

DOI: 10.17516/1999-494X-0446

УДК 004.4

The Method of Choosing the Technology for Creating Parallel Special Software With Time Parameterization

Alexey A. Tolmachev^{*a},
Mikhail A. Aksenov^a, Dmitriy S. Viktorov^a,
Andrey A. Filonov^a and Igor V. Lutikov^b

^a*Military Aerospace Defense Academy
named after Marshal of the Soviet Union G. K. Zhukov
Tver, Russian Federation*
^b*Siberian Federal University
Krasnoyarsk, Russian Federation*

Received 11.11.2022, received in revised form 17.11.2022, accepted 21.11.2022

Abstract. A new method of choosing the technology for creating a parallel program is proposed, which is based on the fact that a parallel program code written once can be executed by different parallel programming technologies, depending on the tasks, cycle parameters and the available time statistics of previous launches. At the same time, one of the parallelization methods can be used during the operation of a ready-made parallel program. Using Bayes' theorem, we move from a priori distributions of unknown magnitude to a posteriori distributions. We have a complete group of incompatible events, if we do not know their probabilities before the experiment, they are equally probable. As a result of the experiment, a certain technology number selection event appears and conditional probabilities of classification selection are known for this event. Then a statistical drawing is made to determine the number of the selected technology.

Keywords: technology selection method, parallel programming, creation of parallel special software, time parameterization.

Citation: Tolmachev, A.A., Aksenov, M.A., Viktorov, D.S., Filonov, A.A., Lutikov, I. V. The method of choosing the technology for creating parallel special software with time parameterization. J. Sib. Fed. Univ. Eng. & Technol., 2022, 15(8), 1000–1008. DOI: 10.17516/1999-494X-0446

Метод выбора технологии создания параллельного специального программного обеспечения с временной параметризацией

А. А. Толмачев^а, М. А. Аксенов^а,
Д. С. Викторов^а, А. А. Филонов^а, И. В. Лютиков^б

^аВоенная академия воздушно-космической обороны
имени Маршала Советского Союза Г. К. Жукова

Российская Федерация, Тверь

^бСибирский федеральный университет

Российская Федерация, Красноярск

Аннотация. Предлагается новый метод выбора технологии создания параллельной программы, который базируется на том, что написанный один раз параллельный программный код может быть выполнен разными технологиями параллельного программирования в зависимости от задач, параметров цикла и имеющейся временной статистики предыдущих запусков. При этом в процессе работы готовой параллельной программы может применяться один из методов распараллеливания. Используя теорему Байеса переходим от априорных распределений на неизвестную величину к апостериорным распределениям. Имеем полную группу несовместных событий, если нам неизвестны их вероятности до опыта, они равновероятны. В результате опыта появляется некоторое событие выбора номера технологии и для этого события известны условные вероятности классификационного выбора. Затем производится статистический розыгрыш для определения номера выбранной технологии.

Ключевые слова: метод выбора технологии, параллельное программирование, создание параллельного специального программного обеспечения, временная параметризация.

Цитирование: Толмачев А. А. Метод выбора технологии создания параллельного специального программного обеспечения с временной параметризацией / А. А. Толмачев, М. А. Аксенов, Д. С. Викторов, А. А. Филонов, И. В. Лютиков // Журн. Сиб. федер. ун-та. Техника и технологии, 2022, 15(8). С. 1000–1008. DOI: 10.17516/1999-494X-0446

В настоящее время в Вооруженных силах Российской Федерации в целом и в Воздушно-космических силах в частности проводятся работы по совершенствованию информационно-управляющих систем в целях повышения оперативности и устойчивости управления, данные факторы становятся решающими в современной войне [1]. Одним из важнейших элементов в этой системе является специальное программное обеспечение (СПО), которое дает возможность исследования сценариев различных форм и способов применения средств воздушно-космического нападения и группировок воздушно-космической обороны в виде имитационного машинного эксперимента в процессе выработки и принятия решения органами военного управления [2].

В результате проведенного анализа было установлено, что масштабируемость специального программного обеспечения моделирующих систем – важный технический аспект, предполагающий расширение их возможностей путем включения новых программных реализаций моделей [3]. Очевидно, что данные модели должны быть адекватны. Подсистема моделирования должна представлять эффективные методы доработки моделей и замены их более совершенными. С усовершенствованием модели растет ее вычислительная сложность. Наличие

множества сложных моделей определяет целесообразность распределения имитационных моделей по нескольким вычислителям (ядрам, процессорам).

Для совершенствования специального программного обеспечения потрачено немало временных, финансовых и интеллектуальных ресурсов. Огромный накопленный опыт и уже разработанное специальное программное обеспечение следует сохранять, модернизировать и повышать эффективность, используя технологии параллельного программирования.

В результате анализа технологий параллельного программирования в интересах преобразования унаследованного программного кода [4] определено, что оптимальным вариантом повышения эффективности является модернизация существующего унаследованного кода путем добавления определенных специальных операторов средств распараллеливания. Выбор технологии для распараллеливания уже имеющегося СПО необходимо проводить на основе сравнения по ряду показателей (метрик) эффективности возможных технологий распараллеливания между собой и с последовательной версией СПО. Показателями эффективности для сравнения могут быть [5]: время отклика, время выполнения, пропускная способность, ускорение, производительность, а также выбор их комбинации.

При модернизации унаследованного СПО для распараллеливания, как правило, используется одна технология параллельного программирования. Проведенный анализ [6] показывает, что разные технологии имеют разную эффективность распараллеливания в зависимости от размера распараллеливаемого цикла как элемента кода.

Предполагаем, что для получения максимальной эффективности распараллеливания при модернизации кода необходимо:

1. Выбор технологии распараллеливания выполнять перед каждым запуском цикла в автоматическом режиме с использованием накопленной статистики.
2. Время выбора технологии должно быть значительно меньше времени обработки информации в цикле.
3. Для оптимального выбора технологии необходим поиск функциональной зависимости количества повторений в цикле и типом технологии [7].

Для разрешения данного предположения разработан и предлагается новый метод распараллеливания программы, который базируется на том, что написанный один раз параллельный программный код может быть выполнен разными технологиями параллельного программирования в зависимости от задач и входных параметров. Такими параметрами могут выступать параметры цикла и имеющаяся временная статистика предыдущих запусков. При этом в процессе работы готовой параллельной программы может применяться один из методов распараллеливания:

1. Последовательный запуск технологии параллельного программирования не используется.
2. Выбирается одна технология, при которой общее время выполнения программы будет минимальным.
3. Технология выбирается при каждом запуске цикла, чтобы время выполнения этого цикла было минимальным.

Для третьего метода распараллеливания, когда технология с минимальным временем выполнения выбирается при каждом запуске цикла, можно добиться ускорения расчетов по сравнению со вторым методом. Суммарное время работы программы обозначено как

$$T_{\text{opt}} = \sum_{i=1}^M T_{\text{min}}^i,$$

где T_{min}^i – минимальное время работы i -ого цикла среди рассматриваемых технологий параллельного программирования, M – общее число вызовов распараллеленного кода.

Предложен метод, позволяющий вероятностно выбирать технологию распараллеливания для каждого вызова распараллеленного кода и одновременно сократить параметр времени выполнения программы путем оптимизации выбора технологии распараллеливания на этапе выполнения программы. Математически можно представить эту задачу в следующей формулировке

$$P_k \sim \frac{e^{-R_k^2}}{\frac{t_k}{O(N_k)} \text{Stat}} = \frac{e^{-\left(\frac{|N-N_k|}{N}\right)^2}}{\frac{t_k}{O(N_k)} \text{Stat}}, \quad (1)$$

где P_k – вероятность выбора k -ой технологии среди n технологий параллельного программирования; $R_k = \frac{|N-N_k|}{N}$ – относительное расстояние между требуемым значением числа итераций

цикла в ближайших статистических данных; $O(N_k)$ – сложность алгоритма; N – число итераций цикла; Stat – число (объем) накопленной статистики.

Метод имеет сходство с методом k ближайших соседей тем, что ближайшие данные в статистике оказывают большее влияние по сравнению с остальными. Вероятности выбора номера технологии параллельного программирования зависят от имеющихся статистических данных и от текущих временных параметров вызова распараллеливаемого кода.

Метод имеет сходство с методом k ближайших соседей тем, что ближайшие данные в статистике оказывают большее влияние по сравнению с остальными. Вероятности выбора номера технологии параллельного программирования зависят от имеющихся статистических данных и от текущих временных параметров вызова распараллеливаемого кода.

Расчет вероятности осуществляется согласно следующим принципам:

– при отсутствии статистических данных вероятности выбора разных технологий распараллеливания одинаковы и равны $1/n$, где n – количество имеющихся технологий;

– при наличии статистических данных вероятности выбора разных технологий распараллеливания зависят от имеющихся статистических данных и от текущих временных параметров вызова распараллеливаемого кода.

– чем больше время работы с указанной технологией, тем меньше вероятность ее выбора (берутся статистические данные для данной технологии с наиболее близким размером цикла), при известной сложности алгоритма $O(N_k)$ (O – нотации) время будет пересчитываться на единицу сложности;

– степень влияния статистических данных разных технологий различна и зависит от «расстояния» от запрашиваемого размера цикла N до размера цикла в статистических данных, вероятность пропорциональна величине $e^{-R_k^2}$;

– чтобы избежать выбора одного и того же значения номера технологии используется зависимость от числа накопленной статистики для данной технологии: чем меньше статистики, тем больше вероятность выбора этого значения.

Процесс выбора номера технологии распараллеливания является случайным событием, в нашем случае он функционально зависит от ряда случайных параметров.

Задача сводится к нахождению функциональной зависимости номера технологии распараллеливания ($N_{\text{тр},i}$) от количества итераций (N_i), а именно $N_{\text{тр},i} = f(N_i)$, что является задачей классификации. Если количество технологий параллельного программирования больше двух, то данная задача является задачей многоклассовой классификации. Для повышения качества

работы классификатора данные текущего запуска распараллеленного кода сразу же необходимо добавлять в статистику для использования при классификации последующих запусков распараллеленного кода. Значит, необходимо разработать метод выбора технологии создания параллельного специального программного обеспечения с учетом параметра времени обработки информации.

Таким образом, номер технологии распараллеливания будет являться функцией от количества итераций запускаемого цикла, данных статистики и времени выполнения цикла:

$$N_{\text{тр},i} = f(N_i, H, T_{\text{пар},i \text{ min}}^*),$$

где N_i – количество итераций цикла, H – статистические данные по ранним запускам распараллеливаемого кода, $T_{\text{пар},i \text{ min}}^*$ – минимальное время выполнения цикла. Статистические данные представляют собой тройки «количество итераций цикла» – «номер технологии» – «время работы», в табл. 1 показан пример статистических данных.

При использовании предлагаемого метода возникает необходимость оценки степени адекватности данного выбора, для этого предлагается использовать метод статистического моделирования (статистических испытаний). Суть данного численного метода состоит в получении оценок вероятностных характеристик, совпадающих с решением аналитических задач. Основной идеей метода статистического моделирования является замена детерминированной (при известных параметрах) задачи эквивалентной схемой некоторой стохастической (со случайными элементами) системы, выходные характеристики которой совпадают с результатом решения детерминированной задачи. Конечно, если произвести такую замену, погрешность выбора будет высокая, однако эта проблема решается увеличением количества испытаний.

Используем теорему Байеса [8] для перехода от априорных распределений на неизвестную величину к апостериорным распределениям. В нашем случае имеется полная группа несовместных событий (гипотез – по выбору номера технологии) $E_1, E_2, E_3, \dots, E_n$, если нам неизвестны их вероятности $P(E_1), P(E_2), P(E_3), \dots, P(E_n)$ до опыта, они равновероятны. В результате опыта появляется некоторое событие A (выбор номера технологии), пусть для этого события

Таблица 1. Статистические данные зависимости количества итераций, номера технологии и времени обработки информации

Table 1. Statistical data on the dependence of the number of iterations, technology number and information processing time

Количество итераций	Номер технологии	Время, сек	Количество итераций	Номер технологии	Время, сек
60	1	2.105	40	3	0.926
60	2	2.425	40	4	0.526
60	3	1.118	30	1	1.811
60	4	0.408	30	2	1.881
40	1	1.921	30	3	0.781
40	2	1.962	30	4	0.526

известны условные вероятности выбора номеров технологий $P(A/E_1)$, $P(A/E_2)$, $P(A/E_3)$, ..., $P(A/E_n)$, в нашем случае это условные вероятности классификационного выбора по (1). Как изменятся вероятности гипотез, если стало известно, что событие A (выбор номера технологии) произошло? Эти искомые вероятности $P(E_1/A)$, $P(E_2/A)$, $P(E_3/A)$, ..., $P(E_n/A)$ называются апостериорными (т.е. после опыта) вероятностями. Если событие A произойдет с одним из событий E_i , то формула Байеса преобразится

$$P(E_i/A) = \frac{P(E_i)P(A/E_i)}{P(A)} = \frac{P(E_i)P(A/E_i)}{\sum_{i=1}^n P(E_i)P(A/E_i)},$$

где $P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + \dots + P(E_n)P(A/E_n)$.

Поскольку сумма условных вероятностей классификационного выбора номера одной из технологий $P(A/E_1)$, $P(A/E_2)$, $P(A/E_3)$, ..., $P(A/E_n)$, согласно (1), не равна единице, то классический розыгрыш одного из n несовместных событий [8] к определению выбора номера технологии применен быть не может. Поэтому предлагается определять номер выбранной технологии при использовании формулы полной вероятности и формулы Байеса [8], а затем произвести статистический розыгрыш для определения номера выбранной технологии.

С этой целью введем интересующие нас события:

A – выбор номера технологии распараллеливания программы и соответствующие гипотезы для примера их четырех технологий:

E_1 – выбор технологии № 1;

E_2 – выбор технологии № 2;

E_3 – выбор технологии № 3;

E_4 – выбор технологии № 4.

При таком подходе гипотезы E_i ($i = \overline{1,4}$) образуют полную группу несовместных событий. Тогда полная вероятность выбора номера технологии распараллеливания может быть определена по формуле

$$P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + P(E_3)P(A/E_3) + P(E_4)P(A/E_4), \quad (2)$$

где условные вероятности выбора номера технологии $P(A/E_i)$ ($i = \overline{1,4}$) связаны с вероятностями $P_1; P_2; P_3; P_4$ (определяемые (1)) соотношениями:

$$P(A/E_1) = P_1; P(A/E_2) = P_2; P(A/E_3) = P_3; P(A/E_4) = P_4.$$

Поскольку данный этап реализуется при условии, что выбор номера технологии осуществлен (т.е. событие A произошло), то априорные вероятности гипотез $P(E_i)$ ($i = \overline{1,4}$) изменяются и становятся апостериорными $P(E_i/A)$ ($i = \overline{1,4}$), которые рассчитываются по формулам Байеса

$$P(E_i/A) = \frac{P(E_i)P(A/E_i)}{P(A)}, \quad (i = \overline{1,4}) \quad (3)$$

События E_i/A ($i = \overline{1,4}$) образуют полную группу несовместных событий и могут быть использованы для розыгрыша выбора номера технологии распараллеливания D в соответствии с известным правилом [3]

$$\begin{cases} 0 \leq D < P(E_1 / A) & \text{выбор технологии № 1;} \\ P(E_1 / A) \leq D < P(E_1 / A) + P(E_2 / A) & \text{выбор технологии № 2;} \\ P(E_2 / A) \leq D < P(E_2 / A) + P(E_3 / A) & \text{выбор технологии № 3;} \\ P(E_3 / A) + P(E_4 / A) \leq D \leq 1 & \text{выбор технологии № 4.} \end{cases} \quad (4)$$

Для наглядности рассмотрим конкретный пример. Если никакой априорной информации о состоянии объекта нет, то

$$P(E_1) = P(E_2) = P(E_3) = P(E_4) = 1/4.$$

Используя метод Монте-Карло [9], найдем значения вероятностей выбора номеров технологий $P(A/E_1)$, $P(A/E_2)$, $P(A/E_3)$, $P(A/E_4)$.

$$P(A/E_1) = P_1 = 0,5; \quad P(A/E_2) = P_2 = 0,6;$$

$$P(A/E_3) = P_3 = 0,2; \quad P(A/E_4) = P_4 = 0,3.$$

Подставив приведенные значения в формулу (2), получим

$$P(A) = 1/4 \cdot 0,5 + 1/4 \cdot 0,6 + 1/4 \cdot 0,2 + 1/4 \cdot 0,3 = 0,4.$$

Тогда в соответствии с (3)

$$P(E_1 / A) = \frac{\frac{1}{4} \cdot 0,5}{0,4} = 0,3125; \quad P(E_2 / A) = \frac{\frac{1}{4} \cdot 0,6}{0,4} = 0,375;$$

$$P(E_3 / A) = \frac{\frac{1}{4} \cdot 0,2}{0,4} = 0,125; \quad P(E_4 / A) = \frac{\frac{1}{4} \cdot 0,3}{0,4} = 0,1875$$

и правило (4) принимает вид:

$$\begin{cases} 0 \leq D < 0,3125 & \text{выбор технологии № 1;} \\ 0,3125 \leq D < 0,6875 & \text{выбор технологии № 2;} \\ 0,6875 \leq D < 0,8125 & \text{выбор технологии № 3;} \\ 0,8125 \leq D \leq 1 & \text{выбор технологии № 4.} \end{cases}$$

Предположим, что в конкретном розыгрыше генератор случайных чисел выдал число $D = 0,564$. В соответствии с (4) в данном розыгрыше выбрана технология № 2. При выдаче $D = 0,754$ выбрана технология № 3. Если в следующем розыгрыше $D = 0,272$, то выбрана технология № 1, а в очередном розыгрыше при $D = 0,913$ выбрана технология № 4. В рассматриваемом случае наибольшие шансы имеет технология № 2.

После реализации предлагаемого метода с накопленной статистикой предыдущих запусков была в качестве примера определена технология № 2 (по $\max P(A/E_2) = P_2 = 0,6$). Далее с применением байесовского подхода было получено апостериорное распределение по оцениваемому параметру (номеру выбираемой технологии распараллеливания по $\max P(E_2/A) = 0,375$ с последующим статистическим розыгрышем данного решения, что подтверждает адекватность предлагаемого метода выбора технологии создания параллельного специального программного обеспечения с временной параметризацией.

Список литературы / References

[1] Сурувикин С. В., Кулешов Ю. В. Особенности организации управления межвидовой группировкой войск (сил) в интересах комплексной борьбы с противником, *Журнал Военная мысль*, 2017, 8, 5–18. [Surovikin S. V., Kuleshov Yu. V. Features of the organization of the management of an interspecific grouping of troops (forces) in the interests of a comprehensive fight against the enemy, *Journal of Military Thought*, 2017, 8, 5–18 (in Rus.).]

[2] Аксенов М. А. Анализ технологий параллельного программирования для применения в моделирующих комплексах военного назначения в интересах повышения эффективности моделирования, *Сборник научных трудов IV Всероссийской научно-практической конференции «Актуальные вопросы состояния и перспектив развития сложных технических систем военного назначения»*. М., ВУЦ МГТУ имени Н. Э. Баумана, 2020, 305–312. [Aksenov M. A. Analysis of parallel programming technologies for use in modeling complexes for military purposes in the interests of improving the efficiency of modeling, *Collection of scientific papers of the IV All-Russian Scientific and Practical Conference «Current issues of the state and prospects of development of complex technical systems for military purposes»*. М., Bauman Moscow State Technical University, 2020, 305–312 (in Rus.).]

[3] Захаров В. Л., Ильин В. А. *Тренажеры Военно-Морского Флота: создание и использование: монография*. СПб–Тверь, 2019, 248. [Zakharov V. L., Ilyin V. A. *Simulators of the Navy: creation and use: monograph*. St. Petersburg–Tver, 2019, 248 (in Rus.).]

[4] Толмачев А. А., Викторов Д. С., Кучин А. А. Проблемные вопросы создания параллельного специального программного обеспечения вычислительной системы Центра контроля космического пространства, *Сборник докладов и выступлений научно-деловой программы Международного военно-технического форума «АРМИЯ – 2021». Секция «Актуальные вопросы развития технологий искусственного интеллекта в Вооруженных силах Российской Федерации»*. М., МО РФ, 2021, 362–366. [Tolmachev A. A., Viktorov D. S., Kuchin A. A. Problematic issues of creating parallel special software for the computer system of the Space Control Center, *A collection of reports and speeches of the scientific and business program of the International Military-Technical Forum «ARMY – 2021». Section «Topical issues of the development of artificial intelligence technologies in the Armed Forces of the Russian Federation»*. Moscow, Ministry of Defense of the Russian Federation, 2021, 362–366 (in Rus.).]

[5] Hennessy J. L., Patterson D. A. *Computer Architecture. A Quantitative Approach. Sixth Edition*. Cambridge: Morgan Kaufmann Publishers, 2019. 1525.

[6] Малышев А. Н., Макарычев А. В., Zubov N. N. и др. *Математическое моделирование и оценка эффективности боевых действий войск ПВО*. Тверь, ВА ПВО, 1995. 62 [Malyshev A. N., Makarychev A. V., Zubov N. N., etc. *Mathematical modeling and evaluation of the effectiveness of combat operations of the air defense forces*. Tver, VA PVO, 1995. 62 (in Rus.).]

[7] Бернгардт Г. В. Частотный анализ использования технологий распараллеливания программ на языке С, *Сборник тезисов трудов V Всероссийского конгресса молодых ученых*. СПб., Университет ИТМО, 2016, 24. [Bernhardt G. V. Frequency analysis of the use of parallelization technologies of programs in C, *Collection of abstracts of the V All-Russian Congress of Young Scientists*. St. Petersburg, ITMO University, 2016, 24. (in Rus.).]

[8] Вентцель Е. С., Овчаров Л. А. *Теория вероятностей и ее инженерные приложения*. М, Высш. шк., 2000, 480. [Wentzel E. S., Ovcharov L. A. *Probability theory and its engineering applications*. M., Higher School, 2000, 480. (in Rus.)].

[9] Некрасов К. А., Поташников С. И., Боярченко А. С., Купряжкин А. Я. *Метод Монте-Карло на графических процессорах*. Екатеринбург, Изд-во Урал. Ун-та, 2016, 60 [Nekrasov K. A., Potashnikov S. I., Boyarchenkov A. S., Kupryazhkin A. Ya. *Monte Carlo method on graphics processors*. Yekaterinburg, Ural Publishing House. Un-ta, 2016, 60 (in Rus.)].