

К вопросу об обеспечении бесперебойной работы электронного архива
07, июль 2011

автор: Марков А. С.

УДК 004.056:681.5

МГТУ им. Н.Э.Баумана

ЗАО «НПО «Эшелон»

mail@cnpo.ru

Эффективным средством обеспечения бесперебойной работы информационных хранилищ и электронных архивов (ЭА) [1] является периодический контроль файловых систем и баз данных (БД), который включает выполнение стандартных процедур обнаружения ошибок, а так же копирования и восстановления БД. Известны два подхода к организации контроля БД [2-5]. Первый основывается на периодической проверке БД через детерминированные промежутки времени. Это согласуется с календарными планами работы информационных хранилищ и позволяет учесть имеющиеся ограничения на вычислительные ресурсы, однако не полностью учитывает стохастическую природу возникновения ошибок БД.

Второй подход связан с возникновением некоторого события (прерывания вычислительного процесса, инициирование обмена данных между задачами и др.). Недостатками указанного подхода являются трудности при определении подмножества контролируемых событий вычислительного процесса и возможность неограниченного роста моментов контроля. Последнее делает его недопустимым в нормальный период функционирования.

В связи со сказанным, представляет интерес подход, сочетающий учет стохастического характера ошибок с ограничениями на число моментов контроля.

Стохастическая модель контроля баз данных

Рассмотрим работу ЭА с учетом проводимого периодического контроля БД. Поток ошибок БД считаем простейшим с плотностью распределения интервала τ между ошибками

$$g_{\tau} = \lambda e^{-\lambda \tau},$$

где λ - интенсивность ошибок.

Поскольку период времени t функционирования системы ЭА во много раз превышает время контроля БД, положим последнее мгновенным. Тогда сохранность БД ЭА характеризуется вероятностью $P(\xi < Q) = F_{\xi}(Q)$ того, что время ξ обнаружения ошибки в межконтрольном интервале T не превосходит допустимое время Q работы ЭА при наличии ошибки. Вероятность P_n обнаружения всех ошибок БД за t период времени работы ЭА можно представить в виде

$$P_n = F_{\xi}(n+1), \quad (1)$$

где n - число моментов контроля.

Диаграмма периодического контроля БД представлена на рис.1 [3].

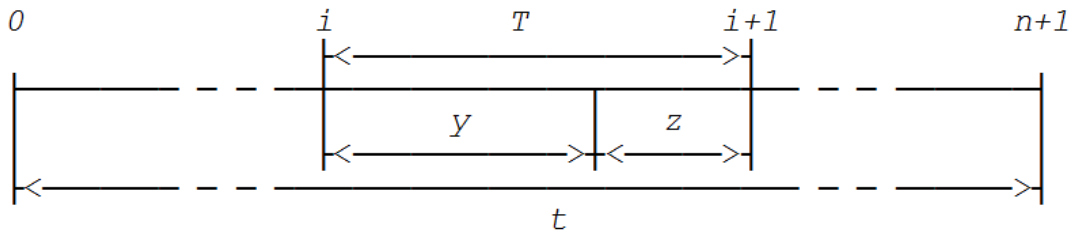


Рис. 1. Диаграмма периодического контроля базы данных

Зададим стохастическую модель обнаружения ошибок БД. В этом случае контроль производится определенное количество раз с равной вероятностью независимо друг от друга. Таким образом, образованный всеми моментами контроля ограниченный поток является потоком Бернулли с плотностью распределения интервала \tilde{T} между моментами контроля

$$f_{\tilde{T}} = n/t(1 - T/t)^{n-1}.$$

Величина задержки $\xi = \tilde{T} - \vartheta$ является функцией двух случайных величин и имеет функцию распределения

$$F_{\xi}^c = \iint_{(s)} n/t(1 - T/t)^{n-1} \lambda e^{-\lambda y} dT dy.$$

Определив пределы интегрирования (рис. 2), получим

$$F_{\xi}^c = \int_0^{t-z} \left(\int_y^{y+z} n/t(1 - T/t)^{n-1} \lambda e^{-\lambda y} dT \right) dy + \int_{t-z}^t \left(\int_y^t n/t(1 - T/t)^{n-1} \lambda e^{-\lambda y} dT \right) dy. \quad (2)$$

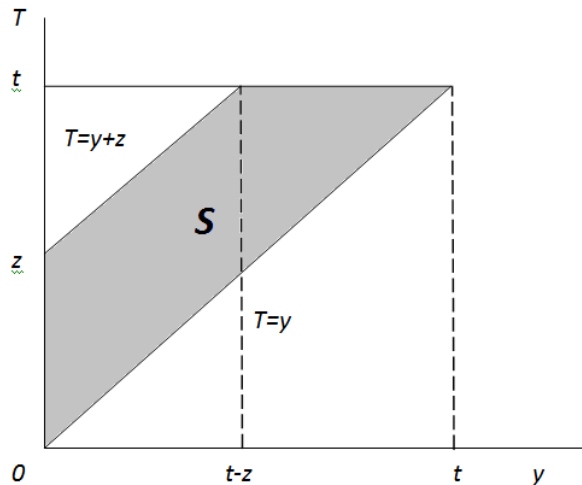


Рис. 2. Область интегрирования интервала задержки времени обнаружения ошибки

После упрощения выражение (2) примет вид

$$F_{\Sigma}^z = \lambda/t^n \left(\int_0^{t-z} e^{-\lambda y} ((t-y)^n - (t-z-y)^n) dy + \int_{t-z}^t e^{-\lambda y} ((t-y)^n) dy \right). \quad (3)$$

Разложив в степенной ряд подынтегральные выражения формулы (3), получим приближенное значение функции распределения

$$F_{\Sigma}^z = \lambda \sum_{i=0}^n \sum_{j=0}^n \sum_{l=1}^{n+j+1} (-1)^{i+j+l+1} C_n^i C_{n+j+1}^l \frac{\lambda^i t^{i+j+1} z^j}{j!(1+j+i)}, \quad (4)$$

которое и является основным расчетным соотношением. Здесь r - число итераций.

Детерминированная модель контроля баз данных

Для сравнения стохастической модели обнаружения ошибок БД (4) с детерминированной рассмотрим последнюю подробнее. В детерминированной модели моменты контроля образуют регулярный поток с постоянной величиной интервала $T=t/(n+1)$ и плотностью распределения времени обнаружения ошибки

$$g_{\Sigma}^z = \lambda e^{-\lambda(T-z)}; \quad 0 < z < T.$$

Можно показать, что выражение для функции распределения в детерминированной модели имеет вид

$$F_{\Sigma}^z = e^{-\lambda T} (e^{\lambda z} - 1); \quad 0 < z < T. \quad (5)$$

Сравнивая выражения (4) и (5), получаем критерий

$$F_{\Sigma}^z(x) \leq F_{\Sigma}^z(x),$$

позволяющий сделать выбор той или иной модели.

Следовательно, при конкретных значениях λ , Q , t и n можно говорить о соответствии рассмотренных моделей процессу появления ошибок БД.

Проведенный автором сравнительный анализ рассмотренных моделей показал преимущество стохастической модели при небольшом числе моментов контроля. Это может быть объяснено тем, что даже при малом числе случайных моментов контроля ресурсов и процессов информационно-вычислительной системы всегда существует вероятность того, что обнаружение ошибки будет произведено сразу же при ее возникновении. При использовании детерминированной модели такой возможности практически нет. Очевидно, что при ограничениях на Q и n численными методами можно определить предпочтительную модель (стохастическую, детерминированную либо комбинированную), ведущую к повышению сохранности информации в ЭА.

Это дает эффект, подобный введению структурной избыточности, то есть можно говорить об особом виде избыточности - стохастической, использование которой практически не увеличивает затрат [6].

Пример использования моделей контроля

Для примера использовались данные одного ведомственного архива за первую половину 2010 года. За это время контроль БД проводился 7 раз, в результате чего было выявлено 12 ошибок, все из которых были восстановлены стандартными средствами, что зафиксировано в журнале администратора. Расчет характеристик ошибок показал следующее:

- среднее время между ошибками $M_2 = 43,83$ час;
- интенсивность ошибок $\lambda = 0,022$ 1/час;
- среднее квадратическое отклонение $\sigma_2 = 30,04$ час;
- критерий согласия Крамера фон Мизиса $k(n) = 0,55$.

Это позволило считать поток ошибок стационарным пуассоновским. Исследование функционирования БД ЭА за период 2000-2010 гг. показало, что ограничение на время Q обнаружения восстанавливаемой ошибки составляет более 1 месяца.

Значения вероятностей восстановления БД, рассчитанные по формулам (1), (4), (5), и зависимость их от числа моментов контроля за первую половину 2010 года ($t = 1052$ час) показаны на рис. 3.

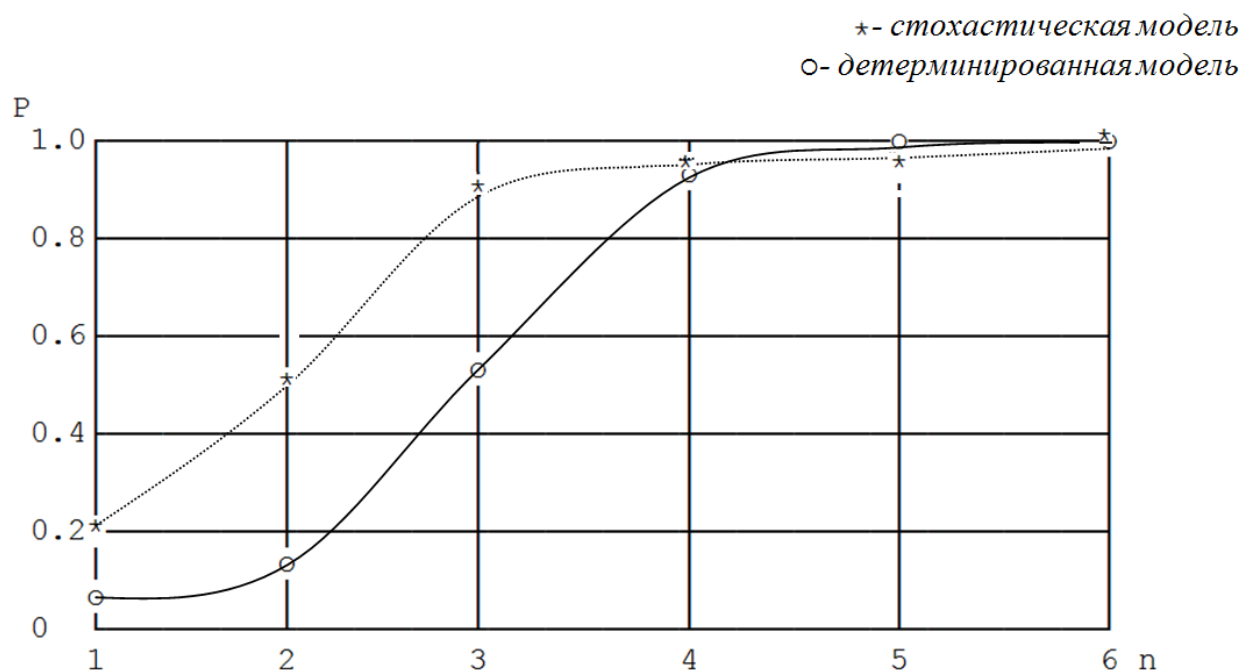


Рис. 3. График зависимости значений вероятности обнаружения ошибок БД от числа моментов контроля

Учитывая требования к готовности функционирования ЭА, при использовании стохастической модели целесообразно ограничиться всего 3 моментами контроля (табл. 1).

Таблица 1

Результаты расчета характеристик контроля

Тип контроля	Число моментов контроля	Затраты (чел/час)	Вероятность восстановления
Традиционный	7	42	1.0
Детерминированный	4	24	0.99996
Стохастический	3	18	0.99997

Заключение

Таким образом, предложенные в работе практические решения учитывают стохастическую природу ошибок БД. Это позволяет выбрать необходимую модель обнаружения ошибок при заданных параметрах БД и ЭА.

Аналогичный подход был положен в основу решения задач построения механизмов резервирования (контрольных точек), предупреждения и обнаружения ошибок

программных средств [7-10]. Для удобства использования стохастической модели в технических системах разработан генератор случайных импульсов, позволяющий формировать случайные ограниченные потоки Бернулли [11].

Помимо контроля информационного обеспечения ЭА изложенные результаты могут быть полезны и в других видах технического контроля и диагностики, а также при решении обратных задач, например, для оценки экономической эффективности реализуемых мер и механизмов безопасности в системе менеджмента информационной безопасности организации [12, 13].

Список литературы

1. Марков А.С. Концепция построения электронного архива // Открытые системы. СУБД. 1997. № 1. С.54-58.
2. Обеспечение надежности сложных технических систем / А.Н.Дорохов [и др.] СПб.: Лань, 2011. 352 с.
3. Кузовлев В.И., Шкатов П.Н. Математические методы анализа производительности и надежности САПР. М.: Высшая школа, 1990. 142 с.
4. Розенберг Е.Н., Шубинский И.Б. Методы и модели функциональной безопасности технических систем. М.: ВНИИАС, 2004. 188 с.
5. Обеспечение безопасности информации в центрах управления полетами КА / Л.М.Ухлинов [и др.] М.: Изд-во МГТУ им. Н.Э.Баумана, 2000. 366 с.
6. Марков А.С. Решение вычислительной задачи при наличии временных ограничений // Известия высших учебных заведений. Приборостроение. 1992. Т. 35, № 5. С. 54-57.
7. Керножицкий В.А., Марков А.С. Об одном подходе к управлению механизмом контрольных точек // Гироскопия и навигация. 1993. № 1. С. 77-80.
8. Керножицкий В.А., Марков А.С. Предварительный контроль программного обеспечения ЛВС // Известия высших учебных заведений. Приборостроение. 1994. Т. 37, № 6. С. 89-92.
9. Markov A.S., Kernozitsky V.A. Economically effective data bases diagnostics method // Advances in Modeling and Analysis. В: Signals, Information, Data, Patterns. 1995. V.33, № 3. P. 5-11.
10. Марков А.С. Метод ограниченного стохастического контроля бортового программного обеспечения // Труды Международной авиакосмической конференции "Человек-земля-космос". М.: Российская инженерная академия, Секция "Авиакосмическая". 1996. Т. 6. С. 344-349.

11. Генератор случайных импульсов: А.с. № 840856 СССР, М.КлЗ, G 07 C 15/00. / В.А.Керножицкий [и др.] Заявл. 06.10.78; опубл. 25.06.81. Бюлл.23.

12. Медведев Н.В., Квасов П.М., Цирлов В.Л. Стандарты и политика информационной безопасности автоматизированных систем // Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия: Приборостроение. 2010. № 1. С. 103-111.

13. Fang W-C. Security Technology, Disaster Recovery and Business Continuity. New York: Springer, 2010. 300 p.