザである社員がファイルサーバに動画ファイルを転送し、ファイルサーバがバックアップサーバに100GBの業務データを転送する事例がある。課題は、ユーザがファイル転送をする際にバックアップが行われていると、ファイルサーバのハードディスク(以下ディスク)が毎秒読み書きができる量に対してプロセスが使用している読み書きができる量の割合(以下帯域幅の使用率)が100%となり、ユーザのファイル転送時間がバックアップをしていない場合と比較して増加することである。本提案では、ユーザのファイル転送時間の増加を抑制するため、企業の業務開始時刻に基づきネットワークの帯域幅に上限を設け、ディスクの帯域幅の使用を抑制する。評価では、100GBの動画ファイルをユーザからファイルサーバ及びファイルサーバからバックアップサーバへ転送し、提案方式を用いない場合と用いる場合の、バックアップ中のユーザのファイル転送時間を比較する。結果、バックアップを73分7秒後に完了させなければいけない場合、バックアップは44分1秒で完了し、ユーザのファイル転送時間の増加を4.5%抑制した。

## 1. はじめに

### 背景

アプリケーション層では、データをバックアップと呼ばれる追加のコピーに複製する。これにより、データが失われた場合に元の情報を復元できるようにする [1]。バックアップは、様々な損害からデータを保護する。損害の例として、アプリケーションやオペレーティングシステムの破損、ストレージの破損、ネットワークの破損、停電、ユーザによるエラー、自然災害が挙げられる [2-4]。

バックアップの利用例として、東映アニメーションを挙げる。ここでは映像、テレビ、キャラクター商品の開発事業が展開され、これらのコンテンツの制作のための画像ファイルや動画ファイルを業務に用いるデータ(以下業務データ)として共有する。社員間でのデータの共有や管理のため、ファイルサーバが使用される。ファイルサーバのデータを保護するため、ファイルサーバの業務データをバックアップされる。複数のコンテンツ制作を同時進行することがあるため、オンプレミスのシステムは24時間稼働している\*1。システムの継続を前提とする企業や組織のその他

の例として、デルタ航空会社が運営する OIS(Operational Information System) が存在する。OISは、企業または組織の日常業務を継続的にサポートする大規模な分散システムである。OISのパフォーマンス要件として、99.99%の可用性が挙げられている [5]。

コンピューティングシステムは指数関数的な速度でより多くの並列処理を活用できるため処理速度が速いが、ストレージインフラストラクチャのパフォーマンスは大幅に低い速度である。例は、スーパーコンピュータである Mira である。Miraのコンピューティングノードでは、ネットワーク帯域幅が1536GB/sで転送できる一方で、ストレージシステムは250GB/sの実行速度である。この状況はI/O輻輳を引き起こし、アプリケーションのパフォーマンスを低下させる [6-8]。バックアップ中にシステムにアクセスできる場合、システムパフォーマンスが影響を受ける可能性を指摘している [9]。これはバックアッププロセスが用いるディスク帯域幅が通常のサービスプロセスと競合するためである。

### 課題

課題は、ファイルサーバからバックアップサーバへの業務データのバックアップ中にユーザがファイルサーバへ動

アニメーション」のアニメ制作を下支えするツールとは”,  
<https://techtaraget.itmedia.co.jp/tt/news/1704/12/news07.html>  
(2022/6/15)

<sup>1</sup> 東京工科大学コンピュータサイエンス学部  
〒192-0982 東京都八王子市片倉町1404-1

<sup>2</sup> 東京工科大学大学院バイオ・情報メディアコンピュータサイエンス専攻  
〒192-0982 東京都八王子市片倉町1404-1

\*1 Tech Target Japan, ”社外とも簡単データ共有、「東映

画ファイルを転送すると、バックアップをしていない場合と比較して、ユーザのファイル転送時間が増加することである。これは、利用者のストレスや、業務進行の妨げに繋がる\*2。原因は、動画ファイルの転送によるファイルサーバのハードディスクの書き込みとバックアップによるハードディスクの読み込みが衝突し帯域幅の使用率が100%となり、時間当たりのハードディスクへの書き込み量が減少することである。業務の妨げとは、修正や確認作業の遅延、伝達の遅れ、進捗の遅れである。映像制作会社の映像制作工程において、ファイル転送時間の長さが課題となる\*3。図1に動画ファイルの転送を示し、図2に、課題となる動画ファイルの転送を示す。

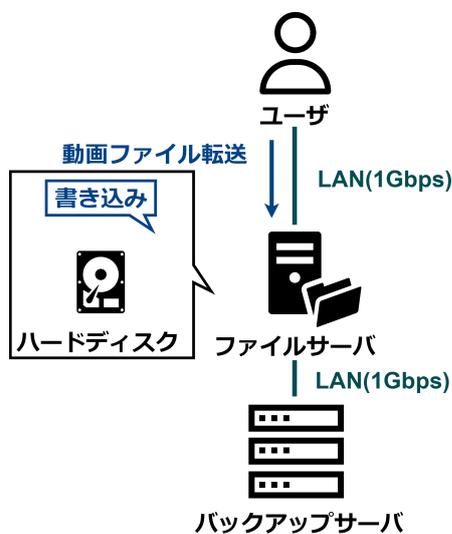


図1 動画ファイルの転送

図1では、ユーザがファイルサーバに動画ファイルを転送し、ファイルサーバのハードディスクが書き込みを行っている。LANの帯域幅はユーザ・ファイルサーバ間及びファイルサーバ・バックアップサーバ間において1Gbpsである。

図2では図1と異なり、ユーザの転送だけでなく動画ファイルをファイルサーバからバックアップサーバへバックアップしている。ファイルサーバのハードディスクはユーザからの動画ファイルを書き込み、バックアップのために業務データを読み取っており、読み取りと書き込みが衝突している。ハードディスクの帯域幅使用率が100%になると読み書きの速度が低下し、ユーザのファイル転送時間がバックアップをしていない場合と比較して増加する。

\*2 NTT Communications, "業務効率に影響するオンラインストレージの転送速度", <https://www.ntt.com/business/services/application/online-storage/bst-sh/lp/article-online-storage-transfer-speed.html>(2022/11/14)

\*3 SKEED, "導入事例", <https://skeed.jp/case/>(2022/11/14)

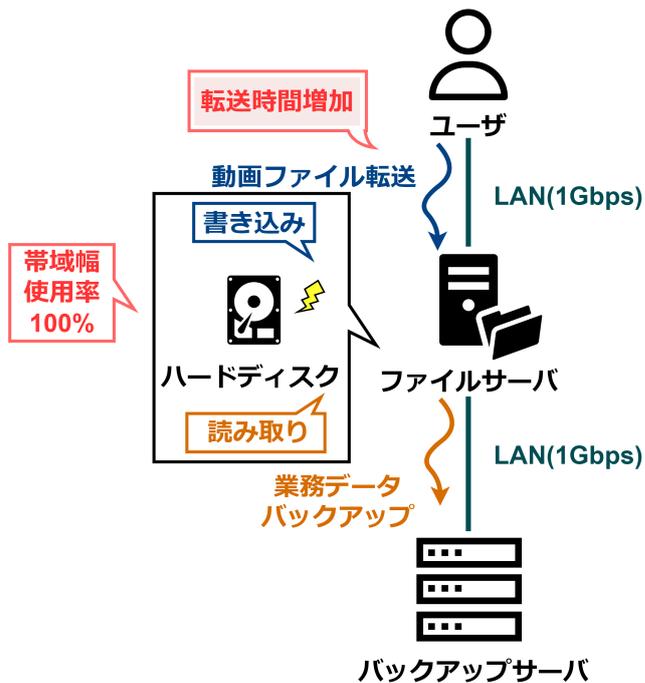


図2 課題となる動画ファイルの転送

### 課題の基礎実験

バックアップによりユーザのファイル転送時間が増加することを示すため、課題の基礎実験を行う。サーバとして、ハイパーバイザのESXiから作成される仮想マシン(以下VM)を用いる。以下に、課題の基礎実験に用いるVMの構成情報を示す。vCPUは、VMに割り当てるCPUを表す。

- VM構成情報(ユーザ)  
OS Ubuntu-20.04.2  
vCPU 2コア  
RAM 8GB  
HDD 120GB
- VM構成情報(ファイルサーバ)  
OS Ubuntu-20.04.2  
vCPU 2コア  
RAM 8GB  
HDD 300GB
- VM構成情報(バックアップサーバ)  
OS Ubuntu-20.04.2  
vCPU 2コア  
RAM 8GB  
HDD 120GB

動画ファイルの転送には、rsyncコマンドを用いる。動画ファイルの容量は、映像制作会社において100GBの動画ファイルを転送する例があるため、ユーザ、ファイルサーバ

ともに 100GB とする\*4。実験では、100GB になるように 10GB の動画ファイルを 10 個複製した。これを 1 つのディレクトリに入れ、動画ファイル、業務データとして用いる。VM は最大通信速度 1Gbps の LAN ケーブルにより接続されている。これは、ネットワークがボトルネックになることを防ぐためである。VM は Network Attached Storage(以下 NAS) に動画ファイルを保存している。各 VM に用いられる NAS は、それぞれ 10Gbps の LAN ケーブルにより接続されている。これにより、VM と NAS の間がボトルネックとなることを防ぐ。

実験では 2 つのファイル転送の方法を比較し、ユーザからファイルサーバへのファイル転送の時間が、バックアップにより増加することを計測する。図 3 と図 4 に、2 つのファイル転送の方法の概要を示す。

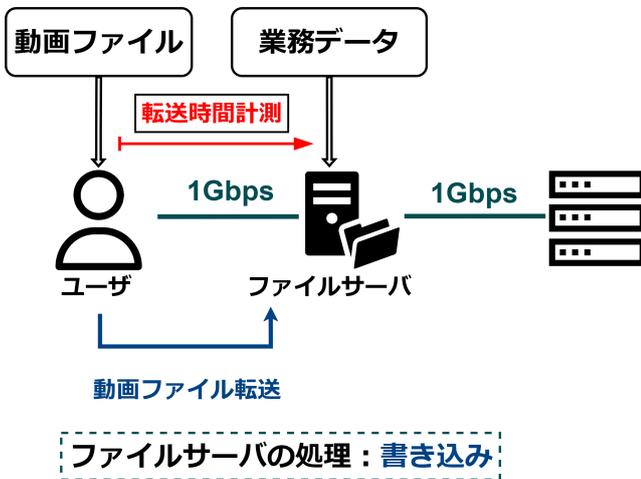


図 3 1つ目のファイル転送

1つ目のファイル転送の方法では、バックアップをせずユーザからファイルサーバへの動画ファイルの転送を行う。

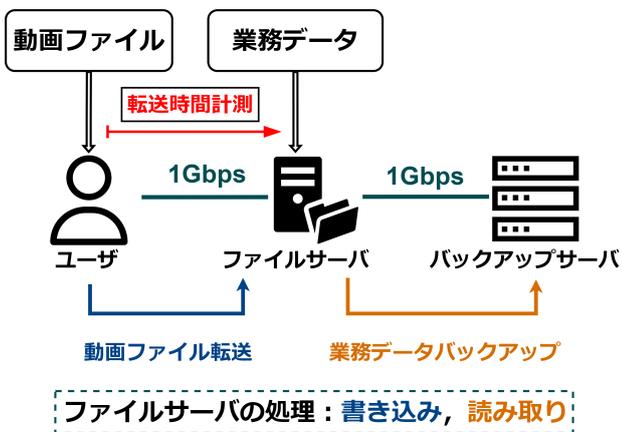


図 4 2つ目のファイル転送

\*4 DIRECT!EXTREME, "導入企業インタビュー映像制作プロダクション様" <https://www.directextreme.com/case/interview2.html> (2022/11/14)

2つ目のファイル転送の方法では、バックアップ中にユーザからファイルサーバへの動画ファイルの転送を行う。図 5 に、2つの実験の書き込み量の差を示す。

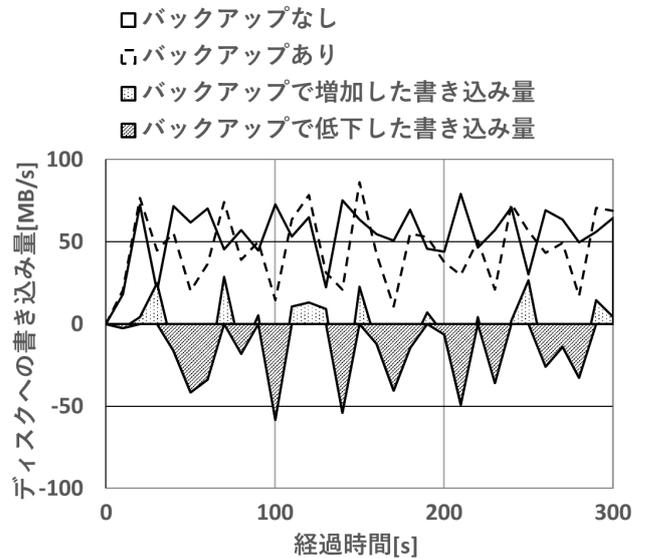


図 5 ファイルサーバのディスク書き込み量の差

書き込み量の計測は iostat コマンドで行い、1秒間に書き込んだセクタ数を取得する。1セクタには 512 バイト保存できるため、iostat で取得した値に 512 を掛けた値を書き込み量とする。縦軸はディスクへの書き込み量 [MB/s] を表し、横軸は時間 [s] を表す。青の点線はファイル転送を表す。これは、1つ目の実験で転送した動画ファイルの書き込み量である。オレンジの点線はファイル転送+バックアップを表す。これは、2つ目の実験で転送した動画ファイルの書き込み量である。黒の面は、1つ目の実験の書き込み量から2つ目の実験の書き込み量を引いたものを表す。バックアップにより、600秒までの書き込み量は合計 5166.75MB 少なくなっている。1秒当たりの書き込み量は、15.5%減少している。これらは、バックアップにより単位時間当たりの書き込み量が減少したことを示している。

次に、2つの実験を 5 回行った際のファイル転送時間の平均を計測する。図 6 に、2つの実験のファイル転送時間を示す。

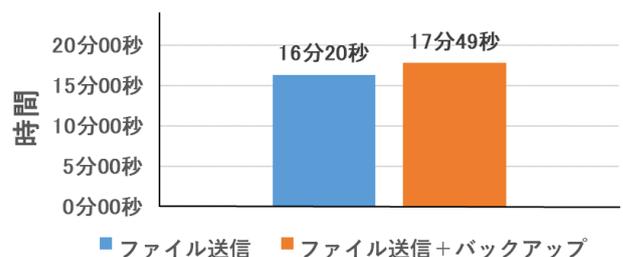


図 6 2つの実験のファイル転送時間

図 6 から、バックアップによりファイル転送時間が 1 分 29 秒増加している。この時間が課題である理由として、電話のやり取り中にファイルを転送する場合を想定する。電話口で 30 秒待たされると 400 人のうち 200 人以上がイライラを感じる\*5。したがって、映像制作会社において電話のやり取り中にファイルを転送する場合、バックアップにより通常より 30 秒以上待たされるとユーザである社員にストレスがかかることを課題とする。したがって、ファイル転送時間の増加を 30 秒短縮することを目標とする。

## 各章の概要

第 2 章では、関連研究について述べる。第 3 章では、本研究の提案方式とユースケースについて述べる。第 4 章では、提案方式の実装と評価のための実験方法について述べる。第 5 章では、提案方式の評価と分析について述べる。第 6 章では、本研究が残す議論について述べる。第 7 章では、本研究の成果とまとめについて述べる。

## 2. 関連研究

Deepavali Bhagwat らは、チャンクベースのファイルバックアップのためのスケーラブルな並列重複排除を提案している [10]。課題は、リアルタイムバックアップを例とするランダムな順序で到着するファイルの重複排除の処理に要する時間が、ディスクのスループットのボトルネックにより低くなることである。提案では、ファイルの類似性を利用し、ファイルごとのチャンク探索に 1 つのディスクアクセスのみを行っている。評価では、フラットチャンクインデックスに必要な RAM と提案手法に必要な RAM 使用量の比率は 86.56:1 であった。ディスクのボトルネックの原因は RAM 使用量がオーバーフローするとディスクにページングする必要があることであり、筆者らはボトルネックの問題を軽減している。しかし、この重複排除をユーザやファイルサーバが用いた際にユーザのファイル転送時間が短縮されることが示されていない。

Huan Chen らは、クラスターの集中ファイルサーバ向けの新しいヒントベースの I/O メカニズムを提案している [11]。課題は、複数の並列アプリケーションが同時に共有ストレージにアクセスするとファイルサーバの I/O パフォーマンスが低下することである。原因は、異なるクライアントからの I/O 要求が干渉するためである。評価では、NFS(Network File System) と比較して読み取りパフォーマンスが約 2 倍向上している。しかし、本研究のユースケースでは 1 ユーザによるファイル転送で課題が発生するため、Huan Chen らの研究には課題が残る。

\*5 CITIZEN, "ビジネスパーソンの「待ち時間」意識", <https://www.citizen.co.jp/research/time/20180531/02.html>

## 3. 提案

### 提案方式

課題の要約を述べる。課題は、ファイルサーバのバックアップ中にユーザがファイルサーバへ動画ファイルを転送することで動画ファイルの転送開始から完了までに要する時間が増加し、業務の進行が妨げられることであった。原因は、ファイルサーバのバックアップによる読み取りと動画ファイルの受信による書き込みが同時に発生しディスクの帯域幅使用率が 100% になることで、単位時間当たりの書き込み量が減少するためである。

図 7 に、提案方式の目的を示す。提案では、①ファイルサーバのバックアップによる読み取りを制限することでディスクの帯域幅使用率を抑制し、②単位時間当たりの書き込み量の減少を防止することを目的とする。

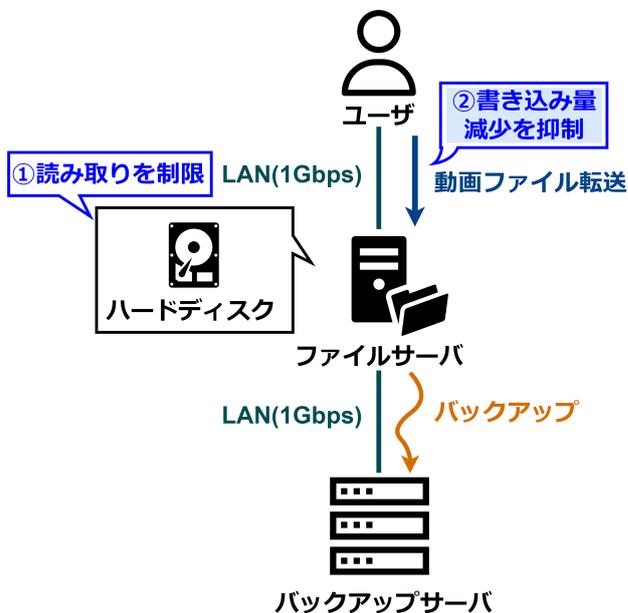


図 7 提案方式の目的

企業の業務開始時刻に基づきバックアップに用いるネットワークの帯域幅に上限を設けることでハードディスクの帯域幅の使用を抑制し、ユーザのファイル転送時間の増加を抑制する。図 8 に、提案方式の流れを示す。

図 8 の①から⑥までの手順を以下に示す。

- ① 提案システム内の制限判定システムはユーザからのファイルの受信を検知する。
- ② ディスクの帯域幅使用率を取得する。
- ③ ディスク使用率が 100%かどうかを判定する。
- ④ バックアップ予定ファイルの合計容量を取得する。
- ⑤ ファイル受信を検知し帯域幅使用率が 100%に達した

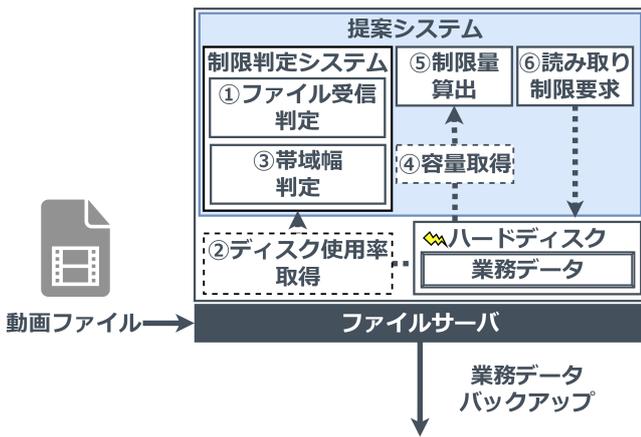


図 8 提案方式の流れ

場合、バックアップのためのプロセスのネットワークの帯域幅の上限を決定する。

⑥ バックアップのためのプロセスのディスク読み取り量の制限を実行する。

⑤ 制限量の算出では、④バックアップ予定ファイルの合計容量とバックアップ期限から、期限内に合うバックアップ速度を求める。バックアップ開始は深夜 0 時で、バックアップ期限は企業の業務開始時刻を想定する。バックアップの実行プロセスが利用するネットワーク帯域幅の制限量決定の条件を以下に示す。1 秒あたりの転送速度の上限を  $S$  とする。バックアップ予定の業務データの容量を  $C$  とする。バックアップ期限を  $L$  とする。

$$S = \frac{C}{L} \quad (1)$$

図 9 に、制限量決定の流れを示す。

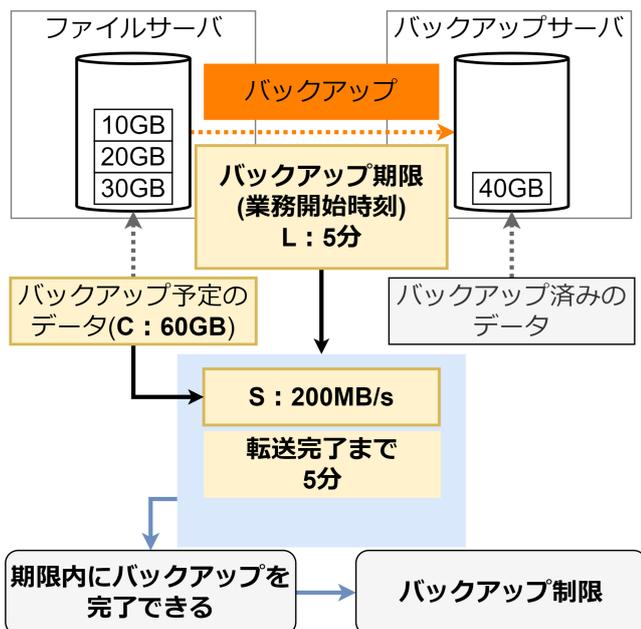


図 9 制限量決定の流れ

バックアップ期限が 5 分で、バックアップ予定のデータが 60GB だった場合を想定する。この場合、1 秒に 200MB 転送すると転送完了までの時間は 5 分となる。この時間であれば、期限内にバックアップを完了できる。期限内にバックアップが完了する場合はネットワークの帯域幅の上限を 200MB/s で制限し、期限内にバックアップが完了しない場合はバックアップの制限をしない。

ファイルサーバがユーザへファイルを送信する場合、バックアップを中断する。ファイル送信の終了を検知すると、バックアップを再開する。理由は、バックアップはユーザの利用が少ない夜間に行われるものであり、ユーザへのファイル送信は、緊急性のあるものとしたためである。例として、動画制作の納期が近く、いち早く動画ファイルを受信する必要がある場合が挙げられる。

### 提案方式の基礎実験

バックアップの実行プロセスが利用するネットワーク帯域幅の制限量決定の条件が、実際にバックアップが期限内に合うかを実験する。実験環境は課題の基礎実験と同様である。図 10 に、提案方式の基礎実験のファイル転送を示す。ファイルサーバからバックアップサーバへ業務デー

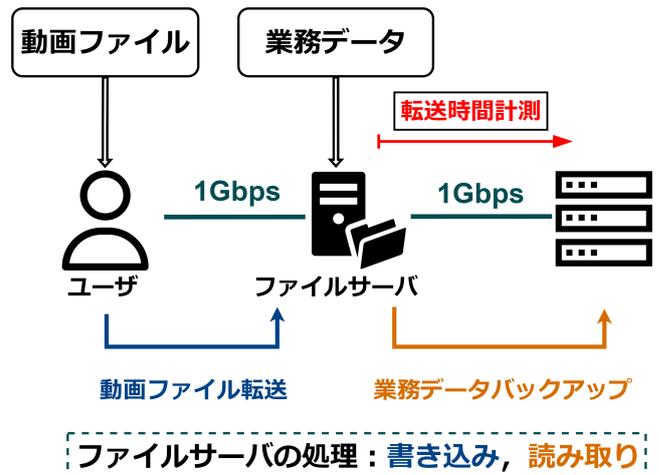


図 10 提案方式の基礎実験のファイル転送

データをバックアップし、その間にユーザからファイルサーバへ動画ファイルを転送する。バックアップには期限を設定し、ネットワーク帯域幅の制限量決定の条件より速度を算出し、その速度を上限してファイル転送する。バックアップ期限は課題のファイル転送時間が 17 分 49 秒であるため、それより長い時間で設定する。今回はバックアップ期限を 20 分から 10 分ずつ増やしていき、70 分までを 1 回ずつ計測した。図 11 に、各バックアップ期限で制限した場合のバックアップ時間を示す。

図 11 は、バックアップ期限が長くなるほど実際のバックアップ時間も長くなる傾向にある。しかし、全ての期限においてバックアップ期限を過ぎている。したがって、ネッ

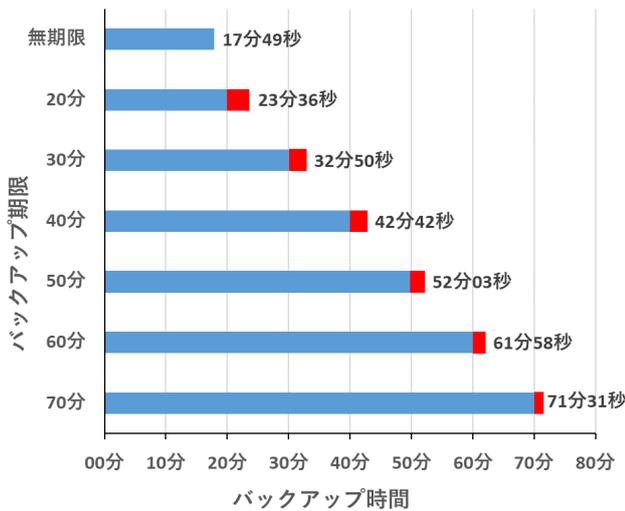


図 11 各バックアップ期限で制限した場合のバックアップ時間

ネットワーク帯域幅の制限量決定の条件を変更し、バックアップ時間をバックアップ期限に収める必要がある。

### 制限量算出方法の補正

バックアップ時間をバックアップ期限に収めるため、制限量算出方法に補正を加える。ファイル転送をして出た1秒当たりの転送速度の最大値と最小値の差(以下範囲)を補正の候補とする。

この補正を  $R$  とした場合の制限量の算出方法は以下となる。1秒あたりの転送速度の上限は  $S$  である。バックアップ予定の業務データの容量は  $C$  である。バックアップ期限は  $L$  である。

$$S = \frac{C}{L} + R \quad (2)$$

この補正はネットワークの帯域幅に上限を設けない場合の転送速度の範囲から求めるが、実際にはネットワークの帯域幅に上限を設けている。したがって、ネットワークの帯域幅に上限を設けた場合の転送速度の範囲が、上限を設けない場合の範囲より大きくなる可能性がある。そのため、ネットワークの帯域幅に上限を設ける場合の範囲を算出する。

図 11 より、バックアップ期限が 30 分の場合は 2 分 50 秒超過、バックアップ期限が 20 分の場合は 3 分 36 秒超過と、バックアップ期限を短くするほど実際のバックアップ時間が期限を超過している。

バックアップ期限が短い方がバックアップ時間がバックアップ期限を超過するならば、ネットワークの帯域幅の上限を可能な限り大きくすることでバックアップ期限の超過の最大値が求まる。したがって、ネットワークの帯域幅に上限を設けずにファイル転送をして出た転送速度の最大値をネットワークの帯域幅の上限の値とする。この上限値でファイル転送をして出た 1 秒当たりの転送速度の最大値と

最小値の差(以下範囲)を補正のもう一つの候補とする。

ネットワークの帯域幅に上限を設けない場合と設ける場合の範囲のうち、より範囲が大きい方を補正  $R$  とする。

補正の算出方法の手順を以下に述べる。

- ① ユーザからファイルサーバへのファイル転送とファイルサーバからバックアップサーバへのファイル転送を同時に行う。この時、転送するファイルはディスクの読み書きによりディスクの帯域幅が 100% となるものとする。
- ② ファイルサーバからバックアップサーバへのファイル転送を行う。
- ③ ①と②から出た転送速度の最大値をネットワークの帯域幅の上限とし、①と②のファイル転送を再度行う。
- ④ ①と②の転送速度の範囲と③の転送速度の範囲のうち、より範囲が大きい方を補正  $R$  とする。

ファイル転送速度は使用しているハードウェアとソフトウェアの性能に依存するため、この補正の算出方法は本実験環境以外にも適用可能である。

### ユースケース・シナリオ

本研究の提案方式は、映像制作会社で適用することを想定している。バックアップは毎日 0 時に開始すると想定する [9]。図 12 に、提案システム適用の想定図を示す。

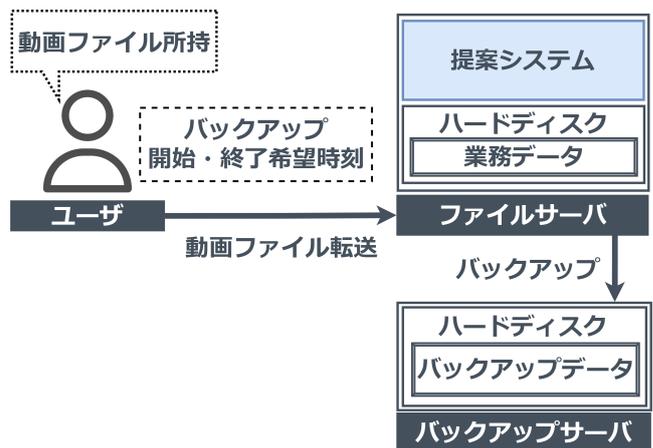


図 12 提案システム適用の想定図

社員はユーザとして動画ファイルを所持している。ユーザはファイルサーバを用いて動画ファイルを共有・管理する。ファイルサーバ内の業務データはバックアップサーバへバックアップされる。バックアップの開始時刻と終了希望時刻を提案システムに入力することで、終了希望時刻をバックアップ期限として提案システムが適用される。

## 4. 実装

提案方式を実現するためのソフトウェアをファイルサーバ内に配置する。バックアップの実行プロセスとして、rsyncを用いる。図13に、実装機能のフローチャートを示す。

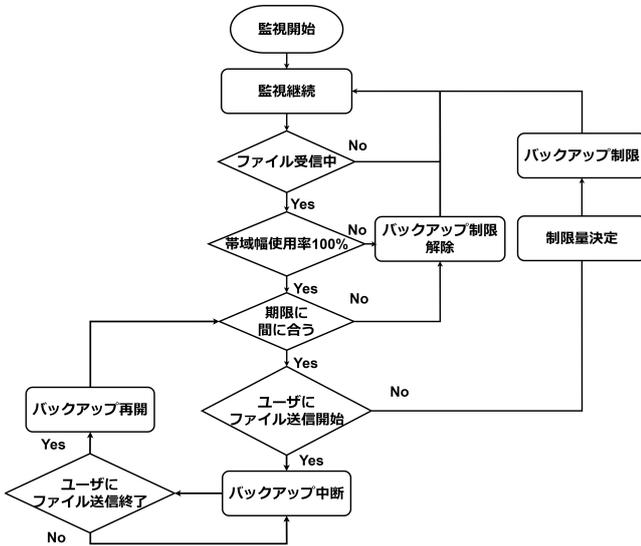


図 13 実装機能のフローチャート

ディスク帯域幅の使用率を監視するため、iostat コマンドを用いる。監視間隔は1秒とする。これは、監視によるディスク帯域幅の使用が0.4%に収まる間隔である。ユーザーが動画ファイルをファイルサーバへ転送すると、提案システムはファイルサーバがファイルを受信したことを検知する。ファイルを受信している場合、ディスクの帯域幅使用率が100%に達しているかを確認する。ディスクの帯域幅使用率が100%に達している場合、バックアップがバックアップ期限に間に合うかを確認する。バックアップが期限に間に合うと判定された場合、ファイルサーバがユーザーにファイルを送信しているかを確認する。ファイルがユーザーに送信していない場合はネットワーク帯域幅の制限量を決定し、制限を実行する。ファイルをユーザーに送信している場合はバックアップをファイル送信が終了するまで中断する。ファイルサーバがユーザーからファイルを受信している場合、帯域幅の使用率が100%に達していない場合、バックアップがバックアップ期限までに間に合わない場合、バックアップの制限を解除する。

## 5. 実験と分析

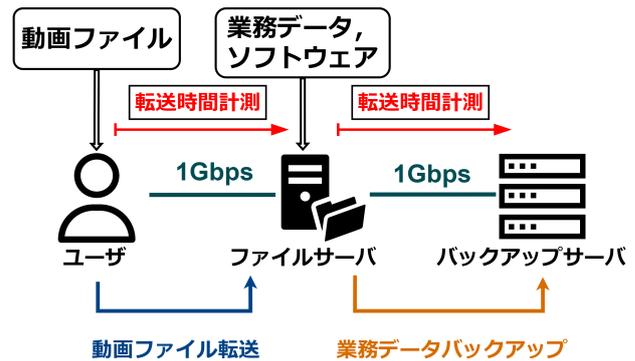
評価実験では、3種類の実験を行う。提案方式の補正の算出、バックアップ期限が超過しないかの確認、ユーザーのファイル転送時間が抑制されているかの確認である。各実験を5回繰り返し、計測した値は5回の平均値で評価する。

提案方式の補正の算出について述べる。バックアップ

予定の容量は、基礎実験と同様100GBである。100GBは10GBの動画ファイル10個で構成されている。実験は5回繰り返すため、10GBの動画ファイルをネットワークの帯域幅を制限しない場合と制限する場合でそれぞれ50回ファイル転送する。ファイル1つあたりの転送速度の平均を取得し、その中から転送速度の最大値と最小値を算出する。

### 実験環境

実験はファイルサーバ内にソフトウェアを配置し、それ以外の環境は基礎実験と同様の環境で行う。図14に、評価実験の環境を示す。



ファイルサーバの処理: 読み書き, ソフトウェア実行

図 14 評価実験の環境

ファイルサーバが業務データをバックアップサーバへバックアップしている間に、ユーザーがファイルサーバへ動画ファイルを転送する。ユーザーが動画ファイル転送に要する時間と、ファイルサーバが業務データのバックアップに要する時間を計測する。

### 実験結果と分析

提案方式の補正の算出を図15に示す。ネットワークの

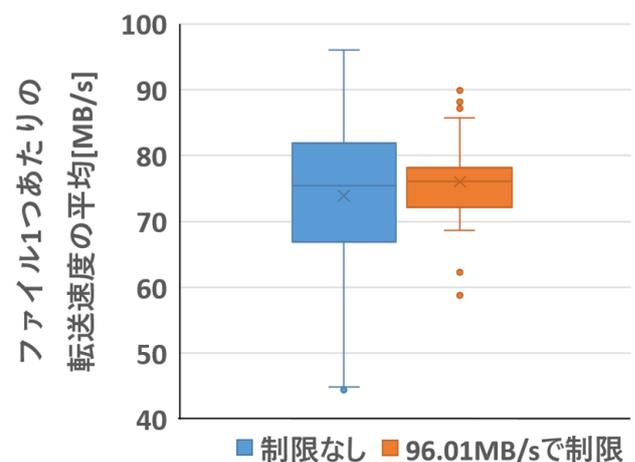


図 15 補正の算出

帯域幅を制限しない場合のファイル転送では、最大値が96.01MB/s、最小値が44.47MB/sとなった。この差である51.54MB/sが補正值の1つ目の候補となる。ネットワークの帯域幅を96.01MB/sで制限した場合のファイル転送では、最大値が89.94MB/s、最小値が58.84MB/sとなった。この差である31.1MB/sが補正值の2つ目の候補となる。この2つの候補の内さ大きい方が補正值となるため、補正值は51.54MB/sである。補正值の決定手順から出た転送速度の最大値と補正值の差/バックアップ予定の容量をバックアップ期限に加算し、これを補正時間とする。本実験環境における補正時間は、100GBのファイルを転送するため、33分7秒である。この値を期限に加算することで、バックアップ期限の超過を防ぐ。

バックアップ期限を評価する実験では、バックアップ期限が71分31秒後の場合を評価する。以下に理由を述べる。図11より、バックアップ期限が短いほど期限を超過する傾向があるため、極力バックアップ期限を短く設定する必要がある。しかし補正值51.54MB/sは最大値96.01MB/sを超えて補正できず、最大値と補正值の差である44.47MB/sから補正できる。したがって、この44.47MB/sを上限としてファイル転送をした場合が極力短くバックアップ期限を設定した場合である。理論上44.47MB/sは100GBのファイルを38分23秒で転送でき、これに補正時間を加算すると71分31秒となる。図16に、バックアップ期限が71分31秒後の場合のファイル転送時間を示す。

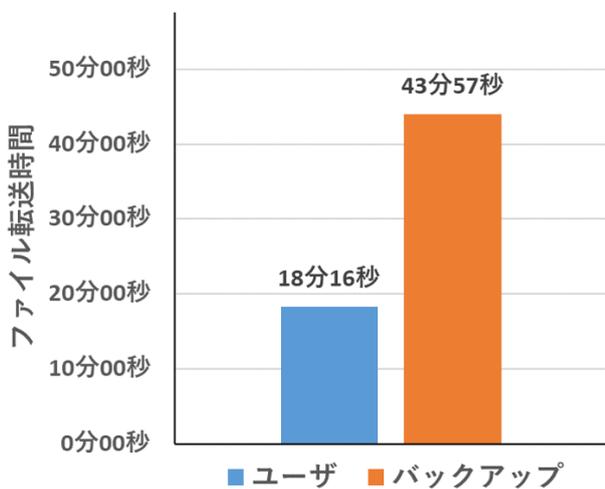


図16 バックアップ期限が71分31秒後の場合

バックアップに要する時間は43分57秒となった。これは、バックアップ期限より27分34秒早くバックアップしている。一方で、ユーザのファイル転送時間は、課題となるファイル転送時間である17分49秒より早くならない。原因は、バックアップ期限を超過させないために、バックアップにディスクの帯域幅の使用を優先していることである。バックアップに要する時間がバックアップ期限

と同時に終了する場合、よりユーザがディスクの帯域幅を使用でき、ユーザのファイル転送時間が早くなる。

ユーザのファイル転送時間を評価する実験では、バックアップ期限が20分+補正時間、30分+補正時間、40分+補正時間後の場合のユーザのファイル転送時間を計測する。本環境における補正時間は33分7秒である。図17に、各期限の場合のユーザのファイル転送時間を示す。

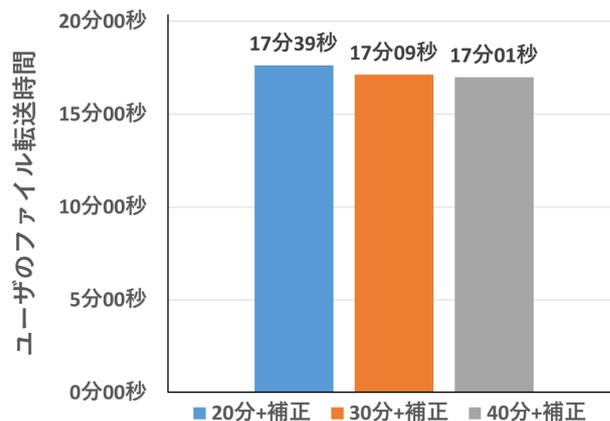


図17 各期限の場合のユーザのファイル転送時間

課題となるファイル転送時間は17分49秒であり、これを30秒短縮することを目標としていたため、バックアップ期限が30分+補正時間の場合と40分+補正時間後の場合に目標を達成している。最もファイル転送時間の短い40分+補正時間、つまりバックアップ期限が73分7秒後の場合、ユーザのファイル転送時間の増加を4.5%抑制している。

各期限の場合のバックアップに要する時間を図18に示す。

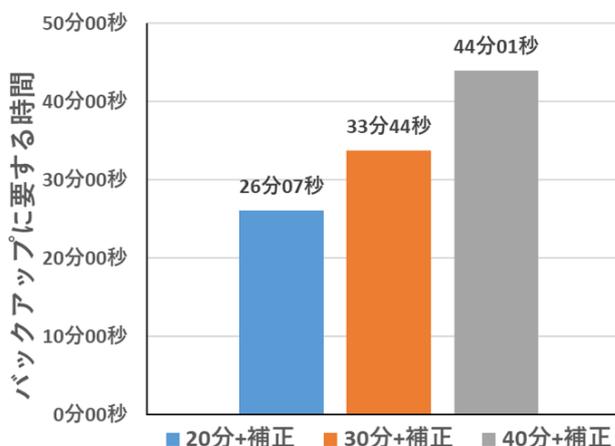


図18 各期限の場合のバックアップに要する時間

バックアップ期限に補正時間である33分7秒を加算することで、各バックアップ時間はバックアップ期限を超過していない。

実験の結果、バックアップを73分7秒後に完了させなければいけない場合、バックアップは44分1秒で完了し、ユーザのファイル転送時間の増加を4.5%抑制した。

## 6. 議論

本提案では、バックアップ期限内にバックアップを終えることができた。しかし、バックアップ期限まで余裕があるため、課題となるユーザのファイル転送時間の増加をより抑制できる余地がある。補正値を転送速度の範囲ではなく転送速度の平均値や中央値へ変更することが、よりバックアップ時間をバックアップ期限に近づける方法の検討となる。また、補正値決定の手順での実験回数を重ねることで、正確な補正値を決定することができる。

## 7. おわりに

課題は、バックアップ中にユーザがファイル転送をする時、ファイルサーバのディスクの帯域幅の使用率が100%となり、ユーザのファイル転送時間がバックアップをしていない場合と比較して増加することである。提案では、企業の業務開始時刻に基づきネットワークの帯域幅に上限を設けることでディスクの帯域幅の使用を抑制し、ユーザのファイル転送時間の増加を抑制する。評価では、100GBの動画ファイルの転送を行い、提案方式を用いない場合と用いている場合の、バックアップ中のユーザのファイル転送時間を比較した。結果、バックアップを73分7秒後に完了させなければいけない場合、バックアップは44分1秒で完了し、ユーザのファイル転送時間の増加を4.5%抑制した。

## 参考文献

- [1] Min, J., Yoon, D. and Won, Y.: Efficient Deduplication Techniques for Modern Backup Operation (2011).
- [2] Cheng, H., Ho, Y. H., Hua, K. A., Liu, D., Xie, F. and Tsaur, Y.-P.: A Service-Oriented Approach to Storage Backup, *2008 IEEE International Conference on Services Computing*, Vol. 2, pp. 413–421 (online), DOI: 10.1109/SCC.2008.132 (2008).
- [3] Pokharel, M., Lee, S. and Park, J. S.: Disaster Recovery for System Architecture Using Cloud Computing, *2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*, pp. 304–307 (online), DOI: 10.1109/SAINT.2010.23 (2010).
- [4] Nayadkar, P. P.: Automatic and secured backup and restore technique in Android, *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–4 (online), DOI: 10.1109/ICIIECS.2015.7193103 (2015).
- [5] Gavrilovska, A., Schwan, K. and Oleson, V.: A practical approach for 'zero' downtime in an operational information system, *Proceedings 22nd International Conference on Distributed Computing Systems*, pp. 345–352 (online), DOI: 10.1109/ICDCS.2002.1022272 (2002).
- [6] Zhou, Z., Yang, X., Zhao, D., Rich, P., Tang, W., Wang, J. and Lan, Z.: I/O-Aware Batch Scheduling for Petascale Computing Systems, *2015 IEEE International Conference on Cluster Computing*, pp. 254–263 (online), DOI: 10.1109/CLUSTER.2015.45 (2015).
- [7] Malensek, M., Pallickara, S. L. and Pallickara, S.: Alleviation of Disk I/O Contention in Virtualized Settings for Data-Intensive Computing, *2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC)*, pp. 1–10 (online), DOI: 10.1109/BDC.2015.32 (2015).
- [8] Kaiser, J., Meister, D., Gottfried, V. and Brinkmann, A.: MCD: Overcoming the Data Download Bottleneck in Data Centers, *2013 IEEE Eighth International Conference on Networking, Architecture and Storage*, pp. 88–97 (online), DOI: 10.1109/NAS.2013.18 (2013).
- [9] Xia, R., Yin, X., Alonso Lopez, J., Machida, F. and Trivedi, K. S.: Performance and Availability Modeling of IT Systems with Data Backup and Restore, *IEEE Transactions on Dependable and Secure Computing*, Vol. 11, No. 4, pp. 375–389 (online), DOI: 10.1109/TDSC.2013.50 (2014).
- [10] Bhagwat, D., Eshghi, K., Long, D. D. E. and Lillibridge, M.: Extreme Binning: Scalable, parallel deduplication for chunk-based file backup, *2009 IEEE International Symposium on Modeling, Analysis Simulation of Computer and Telecommunication Systems*, pp. 1–9 (online), DOI: 10.1109/MASCOT.2009.5366623 (2009).
- [11] Chen, H., Xiong, J. and Sun, N.: A novel hint-based I/O mechanism for centralized file server of cluster, *2008 IEEE International Conference on Cluster Computing*, pp. 194–201 (online), DOI: 10.1109/CLUSTER.2008.4663771 (2008).