2000年度日本オペレーションズ・リサーチ学会
秋季研究発表会

差分バックアップ運用における最適方策

02991833 愛知工業大学 ＊銭 存華 QIAN Cun Hua
愛知工業大学 吉 勇 JI Yong
01402793 名古屋銀行 中村 正治 NAKANMURA Syouji
01400043 愛知工業大学 中川 覃夫 NAKAGAWA Toshio

1 まえがき

近年の高度情報化社会において，情報を管理するデータベースの信頼性が重要視されている．データベース管理システムでは，その障害状況に応じて復旧するための対策をまえもって講じておき，実際に障害が発生した場合，すみやかにデータベースを正常な状態に復旧しなければならない [1]．現在では，バックアップ処理に要するオーバヘッドを大幅に軽減した差分バックアップが多くのシステムで採用されているが [2]，定期的にフルバックアップを実施する必要がある．

著者らは，文献 [3] において，2種類のショックをもつ累積損傷モデルを提案し，データベースのバックアップ方策に応用した．そこでは，データの更新毎に差分バックアップを行う仮定の下に，期待費用を導出し，最適フルバックアップ間隔の時間方策について考察した．しかし，大容量のデータベースなどでは，とくに，オンラインシステムでは，頻繁にバックアップを行うのが事実上不可能である．

ここでは，データベースの情報更新が非定常ポアソン過程に従って発生したとき，ある一定間隔を T とし，T 毎にデータベースの累積更新量をチェックする．累積更新量がある与えられた基準値 $K(0<K<\infty)$ よりも小さい場合は，差分バックアップを，基準値 K よりも大きい場合は，フルバックアップを起動する．ただし，$(N-1)T$ までに累積更新量が基準値 K よりも小さい場合は，差分バックアップを起動するが，NT でフルバックアップを起動する $(N=1,2,\cdots)$[2]．そのとき，フルバップアップに要する費用を c_1，差分バップアップを行う時点までの累積更新量を x としたとき，その費用を $c_2+c_0(x)$ $(c_2<c_1)$ とし，単位時間当たりの期待費用を求める．さらに，それを最小にする最適な K^* と N^* を解析的に求める．最後に，数値例を与え，種々検討する．

2 モデルの解析

データベースに対する情報への更新は，強度関数 $\lambda(t)$ と平均値関数 $R(t)$ をもつ非定常ポアソン過程に従って発生する．故に，任意の区間 $(0,t]$ 間で，j 回の更新が行われる確率は，$H_j(t)\equiv\{[R(t)]^j/j!\}e^{-R(t)}$ $(j=0,1,2,\cdots)$ であり，$R(0)\equiv 0$, $R(\infty)\equiv\infty$ とおく．さらに，j 回目の更新で発生する新規更新量を W_j とし，同一の独立な確率分布 $G(x)\equiv P_r\{W_j\le x\}$ $(j=1,2,\cdots)$ をもつ．従って，j 回の更新による累積更新量を $Z_j\equiv\sum_{i=1}^{j}W_i$ とおくならば，Z_j の分布は，$G(x)$ の j 重たたみこみである．$G^{(j)}(x)$ で与えられ，$G^{(0)}(x)\equiv 1$ とおく．

累積更新量が iT で K を越えていない確率を $F_i(K)$ とすると，$F_i(K)=\sum_{j=0}^{\infty}H_j(iT)G^{(j)}(K)$ となる．ちょうど $[(i-1)T,iT]$ 間で累積更新量が K を越える確率は，$F_{i-1}(K)-F_i(K)$ となり，$F_0(K)\equiv 1$ とおく．したがって，単位時間当たりの期待費用 $C(K,N)$ は，

$$C(K,N)=\frac{c_1+\sum_{i=1}^{N-1}C_i(K,T)}{T\sum_{i=0}^{N-1}F_i(K)}. \qquad (1)$$

ただし，$C_i(K,T)\equiv\sum_{j=0}^{\infty}H_j(iT)\int_0^K[c_2+c_0(x)]dG^{(j)}(x)$.

2.1 レベル方策

差分バックアップ運用において，累積更新量 K でのみフルバックアップするレベル方策を考える．単位時間当たりの期待費用 $C_1(K)$ は，

$$C_1(K)\equiv\lim_{N\to\infty}C(K,N)=\frac{c_1+\sum_{i=1}^{\infty}C_i(K,T)}{T\sum_{i=0}^{\infty}F_i(K)}. \qquad (2)$$

単位時間当たりの期待費用を最小にする最適な K^* を求める．式 (2) から，$C_1(0)=\infty$, $C_1(\infty)=[c_2+c_0(\infty)]/T$. $C_1(K)$ を K で微分して零とおくと，

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}H_j(iT)\int_0^K G^{(j)}(x)dc_0(x)=c_1-c_2 \qquad (3)$$

を得る．式 (3) の左辺を $Q(K)$ とすると，

$$\begin{aligned} Q(0) &\equiv \lim_{K\to 0}Q(K)=0,\\ Q'(K) &= c_0'(K)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}H_j(iT)G^{(j)}(K) \end{aligned}$$

となる．したがって，$c_0(x)$ が x の単調増加関数である場合，$Q(K)$ は 0 から $Q(\infty)$ までの単調増加関数であり，

$Q(\infty) > c_1 - c_2$ ならば，式 (3) を満たす有限かつ唯一の K^* $(0 < K^* < \infty)$ が存在する．そのとき，期待費用は

$$C_1(K^*) = \frac{c_2 + c_0(K^*)}{T}. \tag{4}$$

とくに，$c_0(x) = c_0(1-e^{-sx})$ $(s>0)$ の場合，$Q(K)$ は 0 から $c_0\sum_{i=0}^{\infty} E_i(T)$ までの単調増加関数である．ここで，$E_i(T) \equiv e^{-[1-g(s)]R(iT)}$, $g(s) \equiv \int_0^{\infty} e^{-sx}dG(x)$ である．

2.2 回数方策

差分バックアップ運用において，N でのみフルバックアップする回数方策を考える．単位時間当たりの期待費用 $C_2(N)$ は，

$$C_2(N) \equiv \lim_{K\to\infty} C(K,N) = \frac{c_1 + \sum_{i=1}^{N-1} C_i(T)}{NT}. \tag{5}$$

ただし，$C_i(T) \equiv \sum_{j=0}^{\infty} H_j(iT)\int_0^{\infty}[c_2+c_0(x)]dG^{(j)}(x)$ であり, $c_0(x)$ が x の単調増加関数であるならば，$C_i(T)$ は i の単調増加関数である．

単位時間当たりの期待費用を最小にする最適な N^* を求める．式 (5) から，$C_2(1) = c_1/T$, $C_2(\infty) = [c_2 + c_0(\infty)]/T$. $C_2(N)$ の N についての差分，すなわち，$C_2(N+1) - C_2(N) \geq 0$ とおくと,

$$NC_N(T) - \sum_{i=1}^{N-1} C_i(T) \geq c_1 \tag{6}$$

を得る．式 (6) の左辺を $L(N)$ とすると,

$$\begin{aligned} L(1) &= C_1(T), \\ L(N) - L(N-1) &= N[C_N(T) - C_{N-1}(T)] > 0 \end{aligned}$$

となる．よって，$L(N)$ は $C_1(T)$ からの単調増加関数である．もし，$C_1(T) \geq c_1$ ならば，$N^* = 1$ である．もし，$C_1(T) < c_1 \leq L(\infty)$ ならば，式 (6) を満たす有限かつ唯一の N^* $(1 < N^* < \infty)$ が存在する．

とくに，$c_0(x) = c_0(1-e^{-sx})$ $(s>0)$ の場合，式 (6) は

$$\sum_{i=0}^{N-1} [E_i(T) - E_N(T)] \geq \frac{c_1 - c_2}{c_0} \tag{7}$$

となる．式 (7) の左辺は $1 - E_1(T)$ から $\sum_{i=0}^{\infty} E_i(T)$ までの単調増加関数であることがわかる．

3 数値例

ここでは，前節の解析のもとに，$c_0(x) = c_0(1-e^{-sx})$, $c_1 = c_2 + c_0$ とし，データベースの更新はランダムに発生すると考え，いわば，$\lambda(t) = \lambda$ とする．さらに，データベースの更新による新規更新量が，指数分布 $G(x) = 1 - e^{-\mu x}$ に従う場合を考え，単位時間当たりの期待費用を最小にする K^* と N^* の具体的な数値を求め，最適方策について考察を行う．

この場合，式 (3) は,

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{(i\lambda T)^j}{j!} e^{-i\lambda T} \times \sum_{k=j}^{\infty} \frac{[\mu^j(s+\mu)^{k-j} - \mu^k]K^k}{k!} e^{-(s+\mu)K} = 1 \tag{8}$$

となり，式 (8) の左辺は，K に関して 0 から $1/(1-e^{-\frac{s\lambda T}{s+\mu}})$ までの単調増加関数である．したがって，式 (8) を満たす有限かつ唯一の K^* $(0 < K^* < \infty)$ が存在し，そのときの期待費用は $C_1(K^*) = [c_1 - c_0 e^{-sK^*}]/T$ である．

式 (7) は

$$\frac{1 - e^{-\frac{Ns\lambda T}{s+\mu}}}{1 - e^{-\frac{s\lambda T}{s+\mu}}} - Ne^{-\frac{Ns\lambda T}{s+\mu}} \geq 1 \tag{9}$$

となり，式 (9) の左辺は, 明らかに N に関して $1 - e^{-\frac{s\lambda T}{s+\mu}}$ から $1/(1-e^{-\frac{s\lambda T}{s+\mu}})$ までの単調増加関数である．したがって，式 (9) を満たす有限かつ唯一の N^* $(1 < N^* < \infty)$ が存在し，そのときの期待費用 $C_2(N^*)$ は,

$$\frac{c_1 - c_0 e^{-\frac{(N^*-1)s\lambda T}{s+\mu}}}{T} < C_2(N^*) \leq \frac{c_1 - c_0 e^{-\frac{N^* s\lambda T}{s+\mu}}}{T}. \tag{10}$$

4 むすび

データベースの情報更新が非定常ポアソン過程にしたがって発生し，ある一定間隔 T で差分バックアップする確率モデルを構築し，考察を行った．すなわち，差分バックアップの対象になるデータベースのトラックの累積更新量にある上限値 K を設定し，それ以上になるならば，フルバックアップを行うレベル方策と，N 回目にフルバックアップを行う回数方策を設定した．そして，単位時間あたりの期待費用を求め，それを評価尺度として，最適な K^* と N^* が存在することを解析的に示し，数値例により種々の議論を行った．

参考文献

[1] 南谷崇, フォールトトレラントコンピュータ．オーム社.1996.

[2] K.Suzuki and K.Nakajima, Storage Management Software. *Fujitsu*. 1995, **46**, 389-397.

[3] C.Qian, S.Nakamura and T.Nakagawa, "Cumulative damage model with two kinds of shocks and its application to the backup policy," *JORSJ*, Vol.42, no.4, pp.501-511, 1999.