TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES
NAMED AFTER MUHAMMAD AL-KHWARIZMI

MUHAMMAD AL-XORAZMIY NOMIDAGI
TOSHKENT AXBOROT TEXNOLOGIYALARI
UNIVERSITETI

BULLETIN OF TUIT:

MANAGEMENT AND COMMUNICATION TECHNOLOGIES

ISSN:2181-1083 DOI:
Kuchkarov Muslimjon № 3(26)2025

ALGORITHM FOR DIGITAL PROCESSING OF SEISMIC SIGNALS IN DISTRIBUTED SYSTEMS

Kuchkarov Muslimjon
Department of Artificial Intelligence,
Tashkent University of Information
Technologies named after
Muhammad al Khwarizmi, Tashkent, Uzbekistan
muslimjon1010@gmail.com

Abstract. The problem of misallocation of cache lines in processors is studied based on the specific features of memory organization in multi-core and multi-processor systems of distributed systems. Incorrect allocation of cache lines leads to sequential execution of tasks that should be executed in parallel on several cores. It is difficult to determine the problem, and its presence can cause a sharp decrease in the performance of the entire system. In many seismic studies, scientists' efforts are aimed at finding reliable signs of seismic hazard. Abrupt changes in one or another parameter are called abrupt effects, emissions or anomalies, through which the principle of prediction can be implemented - this means predicting the place, strength and time of a future seismic event. The use of distributed systems in the implementation of predictions with the help of machine learning, digital processing of large volumes of digital signals, ensures fast and high-quality determination of results. To increase the speed, it is necessary to develop parallel algorithms. Synchronization of common parallel streams is necessary to solve parallel computing problems and to ignore cache memory in parallel programming. This causes cache lines to be misallocated. Acceleration of parallel algorithms for digital processing of seismic signals and the dependence of the efficiency of using the system's computing resources on the number of parallel flows were analyzed, a parallel algorithm for determining the epicenter using cubic splines and a software tool was created based on this algorithm, and the proposed spline method was used to determine the characteristics of P and S waves. made it possible to determine the detection time 2.5 times faster than usual.

Keywords: Distributed computing systems, digital signal processing, Spline, multi-processor systems, cache memory, misallocation, multithreaded program.

I. INTRODUCTION

Over the past decade, the main direction of the development of computing systems, especially distributed computing architecture, has been the design of multi-core multiprocessor systems [1]. This was due to the need to further improve the performance of computers, as well as to create high-speed algorithms for sorting, searching and processing large amounts of data. The use of multi-processor systems allows software developers to organize parallel computations within distributed computing systems to further accelerate the execution of software tools. In the design of parallel software, in order to effectively distribute the functionality of the program among the available computing resources, it is necessary to take into account the specific characteristics of the hardware on which it is placed [10]. The following factors that may affect the performance of parallel applications should be considered: the number and configuration of processors, the creation of parallel streams for data, changing the cache, improper distribution of processor cache rows.

II. MAIN SECTION

In the last ten years, the main direction of the development of computing systems, especially the architecture of distributed computing systems, has been the design of multi-core multi-processor systems [1]. The reason for this was the need to further increase the performance of computers and, at the same time, to create fast algorithms for sorting, searching and digital processing of large volumes of data. The use of multiprocessor systems allows software developers to organize parallel computations within distributed computing systems to further accelerate the execution of software tools. When designing parallel software tools, it is necessary to take into account the specific characteristics of the hardware on which it will be placed in order to effectively distribute the functionality of the program among the available computing resources. Electromagnetism, anomalous changes in the gravity field, anomalous disturbances in the ionosphere, seismic noise, various acoustic oscillations, and others can be sudden changes. In recent years, dozens of new methods for predicting seismic events have been proposed. In order to understand the results obtained by these methods, the physical processes applied to the seismic event, it is necessary to create physical and mathematical models of the interdependence of the processes necessary for the practical implementation of the principles of forecasting.

Organization of cache memory of local network computers for earthquake source detection. Each processor is equipped with a local cache, which is much faster than the main memory, to reduce the latency of access to the shared memory. Using cache memory avoids the need to access main memory in multiprocessor systems [2]. The cache memory of multi-core processors has several levels [3]:

- first-level cache (L1) works directly with the processor core, has the shortest access time. It is divided into instruction cache (L1i) and data cache (L1d);

- the second-level cache (L2) also belongs to a specific processor core. This cache is larger and slower than the first-level cache;

- third-level cache (L3) is the largest and slowest, but still works much faster than

RAM. L3 is shared between all processor cores (Intel)

2. The problem of preserving the integrity of cached data when determining the epicenter of an earthquake

Caching shared data requires maintaining its integrity in the cache. If any read operation of a data element performed by any processor returns the last value of this element written by any processor, the memory system preserves the integrity of the data [5]. Thus, there is a need to synchronize cache lines [2]. Changes in one processor cache must be propagated to another processor cache. The value of a shared variable can be repeatedly passed from one cache to another. This phenomenon is called cache modification and can seriously affect the performance of the application [6].

Based on the records (seismogram) recorded at the point, the difference between the arrival times of R and S waves is determined using a hodograph, and on this basis, the distance from the station to the epicenter is determined. This is done as follows: It is known that R-waves arrive at the observation point before S-waves due to their high propagation speed. Let T be the interval of arrival of R- and S-waves measured from the seismogram at the point (determined from the hodograph). The velocities of R and S waves are VP and VS, respectively, and the distance d from the station to the epicenter is determined by the following formula.

For example:

For station A:

$$d = T \cdot \frac{V_p \cdot V_s}{V_p - V_s} \quad V_p = 6\,км/сек;\ V_s = 4\,км/сек$$

$$T = 25\,сек,\ d_A = \frac{6 \cdot 4}{6-4} \cdot 25 = 300\,км$$

For station B:

$$T = 50\,сек,\ d_Б = \frac{6 \cdot 4}{6-4} \cdot 50 = 600\,км$$

A to the epicenter d = 300 km, from station B is d = 600 km. If the observation was made at several stations, that is, d1, d2, d3 are known, then circles are drawn between these distances, and the circles intersect at one point or may not intersect due to inaccuracies in the calculation. . In that case, the center of gravity of the triangle (polygon) formed by the intersection of circles is taken as the location point of the epicenter. During the research, ways to eliminate the problems of efficient allocation of cache memory in the creation of software tools working with parallel streams were identified. Using shared thread synchronization to solve parallel computing problems in multi-threaded software led to misallocation of cache lines. Experiments were conducted on a server equipped with two eight-core Intel Core i5 2.9 GHz processors. Processor cache: L1 - 384 Kb, L2 - 1.5 Mb, L3 - 9 Mb. Operating memory 8 GB. Operating system: Windows 10 Pro. Parallel computing processes were implemented in the Python programming language (Python 3.10.4).

The following indicators are used for analysis. Acceleration is the ratio of the serial execution time of a program to the parallel execution time of the program [8]:

$$S_N = T_1 / T_N \quad (1)$$

Efficiency - The program shows how well the system uses computing resources. To calculate the efficiency of a parallel program, the acceleration observed in the experimental processor must be divided by the number of available cores N.

$$E_N = S_N / N * 100\ \% \quad (2)$$

Parallel currents run on different cores, meaning that the number of cores in the processor is the same as the total number of cores.

3. A method of parallel calculation of spline function in distributed systems

A parallel method of calculating the spline function on a multi-core processor consists of the following sequences. •The four pairs of multiplications presented in the spline are converted into an array for parallel computation on individual processor cores.

For example:

$$b_0\ B_0(x),\ K_j = \sum_{i=0}^{m+1} b_i B_{(i \bmod 10)+20}(x),$$

$$b_1\ B_1(x),\ P_j = \sum_{i=0}^{m+1} b_{i+1} B_{(i \bmod 10)+10}(x),$$

$b_2\ B_2(x)$ is brought to $T_j = \sum_{i=0}^{m+1} b_{i+2} B_{(i \bmod 10)}(x)$.

Sums of four arrays are calculated in parallel after one clock cycle of the processor [7].

The problem of predicting the time and place of earthquakes in advance due to its extreme difficulty (the need to obtain information about the processes in the interior of the earth, the need to obtain information about the low speed of differential tectonic movements that lead to earthquakes) is still incomplete. not resolved internally.

$$P = \max |S(x_i)|, \quad i = 0,1,\ldots,N. \quad (3)$$

S is the maximum value of the digitally processed signal based on the parallel algorithm, which is less than P. S is calculated by the following formula:

$$S = \max |S(x_i)| < P, \quad i = 0,1,\ldots,N. \quad (4).$$

Where is a function approximated by a cubic spline.

Fig. 1 presents a diagram of a digitally processed seismic signal based on the proposed parallel algorithm, and the maximum and minimum values of this signal are marked with the letters P and S[9]. Using the special parameters of the seismic signal determined by seismologists, parameters P and S are defined to determine the range of changes that occurred in the signal. In order to detect changes in other parts of the signal, apart from the parts defined in the algorithm, the software package has the ability to move the P and S parameters along the x and y coordinate axes. Figure 2 shows the block diagram of the parallel algorithm for determining P and S parameters of seismic signals using cubic splines.

Shared memory allocation for parallel streams in distributed systems. Another approach to parallel programming is to create sub-threads for central parallel threads that run independently in shared memory. The main parallel thread allocates a single block of memory for all structures of these threads before creating sub-parallel threads.

Each sub flow can access elements of its own structure. The structures of all parallel streams are located next to each other in memory, and most likely the elements of these structures fall into the same cache line.

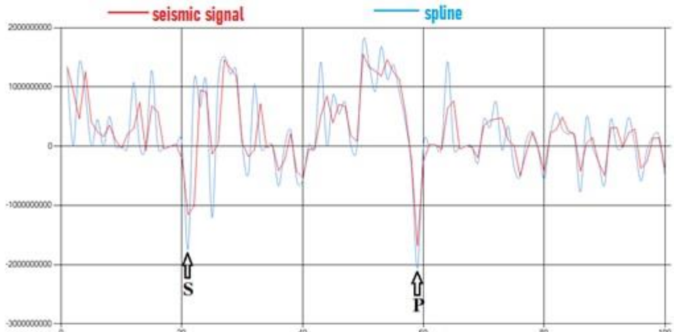

Fig 1. Results of determination of P and S parameters of the seismic signal

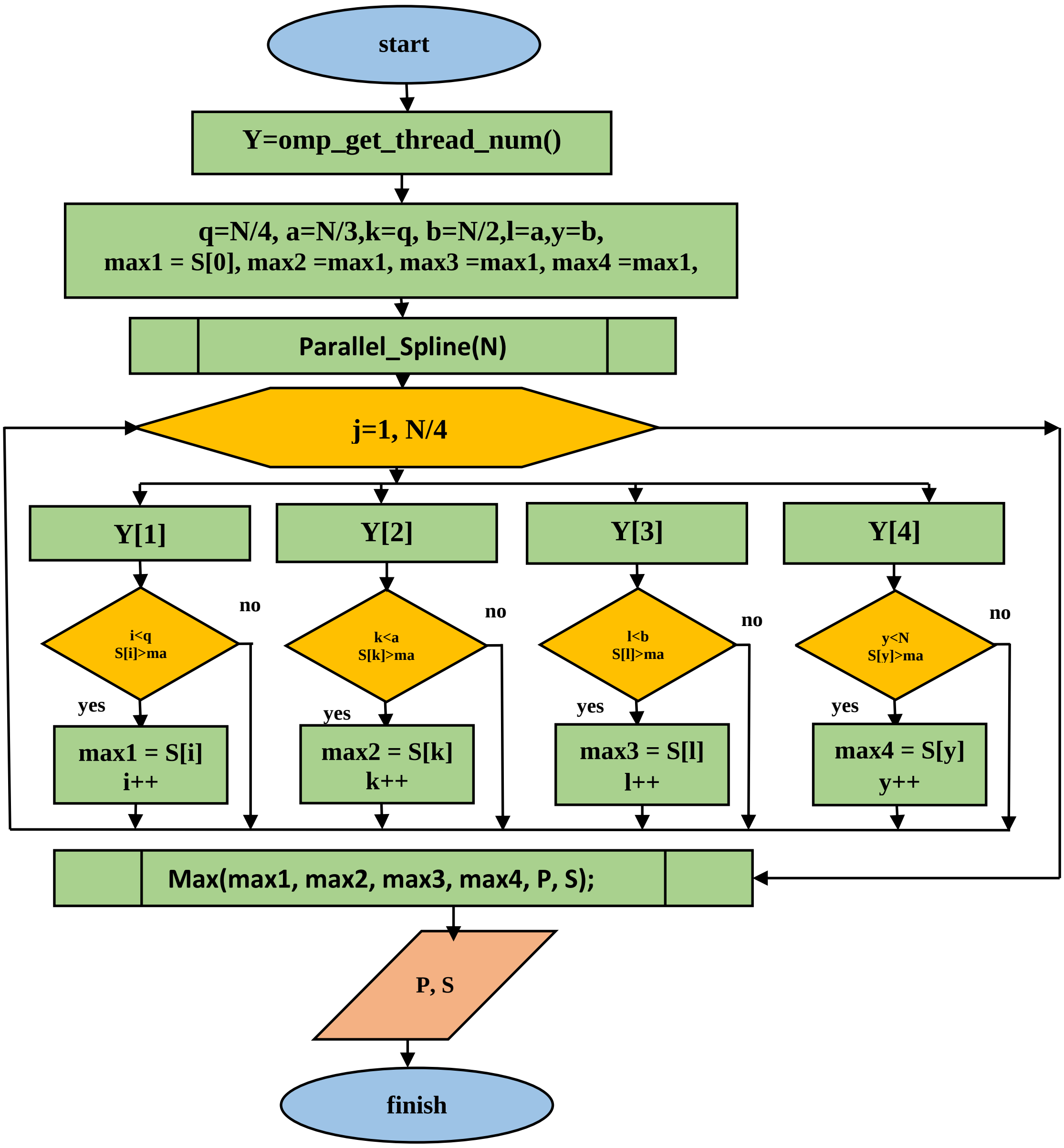

Fig 2. Block diagram of the algorithm for determining P and S parameters of seismic signals in distributed systems

Although parallel threads access unrelated elements of different structures, they cause misallocation of cache lines. In the blue line in Figure 1, it can be seen that the time characteristics of the program have significantly deteriorated with the increase in the number of parallel streams, and in the red line, the execution time of the program in each parallel stream due to the increase in the number of parallel streams and the correct allocation of the cache line was almost the same.

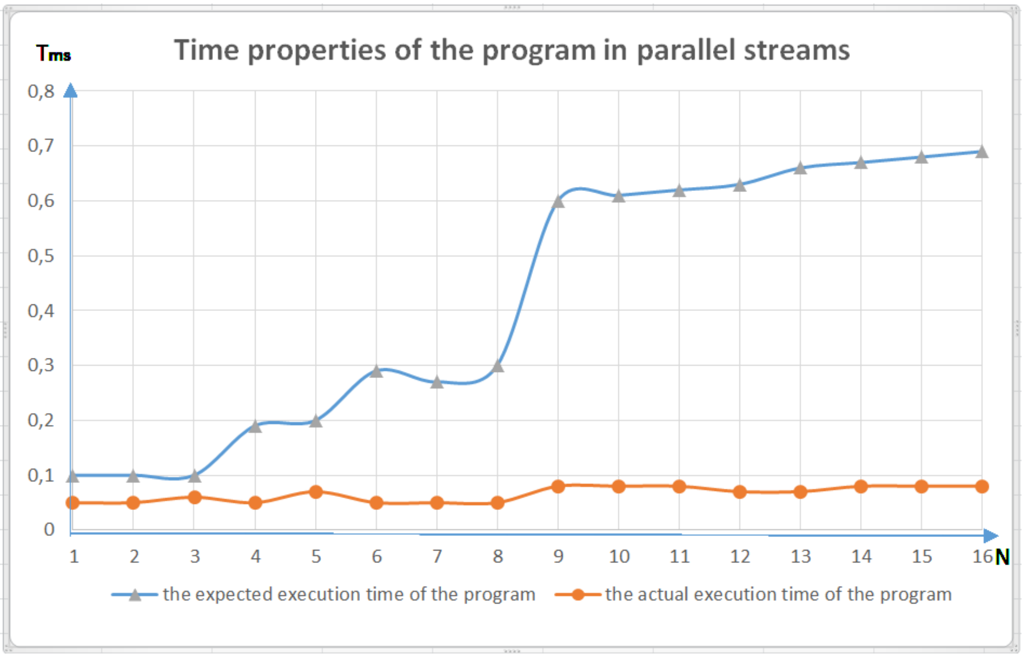

Fig 3. Timing characteristics of the program in parallel streams.

In Figure 3, line 1 shows the actual execution time of the program, line 2 shows the execution time of the program after the correct allocation of the cache line. The OY axis of the diagram in Figure 1 represents the execution time (T) of the program, and the OX axis represents the number of cores (common or main threads) in the processors. This means computing times for distributed arrays in each of the sixteen common streams. For example, the blue line shows the time delay of 0.1 ms to calculate the 1st element of the array in the 1st stream, and the red line shows the time delay of 0.05ms to calculate the 1st element of the array in the 1st stream. As the number of parallel threads increases, it can be seen that the execution time of the program increases due to incorrect cache line alignment, while for correct cache line alignment, each parallel thread takes almost the same amount of time.

Another way to solve the problem is to use alignment: an element is inserted between the elements, the size of which is equal to the size of the cache line. This ensures that data that changes frequently does not end up in the same cache line, eliminating the problem of cache line misallocation.

After solving the problem for the example of allocating shared memory to the computing processes in the processor cores and the example of the queue, the measurement results were calculated similar to the example of the distributed array.

Thus, digital processing of high and low frequency seismic signals based on spline function made it possible to quickly and qualitatively determine the low speed of differential tectonic movements leading to earthquakes. The correct distribution of the cache line of distributed system computers depends on the time characteristics of the parallel algorithms. Running parallel algorithms on dual-core and octa-core CPUs showed that proper cache line allocation can digitize large seismic signals up to 10 times faster than expected. The use of the proposed methods significantly improves the performance of software tools operating on the basis of a parallel algorithm and allows for quick and high-quality identification of seam centers [11].

List of used literatures speed and efficiency of performing parallel algorithms in distributed computing systems.

CONCLUTION

In conclusion, we should note that we have studied the problem of incorrect allocation of cache memory lines in processors due to the peculiarities of the organization of memory in multi-core and multiprocessor systems of distributed systems. The study also identified problems with cache memory in parallel programming, and analyzed the use of common stream synchronization to solve parallel computing problems in software tools leading to incorrect cache memory allocation. Proportional growth was achieved after the time taken to perform large-scale computational processes did not increase proportionally, and after applying the methods proposed in the article to solve this problem. Acceleration of the execution of software algorithms and the dependence of the efficiency of the use of computing resources of the system on the number of parallel flows were analyzed. The study found that misallocation of cache memory had a significant effect on the time characteristics of multi-branch applications, and suggested ways to eliminate it.

REFERENCES

[1] L. Wu, C. Jing, X. Liu, and H. Tian, "Diagnostic methods for communication waiting in MPI parallel programs and applications," *Guofang Keji Daxue Xuebao/Journal Natl. Univ. Def. Technol.*, vol. 42, no. 2, 2020, doi: 10.11887/j.cn.202002006.

[2] U. Khamdamov and H. Zaynidinov, "Parallel Algorithms for Bitmap Image Processing Based on Daubechies Wavelets," 2018, doi: 10.1109/ICCSN.2018.8488270.

[3] D. Singh, M. Singh, and Z. Hakimjon, "Geophysical application for splines," in *SpringerBriefs in Applied Sciences and Technology*, 2019.

[4] D. Singh, H. Zaynidinov, and H. J. Lee,

"Piecewise-quadratic Harmut basis functions and their application to problems in digital signal processing," *Int. J. Commun. Syst.*, vol. 23, no. 6–7, 2010, doi: 10.1002/dac.1093.

[5] D. Singh, M. Singh, and Z. Hakimjon, "Parabolic Splines based One-Dimensional Polynomial," in *SpringerBriefs in Applied Sciences and Technology*, 2019.

[6] Y. A. Üncü, T. Mercan, G. Sevym, and M. Canpolat, "Interpolation applications in diffuse optical tomography system," 2018, doi: 10.1109/BIYOMUT.2017.8478855.

[7] H. N. Zayniddinov and O. U. Mallayev, "Paralleling of calculations and vectorization of processes in digital treatment of seismic signals by cubic spline," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 537, no. 3, doi: 10.1088/1757-899X/537/3/032002.

[8] J. L. Hennessy and D. a Patterson, "Hennessy, J. L., & Patterson, D. a. (2006). Computer Architecture, Fourth Edition: A Quantitative Approach. Computer.Computer Architecture, Fourth Edition: A Quantitative Approach," *Computers*, no. 0, 2006.

[9] S. L. Huang, W. Song, Y. Z. Wang, Y. M. Wu, X. M. Pan, and X. Q. Sheng, "Efficient and accurate electromagnetic angular sweeping of rough surfaces by MPI parallel randomized low-rank decomposition," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, 2020, doi: 10.1109/JSTARS.2020.2981124.

[10] R. Rocco, D. Gadioli, and G. Palermo, "Legio: fault resiliency for embarrassingly parallel MPI applications," *J. Supercomput.*, vol. 78, no. 2, 2022, doi: 10.1007/s11227-021-03951-w.

[11] Z. ZHAO, R. MA, L. HE, X. CHANG, and L. ZHANG, "An efficient large-scale mesh deformation method based on MPI/OpenMP hybrid parallel radial basis function interpolation," *Chinese J. Aeronaut.*, vol. 33, no. 5, 2020, doi: 10.1016/j.cja.2019.12.025.