computing@computingonline.net
www.computingonline.net

ISSN 1727-6209
International Journal of Computing

MODELING AND OPTIMIZATION OF PARALLEL INFORMATION SEARCHING IN FILES

Volodymyr Lisovets, Hryhoriy Tsehelyk

Ivan Franko National University of L'viv,
1, Universytetska str., Lviv, 79000, Ukraine,
kafmmsep@franko.lviv.ua

Abstract: *In this article the m-parallel method of sequential field searching and two variants of m-parallel block field searching method are offered. These methods are oriented to be used in multiprocessing system for information searching in files of database. We research the effectiveness of these methods for different probability distribution law of field access. The mathematical expectation of number of parallel comparisons necessary for field searching in files is taken as a criterion of effectiveness. The effectiveness of the methods is compared and analyzed. The best of offered methods is founded for every considered probability distribution. Optimal strategies of field searching in sequenced files stored in external memory of multiprocessing system are made. In this case the mathematical expectation of total time needed for field searching in files is taken as a criterion of effectiveness.*

Keywords: *multiprocessing system, mathematical modeling, parallel searching, database.*

1. THE METHODS OF PARALLEL SEARCHING AND THEIR EFFECTIVENESS

The following methods of parallel field searching in files of database for multiprocessing system are considered [1-4]:

- method of *m*-parallel sequential field searching;
- the first variant of *m*-parallel block field searching method;
- the second variant of *m*-parallel block field searching method.

We analyze the effectiveness of these methods for different probability distribution law of field access (discrete uniform, binomial, Zipf and generalized the partial occasion of witch is the probability distribution approximately satisfying the rule "80 – 20" [5-7]). The mathematical expectation of number of parallel comparisons necessary for field searching in files is taken as a criterion of effectiveness.

Let's consider the method of *m*-parallel sequential field searching.

Suppose that multiprocessing system consists of *m* processors working parallel and having common memory. Let's enumerate the processors by natural numbers from 1 to *m*. The main point of the method of *m*-parallel sequential field searching is the following. Divide conventionally all fields of file into blocks and each of them includes *m* fields. Let $N = n \cdot m$ is the number of fields in file, where *n* is the number of blocks. When method of *m*-parallel sequential field searching is used the field searching will consist of the next steps. On the first step the processor number *i* searches the value of the key of field number *i*. In this case the process of searching can be successful or failed. Each processor must exchange data with other processors on every step. In case of successful searching the field searching process is finished. In case of failed searching the processor number *i* searches the value of the key of filed number $(m + i)$ on the second step etc. If the field searching is failed on the step number *k* then on the step number $(k + 1)$ the processor number *i* searches the value of the key of filed number $(km + i)$. If the required field is located in the file then it will be found after *n* steps. If p_i is probability of access to field number *i* then mathematical expectation *E* of number of parallel comparisons necessary for field searching in files is calculated by following formula

$$E = \sum_{i=1}^{n} \sum_{j=1}^{m} i p_{(i-1)m+j} \cdot$$

In case of using the first variant of *m*-parallel block field searching method we suppose that the fields of ordered file are divided into *n* blocks each

of them includes *sm* fields. Then the filed searching will be done in this way. First of all we search the block including the needed field by looking the last *m* fields of file. After that we continue the searching in the located block by the method of *m*-parallel sequential field searching. The mathematical expectation *E* of number of parallel comparisons necessary for field searching in files lets write as the sum of mathematical expectation of number of parallel comparisons necessary for block location which includes the needed field and mathematical expectation of number of parallel comparisons necessary for field searching in the located block. Then

$$E = \sum_{k=1}^{n}\sum_{i=1}^{s}\sum_{j=1}^{m}(k+i)\,p_{(k-1)ms+(i-1)m+j}\ .$$

In case of using the second variant of *m*-parallel block field searching method let suppose that the fields of ordered file are divided into *nm* blocks and each of them includes *sm* fields. Then the number of fields of file will be $N = snm^2$. Then the field searching will be done in the following way. First of all we search the block including the needed field by using the method of *m*-parallel sequential searching among the last elements of blocks. After that we continue the searching in the located block by the method of *m*-parallel sequential searching. The mathematical expectation *E* of number of parallel comparisons necessary for field searching in files lets write as the sum of mathematical expectation of number of parallel comparisons necessary for block location which includes the needed field and mathematical expectation of number of parallel comparisons necessary for field searching in the located block:

$$E = \sum_{k=1}^{n}\sum_{l=1}^{m} k\left(\sum_{i=1}^{s}\sum_{j=1}^{m} p_{(k-1)m^2s-(l-1)ms+(i-1)m+j}\right) + \sum_{k=1}^{nm}\sum_{i=1}^{s}\sum_{j=1}^{m} ip_{(k-1)ms+(i-1)m+j}$$

2. THE OPTIMAL STRATEGIES

Using the method of *m*-parallel sequential field searching we create the optimal strategies of parallel information searching in sequenced ordered files of database that are stored in external memory of multiprocessing system. Suppose that multiprocessing system consists of *m* processors working parallel and having common memory.

Suppose that file including *N* fields is divided into *n* blocks and each of them includes *ml* fields. Let $a_0 = b_0 + d_0ml$ is the time of block field reading in the RAM, where b_0, d_0 are some constants; t_0 is the one step time of the method of m-parallel sequential field searching in the RAM; p_i is the probability of access to field number *i*, E_t is the mathematical expectation of total time needed for field searching in file. Let suppose that first of all the field blocks are read to RAM from external memory and then *m*-parallel sequential searching method is used. Then

$$E_t = \sum_{k=1}^{n}\sum_{i=1}^{l}\sum_{j=1}^{m}\{ka_0 + [(k-1)l + i]t_0\}\times \times p_{(k-1)ml+(i-1)m+j}\ .$$

3. CONCLUSION

The *m*-parallel method of sequential field searching and two variants of *m*-parallel block field searching method are considered. We researched the effectiveness of these methods for different probability distribution law of field access (discrete uniform, binomial, Zipf and generalized the partial occasion of witch is the probability distribution approximately satisfying the rule “80 – 20”). The mathematical expectation of number of parallel comparisons necessary for field searching in files was taken as a criterion of effectiveness. The effectiveness of the methods was compared and analyzed. The best of offered methods was founded for every considered probability distribution.

Optimal strategies of field searching in sequenced files stored in external memory of multiprocessing system were made for different probability distribution law of field access. In this case the mathematical expectation of total time needed for field searching in files was taken as a criterion of effectiveness. The values of parameters when mathematical expectation reaches the minimum are founded.