

記憶装置の故障に対する信頼度とコストのモデル化

日大生産工 (学部) ○荒井 裕典

日大生産工 新井 雅之

1. まえがき

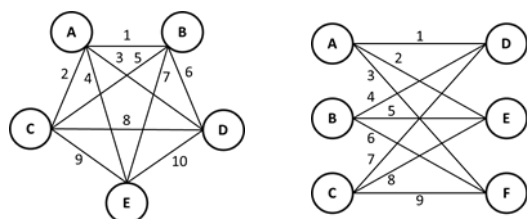
ディスクアレイの高信頼化として、冗長設計に基づく様々な手法が提案されている[1][2][3]. 冗長設計では信頼度とコストがトレードオフの状況にあるため、長期間にわたって使用されるディスクアレイでは、コスト効率の良い解決策を見つけることが、より重要になっていくと考えられる[4]. 本研究では、信頼度とコストのトレードオフを考慮したディスクアレイ設計への適用を目的として、実シナリオに基づく信頼度とコストのモデル化について検討する.

2. ディスクアレイ構成例

本研究では複数の冗長設計に適用可能な信頼性・コストモデルについて検討する. 以下では、冗長設計の例として、RAID5/6 およびダブルパリティ手法について概説する.

RAID5 では、 k 個のデータディスクに対して 1 個の冗長ディスクを追加し、各ディスクに分散してパリティを保存する. 任意の 1 台のディスク故障が発生してもデータロスが発生しない. RAID6 では冗長ディスクを 2 個用意し、任意の 2 台までの故障に耐える. RAID6 と比較して、冗長データ生成のための計算コストが増加する.

ダブルパリティは、RAID5 と同等の計算コストで高い信頼度を目指して提案された手法である[3]. 図 1 にダブルパリティの構成例を



(a) (10, 5)完全グラフ構成 (b) (9, 6)二部グラフ構成
図 1. ダブルパリティ構成例

示す. 図 1(a)は次数 5 の完全グラフであり、10 個のデータディスク(1~10)に対して 5 個のパリティディスク(A~E)が追加される構成を意味する. 図 1(b)の構成では、9 個のディスクに対して 6 個のパリティディスクが追加される. いずれの構成においても、各データディスク上のデータは、2 個のパリティディスクに重複して保存される. 3 台以上のディスクが故障した場合でも、消失誤り訂正により回復可能となる場合がある.

3. 信頼度とコスト計算のためのモデル化

3-1. 定期メンテナンスシナリオに対するディスクアレイ状態遷移

実シナリオに基づいてディスクアレイの故障確率を算出し、信頼度とコストのモデルを作成する. 今回は 1 週間に 1 度メンテナンスが入り、故障したディスクを正常なものに差し替えるモデルについて検討する. 図 2 に本モデルにおけるディスクアレイ状態遷移図を示す. 第 i 日目におけるディスクアレイのある状態を $s_{i,j}$ と表す. 図中の白丸はデータロス無し、黒丸はデータロス有りの状態を示しており、下に行くにつれて故障しているディスク数が増えていく. 図 2 の各状態 $s_{i,j}$ に対する

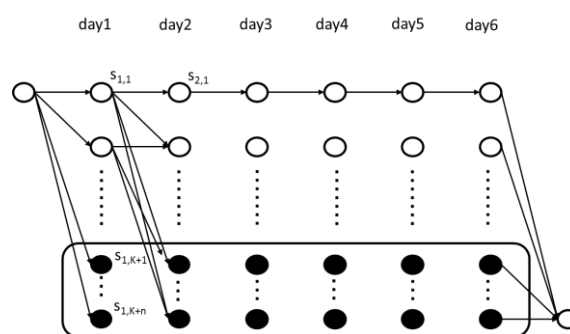


図 2. 定期メンテナンスを考慮したディスクアレイ状態遷移図

Reliability and Cost Modeling for Storage Cluster Considering Device Failure

Hironori ARAI, Masayuki ARAI

状態確率を $Pr(i,j)$ で表す. 状態確率 $Pr(i,j)$ は, 適用される冗長設計とディスク単体の故障率により決まる状態遷移確率 $Pr(i,j \rightarrow i+1,j')$ から算出可能である.

3-2. 状態遷移確率の算出

文献[2]では, マルコフ解析に基づくダブルパリティの MTTF (Mean Time to Failure) 算出法について示されている. 本研究における状態遷移確率 $Pr\{s_{i,j} \rightarrow s_{i+1,j'}\}$ も, 同様の手法により算出される. 故障率 λ , ディスク総数 N , MTTR (Mean Time to Repair) ρ としたダブルパリティにおけるマルコフモデルを図 3 に示す. このマルコフモデルにおける各状態 $0 \sim 5$ は, 故障が発生し, かつデータロスが発生していない状態を示しており, F はデータロス発生状態を示している.

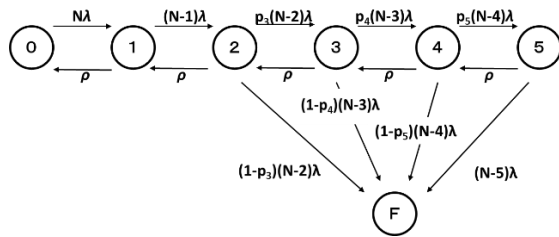


図 3 ダブルパリティに対する MTTF 算出のためのマルコフモデル

3-3. 信頼度およびコストモデル

図 2 に基づいて信頼度とコストのモデルを作成する. まず, 第 i 日目においてデータロスが発生している確率 $S_{i,F}$ およびデータロスが発生していない確率 $S_{i,G}$ を求める. それぞれの日において, k 通りの状態 $s_{i,1}, \dots, s_{i,k}$ ではデータロスが発生しておらず, n 通りの状態 $s_{i,k+1}, \dots, s_{i,k+n}$ ではデータロスが発生しているため,

$$S_{i,F} = \sum_{i=1}^6 \sum_{j=1}^n Pr(i,j+k) \quad (1)$$

$$S_{i,G} = \sum_{i=1}^6 \sum_{j=1}^k Pr(i,j) \quad (2)$$

となる. 第 i 日目におけるディスクアレイの信頼度 $R(i)$ は, $R(i) = S_{i,G} = 1 - S_{i,F}$ となる.

定期メンテナンスは, 1 台以上のディスク故障が発生している場合に必要となり, 固定コスト $C_{constant}$ と, ディスク 1 台毎のメンテナンスコスト C_{disk} が発生する. また, ディスクアレイが故障により停止している期間もペナルティコストとして考慮する. 停止 1 日あたりのペナルティコストを $C_{penalty}$ とおく. このとき, 期間

中の総期待コスト C は,

$$C = C_{constant} \cdot \{1 - Pr(6,1)\} + C_{disk} \cdot \sum_{j=2}^{k+n} Fu(s_{6,j}) \cdot Pr(6,j) + C_{penalty} \cdot T_{MTBF} \quad (3)$$

となる. ここで, T_{MTBF} は本モデルにおける MTBF (Mean Time Between Failure) であり,

$$T_{MTBF} = \sum_{i=1}^5 \left\{ \left(\prod_{j=1}^{i-1} S_{j,G} \right) \cdot S_{i,F} \cdot (7-i) \right\}$$

と求まる. また, $Fu(s_{i,j})$ は, 各状態における故障ディスク台数を表す.

4. まとめ

本報告では図 3 のモデルを用いて, 計算式を立式した. 今後は複数のモデルを作成していく. 例として, 予備のディスクをホットスワップして用意しておくもの[4], メンテナンスの頻度が異なるもの, データロスが発生した時点で緊急メンテナンスを行うもの, データロス一歩手前の状態でも緊急メンテナンスを行うものが挙げられる. また, 複数のディスク故障でもデータロスが発生する場合もあれば最小の 3 個のディスク故障でデータロスが発生する場合もあることから, いつ, 何個のディスク故障が発生してデータロスが発生したのかどうかのシナリオも考える.

参考文献

- [1] 米田友洋, 梶原誠司, 土屋達弘, ディペンダブルシステム, 共立出版, 2005 年.
- [2] S. J. Thomas Schwarz, Darrell D. E. Long, Jehan-Francois Paris, "Reliability of Disk Arrays with Double Parity," 19th Pacific Rim International Symposium on Dependable Computing, pp. 108-117, 2013.
- [3] Jehan-Francois Paris, Darrell D. E. Long, Witold Litwin LAMSADE, "Three-Dimensional Redundancy Codes for Archival Storage," 21st International Symposium on Modelling, Analysis & Simulation of Computer and Telecommunication Systems, pp. 328-332, 2013.
- [4] Jehan-Francois Paris, Darrell D. E. Long, S. J. Thomas Schwarz, "Zero-Maintenance Disk Arrays," 19th Pacific Rim International Symposium on Dependable Computing, pp. 140-141, 2013.