The Tenth International Symposium on Operations Research and Its Applications (ISORA 2011)
Dunhuang, China, August 28–31, 2011

Studies on Tenuring Collection Times for a Generational Garbage Collector*

Xufeng Zhao[1,2] Cunhua Qian[2] Syouji Nakamura[3]

[1]Department of Business Administration
Aichi Institute of Technology, Toyota 470-0392, Japan
[2]School of Economics and Management
Nanjing University of Technology, Nanjing 210009, China
[3]Department of Human Life and Information
Kinjo Gakuin University, Nagoya 463-8521, Japan

Abstract It is an important problem to analyze the collection costs and determine tenuring collection times to meet the pause time goal for a generational garbage collector. From such viewpoints, this paper analyzes the costs suffered for collections according to the working schemes firstly, and then, proposes one garbage collection model. Garbage collections occur at a nonhomogeneous Poisson process, and tenuring collections are made at time T and at level K. Applying the techniques of degradation processes and continuous wear processes, the expected cost rate is derived, four cases of optimal policies for the model are discussed analytically and numerically.

Keywords Garbage Collection; Tenuring Collection; Minor Collection; Continuous Process; Optimal Policy

1 Introduction

In recent years, *generational garbage collection* [1-5] has been popular with programmers for the reason that it can be made more efficient. Compared with classical tracing collectors, *e.g.*, reference counting collector, mark-sweep collector, mark-compact collector and copying collector, a *generational garbage collector* is effective in applications of the computer programs with character that it is unnecessary to mark or copy all active data of the whole heap for every collection, *i.e*, the collector concentrates effort on those objects most likely to be garbage. Based on the weak generational hypothesis [2] which asserts that most objects are short-lived after their allocation, a generational garbage collector segregates objects by age into several regions called *generations* or *multiple generations*. The survival rates of younger generations are always much lower than those of older ones, which means that younger generations are more likely to be garbage and can be collected more frequently than older ones. Although such collections cost much shorter time than those of full collections, the problems of pointers from older generations to younger ones and the size of root sets for younger generations will become more

*This project is supported by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science under Grant No. 22500897; National Natural Science Foundation of China (70471017, 70801036); Humanities and Social Science Research Foundation of China (05JA630027).

complicated. For these reasons, many collectors are limited to just two or three generations [1, p.153]. For instance, the collector which is used in Sun's HotSpot Java Virtual Machine (JVM) manages heap space for both young and old generations [5]: new objects space *Eden*, two equal survivor spaces SS♯1 and SS♯2 for surviving objects, and tenured objects space *Old* (*Tenured*), where Eden, SS♯1 and SS♯2 are for young generation, and Old (Tenured) is for old one.

A generational garbage collector uses *minor collection* and *tenuring collection* for young generation and *major collection* for multi-generations. However, for every garbage collection, the manner of *stop and copy* pauses all application threads to collect the garbage. The duration of time in which a garbage collector has worked is called *pause time* [1, p.148], which is an important parameter for interactive systems, and depends largely upon the volume of surviving objects and the type of collections. That is, pause time suffered for minor collections increases with the number of collections and is less than that of tenuring collections, major collection pause time is the longest among the three ones.

There have been very few research papers that studied analytically optimal collection times for a generational garbage collector. Most problems were concerned with several ways to introduce garbage collection methods and how to tune the garbage collector, which are complex to operate due to random accesses of programs in the memory [6-10]. We propose that garbage collection is a *stochastic decision making process* and should be analyzed by the theory of stochastic processes. As the applications of damage models, a garbage collection model for a database in the computer system [11] was studied, but the theoretical point of collection was not considered essentially, and optimal policies for a generational garbage collector with tenuring threshold [12] according to practical working schemes were studied recently, however, it is may be difficult to inspect the survivor rates exactly at collection times.

Furthermore, increase in objects might be unclear at discrete times for the high frequency of computer processes, it would be more practical to assume that surviving objects that should be copied increase with time continuously and roughly according to some mathematical laws. On the other hand, degradation processes [13] and continuous wear processes [14] paly an important role in describing the degradation of the unit. Life distribution properties, failure rates and some maintenance policies subjected to such processes were summarized in [13, 14] and such processes has been applied to various fields such as material, physics, finance and statistics [15-18].

This paper considers a pause time goal which is called time cost or cost for simplicity. Our problem is to obtain optimal tenuring collection times which minimize the expected cost rates. First, the working schemes and some analyses about collection costs are given. Second, garbage collections occur at a nonhomogeneous Poisson process, and tenuring collections are made at time T and at level K. Applying the techniques of degradation processes and continuous wear processes, the expected cost rate is derived, four cases of optimal policies for the model are discussed analytically. Fourth, numerical examples of the policies are given.

2 Models and Optimal Policies

2.1 Working Schemes

Detailed working schemes of a generational garbage collector in general which has been introduced in [1, 5, 12] are given as following steps:

1. New objects are allocated in Eden.
2. When the first minor collection occurs, surviving objects are copied from Eden into SS♯1.
3. When the second minor collection occurs, surviving objects from Eden and SS♯1 are copied into SS♯2.
4. In the fashions of 1-3, minor collection copies surviving objects between the two survivor spaces until they become tenured, *i.e.*, tenuring collection occurs when some parameter meets the tenuring threshold, and then, the older or oldest objects are copied into Old.
5. When Old fills up, major collection of the whole heap occurs, and surviving objects from Old are kept in Old, while objects from Eden and survivor space are kept in a survivor space.

Based on [2], from the viewpoints of collection efficiency, new objects can be tenured only if they survive at least one minor collection, because objects that survive two minor collections are much less than those survive just one minor collection. In other words, surviving objects are likely to reduce slightly with the number of minor collections beyond two. Although major collection spends much time, Old will be filled with tenured objects slowly, especially when the tenuring threshold is high and the survivor rate is low, that is, major collection occurs rarely in this case. So that, this paper concentrates only on minor and tenuring collections and considers tenuring collections as renewal points of the collection processes, that is, after tenuring collection, the same collection cycle begins with step 1: $1 \to 2 \to 3 \to 4 \to 1 \to \cdots$.

It is assumed that the total volume of surviving objects in Eden and survivor space at time t is $Z(t) = A(t)t + \sigma B(t)$ with distribution $\Pr\{Z(t) \le x\} = W(t,x)$, where $A(t)$ and $B(t)$ are random variables. Letting $c_S + c_M(x)$ be the cost suffered for every minor collection, where c_S is the constant cost of scanning surviving objects and x is the surviving objects that should be copied, $c_M(x)$ increases with x and $c_M(0) \equiv 0$. c_K, c_T and c_N ($c_K, c_T, c_N > c_S + c_M(K)$) be the cost suffered for tenuring collection when surviving objects have exceeded a threshold level K ($0 < K < \infty$), at time T and at Nth collection, respectively. Then, the expected cost of minor collection at time t is

$$C(t,K) = \frac{1}{W(t,K)} \int_0^K \Big[c_S + c_M(x)\Big] \mathrm{d}W(t,x), \tag{1}$$

where $C(0,K) \equiv 0$. It is clear that if $-W'(t,x)/W(t,x)$ increases with t for any $x \ge 0$, $C(t,K)$ increases with t for any $K \ge 0$.

2.2 Modeling and Optimization

It is assumed that garbage collections occur at a nonhomogeneous Poisson process with an intensity function $\lambda(t)$ and a mean-value function $R(t) \equiv \int_0^t \lambda(u)\mathrm{d}u$. Then, the probability that collections occur exactly j times in $(0,t]$ is

$$H_j(t) \equiv \frac{[R(t)]^j}{j!} e^{-R(t)} \quad (j = 0, 1, 2, \cdots).$$

Letting $F_j(t)$ denote the probability that collections occur at least j times in $(0,t]$. Then,

$$F_j(t) = \int_0^t H_{j-1}(u)\lambda(u)du = \sum_{i=j}^{\infty} H_i(T), \quad (j = 1, 2, \cdots), \tag{2}$$

where $F_0(t) \equiv 1$.

Suppose that minor collections are made when the garbage collector begins to work, tenuring collection is made at a planned time T $(0 < T \leq \infty)$, or when surviving objects have exceeded a threshold level K $(0 < K \leq \infty)$, whichever occurs first. Then, the mean time to tenuring collection is

$$E(L) = TW(T,K) + \int_0^T t d\overline{W}(t,K) = \int_0^T W(t,K)dt, \tag{3}$$

where $\overline{V}(t,x) \equiv 1 - V(t,x)$ for any distribution $V(t,x)$.

The expected cost suffered for minor collections until tenuring collection is

$$\begin{aligned} C_M(T,K) &= W(T,K) \sum_{j=1}^{\infty} \int_0^T C(t,K) dF_j(t) + \int_0^T \left[\sum_{j=1}^{\infty} \int_0^t C(u,K) dF_j(u) \right] d\overline{W}(t,K) \\ &= \sum_{j=1}^{\infty} \int_0^T C(t,K)W(t,K) dF_j(t) = \int_0^T \lambda(t)C(t,K)W(t,K)dt. \end{aligned} \tag{4}$$

Then, the expected cost until tenuring collection is

$$E(C) = c_K - (c_K - c_T)W(T,K) + \int_0^T \lambda(t)C(t,K)W(t,K)dt. \tag{5}$$

Therefore, from (3) and (5), by using the renewal reward process [19], the expected cost rate is

$$C(T,K) = \frac{\begin{array}{c} c_K - (c_K - c_T)W(T,K) \\ + \int_0^T \lambda(t)C(t,K)W(t,K)dt \end{array}}{\int_0^T W(t,K)dt}. \tag{6}$$

It can be seen that $C(T,K)$ includes the following collection polices:

- Tenuring collection is made at time T for a given K, the reason why making such a policy is $c_T < c_K$.
- Tenuring collection is made at level K for a given T. In this case, $c_K < c_T$.
- Tenuring collection is made only at time T or only at level K. In these two cases, $c_K = c_T$.

2.2.1 Optimal T^*

When $c_T < c_K$, we find an optimal T^* which minimizes $C(T,K)$ in (6) for a given K. Letting $r(t,x)$ be the failure rate of $W(t,x)$, *i.e.*, $r(t,x) \equiv -W'(t,x)/W(t,x)$ [20]. Then, differentiating $C(T,K)$ with respect to T and setting it equal to zero,

$$(c_K - c_T)\left[r(T,K)\int_0^T W(t,K)\mathrm{d}t - \overline{W}(T,K)\right] + \int_0^T \Big[\lambda(T)C(T,K) - \lambda(t)C(t,K)\Big] W(t,K)\mathrm{d}t = c_T. \tag{7}$$

Letting $L_1(T,K)$ be the left-hand side of (7),

$$\begin{aligned} L_1(0,K) &\equiv \lim_{T\to 0} L_1(T,K) = 0, \\ L_1'(T,K) &= (c_K - c_T)r'(T,K)\int_0^T W(t,K)\mathrm{d}t \\ &\quad + \Big[\lambda'(T)C(T,K) + \lambda(T)C'(T,K)\Big]\int_0^T W(t,K)\mathrm{d}t. \end{aligned}$$

Thus, if $r(t,K)$ and $\lambda(t)$ increase with t, then the left-hand side of (7) increases with t from 0. Therefore, there exists a unique optimal T^* $(0 < T^* \leq \infty)$ which satisfies (7), and the resulting cost rate is

$$C(T^*,K) = (c_K - c_T)r(T^*,K) + \lambda(T^*)C(T^*,K).$$

2.2.2 Optimal K^*

When $c_K < c_T$, we find an optimal T^* which minimizes $C(T,K)$ in (6) for a given T. Letting $w(t,x)$ be a density function of $W(t,x)$, *i.e.*, $w(t,x) \equiv \mathrm{d}W(t,x)/\mathrm{d}x$. Then, differentiating $C(T,K)$ with respect to K and setting it equal to zero,

$$(c_T - c_K)\left[Q_1(T,K)\int_0^T W(t,K)\mathrm{d}t - W(T,K)\right] + \int_0^T \Big[Q_2(T,K) - \lambda(t)C(t,K)\Big] W(t,K)\mathrm{d}t = c_K, \tag{8}$$

where

$$Q_1(T,K) \equiv \frac{w(T,K)}{\int_0^T w(t,K)\mathrm{d}t},\ Q_2(T,K) \equiv \frac{[c_S + c_M(K)]\int_0^T \lambda(t)w(t,K)\mathrm{d}t}{\int_0^T w(t,K)\mathrm{d}t}.$$

Letting $L_2(T,K)$ be the left-hand side of (8),

$$\begin{aligned} L_2(T,0) &\equiv \lim_{K\to 0} L_2(T,K) = 0, \\ L_2'(T,K) &= (c_T - c_K)Q_1'(T,K)\int_0^T W(t,K)\mathrm{d}t \\ &\quad + Q_2'(T,K)\int_0^T W(t,K)\mathrm{d}t. \end{aligned}$$

Thus, if $Q_1(T,K)$ and $Q_2(T,K)$ increase with K, then the left-hand side of (8) increases with K from 0. Therefore, there exists a unique optimal K^* $(0 < K^* \leq \infty)$ which satisfies (8), and the resulting cost rate is

$$C(T,K^*) = (c_T - c_K)Q_1(T,K^*) + Q_2(T,K^*).$$

2.2.3 Optimal $\widetilde{T}^*$

When $c_K = c_T$, putting that $K = \infty$ in (6), the expected cost rate is

$$\widetilde{C}(T) \equiv \lim_{K\to\infty} C(T,K) = \frac{1}{T}\left[\int_0^T \lambda(t)C(t,\infty)\mathrm{d}t + c_T\right], \tag{9}$$

From (7), if $\lambda(t)$ increases with t, then an optimal tenuring collection time $\widetilde{T}^*$ which minimizes (9) is given by a unique solution of the equation

$$\int_0^T \Big[\lambda(T)C(T,\infty) - \lambda(t)C(t,\infty)\Big]\mathrm{d}t = c_T, \tag{10}$$

and the resulting cost rate is

$$\widetilde{C}(\widetilde{T}^*) = \lambda(\widetilde{T}^*)C(\widetilde{T}^*,\infty).$$

In particular, when $\lambda(t) = \lambda$, (10) becomes

$$\int_0^\infty \left[\int_0^T \big[W(t,x) - W(T,x)\big]\mathrm{d}t\right]\mathrm{d}c_M(x) = \frac{c_T}{\lambda}, \tag{11}$$

which increases with T.

2.2.4 Optimal $\widetilde{K}^*$

When $c_K = c_T$, putting that $T = \infty$ in (6), the expected cost rate is

$$\widetilde{C}(K) = \lim_{T\to\infty} C(T,K) = \frac{\int_0^\infty \lambda(t)C(t,K)W(t,K)\mathrm{d}t + c_K}{\int_0^\infty W(t,K)\mathrm{d}t}. \tag{12}$$

From (8), if $Q_2(\infty,K)$ increases with K, then an optimal tenuring collection time $\widetilde{K}^*$ which minimizes (12) is given by a unique solution of the equation

$$\int_0^\infty \Big[Q_2(\infty,K) - \lambda(t)C(t,K)\Big]W(t,K)\mathrm{d}t = c_K, \tag{13}$$

and the resulting cost rate is

$$\widetilde{C}(\widetilde{K}^*) = Q_2(\infty,\widetilde{K}^*).$$

In particular, when $\lambda(t) = \lambda$, (13) becomes

$$\int_0^\infty \left[\int_0^K W(t,x)\mathrm{d}c_M(x)\right]\mathrm{d}t = \frac{c_K}{\lambda}, \tag{14}$$

which increases with K.

Table 1: Optimal T^* and $C(T^*,K)/c_T$ when $c_S/c_T = 0.1$ and $\lambda = \mu = \sigma = 1$.

		$c_M/c_T = 0.01$		$c_M/c_T = 0.05$		$c_M/c_T = 0.1$	
K	c_K/c_T	T^*	$C(T^*,K)/c_T$	T^*	$C(T^*,K)/c_T$	T^*	$C(T^*,K)/c_T$
5	2	4.73	0.4723	4.09	0.5462	3.57	0.6324
	3	3.25	0.5473	3.06	0.6087	2.87	0.6822
	4	2.80	0.5891	2.69	0.6453	2.56	0.7133
	5	2.57	0.6191	2.49	0.6718	2.40	0.7361
10	2	7.43	0.2990	5.77	0.4285	4.50	0.5586
	3	6.37	0.3160	5.40	0.4339	4.42	0.5596
	4	5.91	0.3252	5.18	0.4378	4.36	0.5604
	5	5.63	0.3320	5.03	0.4408	4.31	0.5612

Table 2: Optimal K^* and $C(T,K^*)/c_K$ when $c_S/c_K = 0.1$ and $\lambda = \mu = \sigma = 1$.

		$c_M/c_K = 0.01$		$c_M/c_K = 0.05$		$c_M/c_K = 0.1$	
T	c_T/c_K	K^*	$C(T,K^*)/c_K$	K^*	$C(T,K^*)/c_K$	K^*	$C(T,K^*)/c_K$
5	2	4.35	0.4677	3.77	0.5342	3.23	0.6063
	3	3.45	0.5500	3.14	0.6048	2.80	0.6661
	4	3.01	0.6153	2.79	0.6633	2.54	0.7183
	5	2.67	0.6719	2.55	0.7155	2.36	0.7659
10	2	7.43	0.2973	5.30	0.4209	3.92	0.5328
	3	6.37	0.3202	4.93	0.4221	3.80	0.5390
	4	5.80	0.3372	4.67	0.4318	3.71	0.5448
	5	5.42	0.3512	4.48	0.4405	3.62	0.5503

3 Numerical Examples

We compute numerical examples of the models discussed above for $Z(t) = \mu t + \sigma B(t)$ when $B(t)$ is normally distributed with mean 0 and variance t or for $Z(t) = A(t)t$ when $A(t)$ is normally distributed with mean μ and variance σ^2/t, that is,

$$W(t,x) = \Phi\left(\frac{x-\mu t}{\sigma\sqrt{t}}\right), \tag{15}$$

where $\Phi(x)$ is the standard normal distribution with mean 0 and variance 1, *i.e.*, $\Phi(x) \equiv (1/\sqrt{2\pi})\int_{-\infty}^{x} \mathrm{e}^{-u^2/2}\mathrm{d}u$.

Table 3: Optimal $\widetilde{T}^*$, $\widetilde{C}(\widetilde{T}^*)/c_T$, $\widetilde{K}^*$ and $\widetilde{C}(\widetilde{K}^*)/c_K$ when $\lambda = \mu = \sigma = 1$.

c_M/c_i	$c_S/c_T = 0.1$		$c_S/c_K = 0.1$	
$i{=}T,K$	$\widetilde{T}^*$	$\widetilde{C}(\widetilde{T}^*)/c_T$	$\widetilde{K}^*$	$\widetilde{C}(\widetilde{K}^*)/c_K$
0.01	14.24	0.2424	14.15	0.2413
0.02	10.11	0.3021	9.99	0.2997
0.03	8.28	0.3486	8.15	0.3444
0.04	7.20	0.3882	7.05	0.3820
0.05	6.46	0.4233	6.30	0.4151
0.06	5.91	0.4554	5.75	0.4449
0.07	5.49	0.4851	5.32	0.4722
0.08	5.15	0.5130	4.97	0.4976
0.09	4.86	0.5394	4.68	0.5114
0.10	4.62	0.5546	4.44	0.5239

References

[1] R. Jones and R. Lins. Garbage Collection: Algorithms for Automatic Dynamic Memory Management. John Wiley & Sons, Chichester, 1996.

[2] D. Ungar. Generation scavenging: A non-disruptive high performance storage reclamation algorithm. *ACM Sigplan Notices* 19, 157–167, 1984.

[3] A. W. Appel. Simple generational garbage collection and fast allocation. *Software Practice and Experience*, 19, 171–183, 1989.

[4] P. R. Wilson. Uniprocessor garbage collection techniques. In: International Workshop on Memory Management, Vol. 637 in Lecture Notes in Computer Science, 1–42, Springer, 1992.

[5] D. Vengerov. Modeling, analysis and throughput optimization of a generational garbage collector. Technical Report, TR–2009–179, Sun Labs, 2009.

[6] D. Ungar and F. Jackson. An adaptive tenuring policy for generation scavengers. *ACM Transactions on Programming Languages and Systems*, 14, 1–27, 1992.

[7] A. Kaldewaij, and L. Vries. Optimal real-time garbage collection for acyclic pointer structures. *Information Processing Letters*, 77, 151–157, 2001.

[8] W. H. Lee and J. M. Chang. A garbage collection policy based on empirical behavior. *Information Sciences*, 167, 129–146, 2004.

[9] W. D. Clinger and F. V. Rojas. Linear combinations of radioactive decay models for generational garbage collection. *Science of Computer Programming*, 62, 184–203, 2006.

[10] S. Soman and C. Krintz. Application-specific garbage collection. *Journal of Systems and Software*, 80, 1037–1056, 2007.

[11] T. Satow, K. Yasui and T. Nakagawa. Optimal garbage collection policies for a database in a computer system. *RAIRO Operations Research* 30, 359–372, 1996.

[12] X. F. Zhao, S. Nakamura and T. Nakagawa. Optimal policies for random and periodic garbage collections with tenuring threshold. In: Proceedings of the 2nd International Conference on Advanced Science and Technology, Vol. 74 of Communications in Computers and Information Science, 125–135, Springer, 2010.

[13] M.S. Nikulin, N. Balakrishnan. Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finanace. Birkhöuser, Boston, 2010.

[14] T. Nakagawa. Shock and Damage Models in Reliability Theory. Springer, London, 2007.

[15] D.B. Kececioglu, M.X. Jiang, F.B. Sun. A unified approach to random-fatigue reliability quantification under random loading. In:Procedings of Annual Reliabibility Maintainability Symposium, 308–313, 1998.

[16] W.J. Owen, W.J. Padgett. Accelerated test models with the Birnbaum- Saunders distribution. In: Pham H (ed) Handbook of Reliability Engineering, 429–439, Springer, London, 2003.

[17] C. Park, W.J. Padgett. New cumulative damage models for failure using stochastic processes as initial damage. *IEEE Trans. Reliab.* 54, 530–540, 2005.

[18] K. Sato. Basic results on Lévy processes. In: Bandorff-Nielsen O, Mikosch T, Resnick S (eds) Lévy Processes, Theory and Applications, Birkhöuser, Boston, 2001.

[19] S.M. Ross. Stochastic processes. John Wiley & Sons, New York, 1983.

[20] R.E. Barlow and F. Proschan. Mathematical Theory of Reliability. John Wiley & Sons, New York, 1965.