

Р. В. Р а з у м ч и к (Москва, ИПИ ФИЦ ИУ РАН). Оценка среднего времени отклика в суперкомпьютерных системах с неточным заказанным временем обслуживания.

УДК 519.87+519.24

Резюме: Предлагается новый подход для построения оценок стационарного среднего времени отклика в одном классе моделей суперкомпьютерных систем, в которых заказанное пользователем время выполнения задания не совпадает с фактическим.

Ключевые слова: суперкомпьютерные системы, неточное заказанное время обслуживания, среднее время отклика.

Средства, позволяющие определять показатели эффективности функционирования суперкомпьютерных систем, ресурсы которых предоставляются в аренду (для проведения расчётов, моделирования, анализа данных и пр.), имеют большое значение для их владельцев. Одна из причин — необходимость формировать и соблюдать требования, фиксируемые в соглашениях об уровне обслуживания. Разработка подобных средств сопряжена с известными трудностями. В частности, в современных суперкомпьютерных системах может наблюдаться следующий эффект: заказанное пользователем время выполнения задания не совпадает с фактическим. Значит локальная (и глобальная) система управления заданиями использует заведомо неточную¹ информацию о временах обслуживания. В итоге, показатели производительности системы «скачут»: без вмешательства извне системе не удастся их удерживать на заданном уровне. Таким образом, с практической точки зрения планировщики и методы, позволяющие устранять эффекты, вызываемые пользовательскими ошибками в априорной информации о вычислительных заданиях, представляют большой интерес.

Судя по публикациям в открытой периодической печати, в настоящее время эта проблема остается по большей части открытой. За кратким обзором результатов в этом направлении можно обратиться, например, к [1, Раздел 1]. На основе некоторых новых результатов для систем массового обслуживания с неточным временем обслуживания, в докладе предлагается новый подход, который для определенного класса моделей суперкомпьютерных систем (лишь на основе имеющейся априорной информации) позволяет строить более точные оценки стационарного среднего времени отклика — одного из главных показателей производительности суперкомпьютерных систем. Суть подхода заключается в построении (математической или имитационной) модели исходной системы, реализации в ней дисциплины ресамплинга (см. [1, Раздел 2]) и использовании показателей производительности этой измененной модели в качестве оценок показателей исходной системы. Эксперименты на имитационной модели с системой из [2] — простейшей моделью суперкомпьютера с вычислительными заданиями с фиксированным ресурсным запросом — показывают, что, при наличии мультипликативной ошибки в информации о заказанных временах выполнения заданий, новый подход приводит к более точным оценкам фактического стационарного среднего времени отклика (см. Рис. 1).

© Редакция журнала «ОПиПМ», 2020 г.

¹Т.е. фактическое время обслуживания может «оказаться» как больше, так и меньше указанного изначально пользователем.

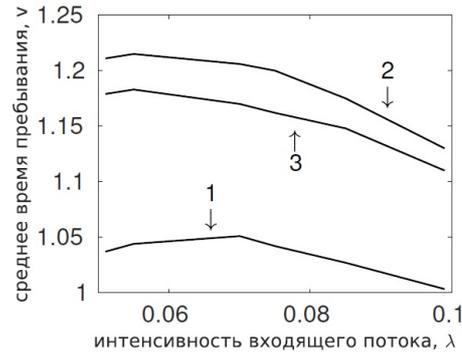


Рис. Среднее время пребывания в системе из [2]. Исходные данные соответствуют [2, Fig.9(b)]. Входящий поток — пуассоновский с параметром λ . 1 — экспоненциальное распределение заказанного времени обслуживания, 2 — экспоненциальное распределение с мультипликативной ошибкой с распределением $\ln \mathcal{N}(0, 0.25)$, 3 — оценка с помощью нового подхода.

Отметим, что для применения нового подхода необходимо наличие у априорного распределения заказанного пользователем времени обслуживания «тяжелого хвоста». Общий же критерий пока найти не удастся: каждый отдельный случай требует отдельного исследования².

Исследование выполнено при финансовой поддержке РФФИ (проект 20-07-00804).

СПИСОК ЛИТЕРАТУРЫ

1. *Horvath I., Miklos T., Razumchik R.* The resampling $M/G/1$ non-preemptive LIFO queue and its application to systems with uncertain service time. — Performance Evaluation, 2019, Art. ID: 102000.
2. *Filippopoulos D., Karatza H.* An $M/M/2$ parallel system model with pure space sharing among rigid jobs. — Mathematical and Computer Modelling, 2007, v. 45, p. 491–530.

УДК 519.87+519.24

Razumchik R. V. (Moscow, Institute of Informatics Problems, FRC CSC RAS).
Estimation of mean response time in supercomputer systems with rigid jobs and inaccurate job size information.

Abstract: New approach for the estimation of mean response time in in one class of models of supercomputer systems with rigid jobs and inaccurate job size information is being suggested.

Keywords: supercomputer systems, inaccurate job size information, mean response time.

²Которое, при соблюдении некоторых дополнительных условий, может проводиться и аналитически.