

非同期リモートコピーにおける書き込み遅延時間監視方式

江丸 裕教[†] 矢川 雄一[†] 原 純一[†] 木原 健一[†] 三井小吾[‡]

(株) 日立製作所 システム開発研究所[†]

(株) 日立製作所 ソフトウェア事業部[‡]

1. はじめに

企業情報システムにおける無停止運用やデータ保護の重要性は、マーケットのグローバル化や、Web による 24 時間 365 日のサービス提供などを背景に高まる一方である。ところが、テロや自然災害など、企業情報システムの停止やデータ損失に繋がりがねないリスクは数多く存在する。これらのリスクを低減するには、災害や障害を起こりうるものと想定し、災害や障害の発生時に停止したシステムをいつまでにどうやって復旧するかを予め計画しておく必要がある^[1]。これをディザスタリカバリ (Disaster Recovery 以下 DR) と呼ぶ。DR では、どの時点のデータをいつまでに復旧するかを指標として事前に定めておき、その指標が遵守されていることを監視しつつ運用することが重要である。この指標のうち、前者、すなわち、どの時点のデータを復旧できるべきかを目標復旧時点 (Recovery Point Objective, 以下 RPO)、後者、すなわち、被災してからいつまでに業務を再開できるべきかを目標復旧時間 (Recovery Time Objective) と呼ぶ。

ストレージシステム (以下ストレージ) における DR 技術の一つに、データを多重化することによって災害や障害からデータを守るデータ保護があり、更にその一技術としてデータを別サイトにコピーする非同期リモートコピーがある^[2]。非同期リモートコピーでは、書き込み遅延時間と呼ばれる性能情報の監視が RPO を遵守しているか否かを監視することに該当するが、特にオープンシステムにおいて、その取得が困難であるという問題点があった。そこで、本研究は、オープンシステムにおいても、一般的なストレージにおいて取得容易な性能情報を組み合わせることにより、ホストやアプリケーションに影響を与えずに遅延時間の監視を実現することを目的とする。

2. 非同期リモートコピーとその監視

リモートコピーとは、業務を行っているサイ

ト (ローカルサイト) から地理的に離れた位置にリモートサイトを用意し、ローカルサイトと全く同じデータをリモートサイトに作成する、DR を実現するための基盤技術の一つである。

リモートコピーには、業務ホストから I/O が発行された際に、リモートサイト側でのデータ書き込み完了応答を待ってサーバに書き込み完了応答を返す同期方式と、リモートサイト側でのデータ書き込み完了応答を待たずにサーバに書き込み完了応答を返す非同期方式 (図 1) がある。

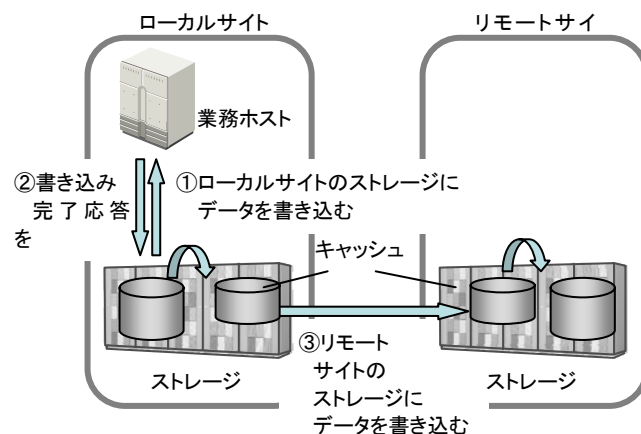


図 1 非同期リモートコピーの構成

非同期方式は、ネットワークの帯域変動や急激な業務負荷変動時にも業務の応答性能に対する影響を抑えられるため、リモートサイトを遠距離に配置する場合や性能が不安定なネットワーク回線を用いる場合に有効である。しかし、ネットワークの帯域変動や急激な業務負荷変動が発生した場合には、業務ホストで I/O が発行されてからリモートサイトにデータが書き込まれるまでの遅延時間 (以下書き込み遅延時間) に影響が出る。この書き込み遅延時間は、ある時点において業務ホストから発行されローカルサイトに書き込まれた I/O が、どれだけの遅延でリモートサイトに書き込まれるかを示す値である。したがって、ローカルサイトに障害、災害

Write Delay Time Monitoring Method for Asynchronous Remote Copy

[†] Hironori Emaru, Yuichi Yagawa, Junichi Hara, Kenichi Kihara, Systems Development Laboratory, Hitachi Ltd.

[‡] Shogo Mitsui, Software Division, Hitachi, Ltd.

が発生した場合、リモートサイトでは書き込み遅延時間分だけデータが失われた状態で業務復旧の作業を行うことになる。すなわち、非同期リモートコピーにおいて書き込み遅延時間を監視することは、RPO を遵守しているか否かを監視することと同義である。

業務ホストが I/O にタイムスタンプ情報を埋め込むメインフレームシステムでは、その情報を利用して書き込み遅延時間を取得することが可能である。つまり、リモートサイト側で取得した現在の時刻と、I/O に埋め込まれている時刻の差が書き込み遅延時間である。しかしながら、オープンシステムでは、業務ホストやアプリケーションに手を入れる必要があるなど、タイムスタンプ情報を I/O に埋め込むことが現実的ではないため、非同期リモートコピーの書き込み遅延時間を監視することは困難であった。

3. 書き込み遅延時間の監視方法

前章で述べたように、オープンシステムにおいて書き込み遅延時間を取得するために I/O 内のタイムスタンプ情報を利用する方法は現実的ではない。そこで、取得可能な性能情報を利用して近似的な値を求めるアプローチをとる。

非同期リモートコピーでは、図 1 に示すようにローカルサイトのボリュームへの書き込み処理とは非同期に、バッファを介してリモートサイトのボリュームに対する書き込みを行う。バッファはローカルサイト側、リモートサイト側双方のストレージに設けられるが、遅延発生の原因となるのはネットワークであることが大半であるため、遅延時間の計算にはローカルサイト側のバッファのみを考慮すればよい。

この前提により、ある時点において、ローカルサイト側のバッファに滞留している I/O がどれくらいの時間で掃けるかによって書き込み遅延時間を求める方式を提案する。具体的には(a)に示すように、バッファにある時点で滞留している I/O 数 C [IOs] をローカルサイトからリモートサイトへの単位時間(1 秒)あたりの平均 I/O 流出量 X [IOs/sec] で割ることにより書き込み遅延時間 De [sec] を求める。

$$De [\text{sec}] = C [\text{IOs}] \div X [\text{IOs/sec}] \cdots (a)$$

書き込み遅延時間を求めるのに、I/O 流出量といったリアルタイムに変動する情報を絶えず取得するのは、システムに掛かる負荷を考えても現実的ではない。そこで、定期的に取り得た情報を元に書き込み遅延時間を求める方法を検討した。ここではその一方法として、バッファに

滞留している I/O 数およびホストからの I/O に対してストレージ側で到着順に割り振った番号(シーケンス番号)を定期的に取り得し、これを元に書き込み遅延時間を求める方法を示す。

1. ある時刻(t_1 とする)にバッファ滞留量とシーケンス番号を取得、それぞれ C_1 [IOs], S_1 とする
2. t_1 から T [sec] 後の時刻 t_2 に、バッファ滞留量とシーケンス番号を取得、それぞれ C_2 [IOs], S_2 とする
3. T [sec] 間の平均 I/O 流入量を I [IOs/sec] とすると、下記の式が成り立つ
 $I = (S_2 - S_1) \div T \cdots (b)$
4. T [sec] 間の平均 I/O 流出量を X [IOs/sec] とすると、
 $C_2 = C_1 + (I - X) \times T \cdots (c)$
の関係が成り立つ。(b)と(c)より、
 $X = (C_1 - C_2 + S_2 - S_1) \div T \cdots (d)$
が成り立つ。
5. (a)より、時刻 t_2 における書き込み遅延時間 De [sec] は
 $De = C_2 \div X$
であるから、 X を(d)により置き換えることにより、書き込み遅延時間 De [sec] は下記の式で求まる。
 $De = C_2 \times T \div (C_1 - C_2 + S_2 - S_1) \cdots (e)$

これにより、 T [sec] ごとにローカルサイト側のストレージのバッファ滞留数とシーケンス番号を取得し、前回取得した値とともに(e)の計算式に代入することによって、ローカルサイト側のストレージから取得した情報のみによって書き込み遅延時間を求めることが可能になる。

4. おわりに

本報告では、業務ホストやアプリケーションに影響を与えずに書き込み遅延時間の監視を実現する方式を提案した。今後は、本方式の有効性を検証するため、実機による評価を実施する。

参考文献

- [1] Rudolph, C.G, "Business continuation planning/disaster recovery", Communications Magazine, IEEE, Volume 28, Issue 6, PP.25-28, June 1990.
- [2] Advanced functions for storage subsystems: Supporting continuous availability, IBM SYSTEM JOURNAL, Volume 42, Number 2, P.268, 2003.