

М. Г. Коновалов, Р. В. Разумчик (Москва, ФИЦ ИУ РАН).
Алгоритмы распределения ресурсов для частично наблюдаемых систем конечной емкости с параллельным обслуживанием.

УДК 519.87+519.24

DOI https://doi.org/10.52513/08698325_2022_29_1_1

Резюме: Для частично наблюдаемых систем с параллельным обслуживанием, составленных из серверов с очередями возможно конечной емкости и консервативными дисциплинами обслуживания, предложено семейство алгоритмов диспетчеризации.

Ключевые слова: Системы с параллельным обслуживанием, конечная емкость, диспетчеризация, управление при неполном наблюдении.

В этой заметке речь идет об одной принципиально новой постановке задачи в продолжающемся цикле исследований, посвященных распределению ресурсов (или, по-другому, стратегиям маршрутизации, распределению заданий, диспетчеризации) в частично наблюдаемых системах с параллельным обслуживанием. Типичная система представляет собой конечную совокупность параллельно и независимо друг от друга работающих обслуживающих ресурсов (серверов) различной производительности, которые выполняют задания, направляемые на них диспетчером. При этом диспетчер, осуществляя выбор сервера для выполнения очередного задания, не имеет возможности отложить решение и может руководствоваться только:

- априорной информацией о системе, т. е. распределениями времен между поступлениями заданий, распределениями размеров заданий, производительностями серверов, информацией о дисциплине обслуживания заданий в каждом сервере;
- значениями моментов времени поступления предыдущих заданий и информацией об уже принятых решениях.

Одна из главных возникающих в таких условиях проблем — это проблема построения оптимальной процедуры диспетчеризации. Наиболее естественным приложением результатов таких теоретических исследований являются системы добровольных вычислений, новая волна интереса к которым возникла в 2020 году (в связи с проектом Foldinghome). Наличие хороших «слепых» процедур распределения заданий не только упрощает масштабирование систем, но и способствует повышению их энергоэффективности (по данным [1] сейчас к платформе для организации добровольных вычислений BOINC подключено более 150 тыс. серверов).

Если емкости очередей в серверах неограничены, то вне зависимости от критерия оптимальности для диспетчера, из всех известных из научной периодической печати маршрутизаций, доступны лишь три: рандомизированная [2], циклическая [3] и маршрутизация по предыстории [4]. Из них, как показывают вычислительные эксперименты, последняя, хотя и уступает первым двум в простоте реализации, в остальном является наилучшей. Такой скудный выбор правил действия для диспетчера является следствием тех жестких ограничений, в которых осуществляется распределение ресурсов; в частности, диспетчером не наблюдается даже функционал, подлежащий оптимизации. Отказ от предположения неограниченности размеров очередей в серверах еще больше осложняет дело. Применение любой из упомянутых выше диспетче-

ризаций не может быть оправдано ни теоретически, ни экспериментально, поскольку заложенные в них правила выбора очередного действия не учитывают наличие потерь — особенность, существенно влияющую на значение традиционных для распределенных компьютерных систем целевых функционалов.

В докладе речь пойдет о семействе алгоритмов диспетчеризации (основанных на некоторых идеях из [4, 5, 6]), которые (при консервативных дисциплинах обслуживания в серверах) позволяют учитывать при выборе очередного действия ограничения на размеры очередей. Ввиду отсутствия эталонных правил диспетчеризации, вопрос эффективности (и тем более оптимальности) предложенных решений остается на данный момент открытым.

диспетчеризации (пока не поддающихся теоретическому анализу) [5, 6], который свободен от серьезных вычислительных недостатков, присущих существующим диспетчеризациям, и превосходит их при всех допустимых значениях загрузки.

СПИСОК ЛИТЕРАТУРЫ

1. *BOINC computing power* <https://boinc.berkeley.edu/computing.php>
2. Bell C. H., Stidham S. Individual versus social optimization in the allocation of customers to alternative servers. — *Management Science*, 1983, v. 29, № 7, p. 831–839.
3. Hordijk A., van der Laan D. Periodic routing to parallel queues and billiard sequences. — *Mathematical Methods of Operations Research*, 2004, v. 59, № 2, p. 173–192.
4. Konovalov M., Razumchik R. A Improving routing decisions in parallel non-observable queues. — *Computing*, 2018, v. 100, № 10, p. 1–21.
5. Konovalov M., Razumchik R. A simple dispatching policy for minimizing mean response time in non-observable queues with SRPT policy operating in parallel. — *Communications of the ECMS*, 2020, v. 34, № 1, p. 398–402.
6. Konovalov M., Razumchik R. Minimizing mean response time in batch-arrival non-observable systems with single-server FIFO queues operating in parallel. — *Communications of the ECMS*, 2021, v. 35, № 1, p. 272–278.

UDC 519.87+519.24

DOI https://doi.org/10.52513/08698325_2022_29_1_1

Konovalov M. G., Razumchik R. V. (Moscow, FRC CSC RAS). **Routing of jobs to non-observable parallel queues with finite capacities.**

Abstract: The problem of routing of homogeneous jobs to non-observable parallel queues, having finite capacities, is discussed briefly. New approach is sketched.

Keywords: Parallel service, finite capacity, dispatching, decision making under non-observability.