

**Оптимизация автомасштабирования контейнеров на основе анализа нагрузки**  
**Мельничук М. (ИТМО)**  
**Научный руководитель – кандидат технических наук, Богатырев Владимир**  
**Анатольевич (ИТМО)**

**Введение.** Проблема оптимизации вычислительных ресурсов в контейнеризованных системах приобретает все большее значение в контексте облачных вычислений и микросервисной архитектуры. Основной задачей в данных системах является определение оптимального количества контейнеров для достижения баланса между производительностью системы и эффективностью использования ресурсов [1]. Традиционные подходы к автомасштабированию часто не учитывают специфику рабочей нагрузки, что приводит к неэффективному распределению ресурсов [2]. Данная работа предлагает алгоритм автомасштабирования, основанный на анализе неоднородности нагрузки через параметр, который характеризует соотношение между нагрузками ориентированными на ввод-вывод и нагрузками ориентированных на процессор.

**Основная часть.** Предыдущие исследования установили связь между количеством контейнеров и временем обслуживания [3] атак же и типом нагрузки [4] что имеет большое практическое значение для оптимизации затрат в облачной инфраструктуре. В данной публикации разработан алгоритм нахождения оптимального количества контейнеров  $n_{opt}$  в зависимости от параметра неоднородности нагрузки  $\alpha \in [0.0, 1.0]$ , где  $\alpha = 0.0$  соответствует нагрузке исключительного ввода-вывода (ВВ), а  $\alpha = 1.0$  соответствует нагрузке только на центральный процессор (ЦП). Алгоритм использует адаптивную стратегию поиска, которая варьируется в зависимости от типа нагрузки.

**Ключевые особенности алгоритма.** Для ЦП интенсивных нагрузок ( $\alpha \geq 0.6$ )

алгоритм выявляет V-образную кривую производительности с резким минимумом. Применяется глобальный поиск в комбинации с локальной оптимизацией и избегает перерасхода ресурсов в областях, где деградация производительности резкая. Для ВВ интенсивных нагрузок ( $\alpha \leq 0.4$ ) характеризуются плоским минимумом (плато), где производительность слабо зависит от количества контейнеров. Алгоритм применяет агрессивную консолидацию ресурсов прогнозируя сдвиг оптимума к левому краю спектра. Это позволяет достичь значительной экономии затрат при минимальной потере производительности. Для смешанных нагрузок ( $\alpha \in [0.4, 0.6]$ ) переходная зона с простым унимодальным минимумом. Алгоритм использует консервативную стратегию поиска для надежного нахождения точки оптимума.

**Выводы.** Разработан алгоритм оптимизации количества контейнеров в зависимости от характера нагрузки ( $\alpha$ ), демонстрирующий высокую точность, при которой средняя деградация производительности составляет менее 4.5% от теоретического оптимума, при этом адаптивная стратегия поиска позволяет эффективно обрабатывать различные типы рабочих нагрузок. Практическое применение алгоритма дает возможность существенно снизить затраты на облачную инфраструктуру, особенно для ВВ интенсивных приложений за счет агрессивной консолидации контейнеров.

**Список использованных источников:**

1. Татарникова Т., Архипцев Е., Кармановский Н. Определение размера кластера и числа реплик высоконагруженных информационных систем // Известия высших учебных заведений. Приборостроение. 2023. № 66. С. 646–651. <https://doi.org/10.17586/0021-3454-2023->

[66-8-646-651](#)

2. Богатырев В. А., Богатырев С. В., Богатырев А. В. Граничная оценка надежности кластерных систем на основе декомпозиции марковской модели при ограниченном восстановлении узлов с накоплением отказов // Научно-технический вестник информационных технологий, механики и оптики. 2025. № 25. С. 574–583. <https://doi.org/10.17586/2226-1494-2025-25-3-574-583>

3. Фунг В. К., Богатырев В. А., Кармановский Н. С., Лэ В. Х. Оценка вероятностно-временных характеристик компьютерной системы с контейнерной виртуализацией // Научно-технический вестник информационных технологий, механики и оптики. 2024. № 24. С. 249–255. <https://doi.org/10.17586/2226-1494-2024-24-2-249-255>

4. Мельничук М. и Богатырев В. А. Анализ кластеров контейнерной виртуализацией при сочетании нагрузки, ориентированной на вычисления и ввод-вывод информации // Научно-технический вестник информационных технологий, механики и оптики. 2026.