



Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра суперкомпьютеров и квантовой информатики

Скрябин Глеб Денисович

**Воспроизведение результатов исследования влияния
системы мониторинга производительности на выполнение
коллективных MPI операций**

КУРСОВАЯ РАБОТА

Научный руководитель:
к.ф.-м.н., с.н.с. НИВЦ МГУ
Стефанов К.С.

Москва 2022

Содержание

1. Введение.....	3
2. Характеристика исходной работы.....	3
3. Метод проверки	3
4. Критерий обнаружения шума системы мониторинга	4
5. Влияние шума на детектор	6
6. Влияние шума на детектор, стандартная постановка задачи	7
7. Влияние шума на детектор, привязка к ядрам.....	9
8. Заключение.....	11
9. Ссылки.....	12

1. Введение

Для выполнения данной курсовой работы была выбрана ВКР по исследованию влияния системы мониторинга производительности суперкомпьютера Ломоносов-2 [1] на сильно синхронизированные программы, содержащие коллективные MPI операций All-to-All и Barrier. [2]

2. Характеристика исходной работы

Параллельное приложение, запускаемое на суперкомпьютере, может быть исследовано с помощью системы мониторинга производительности. Агент системы мониторинга собирает информацию о состоянии программно-аппаратной среды во время запуска параллельного приложения. Однако агент системы мониторинга производительности оказывает дополнительную нагрузку, помехи на пользовательскую задачу, так как для сбора данных также использует ресурсы вычислительной системы. В выбранной ВКР изучается влияние агента системы мониторинга производительности суперкомпьютера Ломоносов-2 на представленный детектор шума. Детектор, основанный на выполнении цикла, содержащего коллективные MPI операций All-to-All и Barrier, показал себя чувствительным к малому шуму этой системы.

3. Метод проверки

Проверка состоит в воспроизведении опытов и результатов исследования влияния системы мониторинга производительности суперкомпьютера на сильно синхронизированные программы, при помощи методы, используемых в изначальных опытах.

Запуски программ проводились в тестовом разделе, в котором имеется ограничение по времени работы программ, чего не было во всех изначальных опытах. Для того, чтобы в таких условиях получить результаты, схожие с изначальными, требуется уменьшить параметры, задающие уровень нагрузки на вычислительный узел суперкомпьютера. Это может повлиять на точность, но при этом воспроизведение и сравнения результатов все еще возможны. Так же тестовый раздел отличается от основного тем, что в нем доступно лишь 32 вычислительных узла, в остальном они одинаковые. [1]

Таблица 1: Характеристики раздела Test. [1]

Общее число узлов	32
Центральный процессор	Intel Haswell-EP E5-2697v3, 2.6 GHz, 14 cores
Графический ускоритель	NVidia Tesla K40M
Объем оперативной памяти на узле	64 В

Все опыты проводились при помощи скриптов постановки детектора в очередь для исполнения на суперкомпьютере. Один скрипт использовал только детектор, а второй одновременно с детектором создавал искусственный шум, для имитации реального шума системы мониторинга производительности. Для каждого из вариантов операций All-to-All и Barrier использовались по два таких скрипта. Скрипты использовали следующие пользовательские параметры для запусков:

1. Количество используемых узлов;
2. Количество используемых логических ядер на каждый узел;
3. Номера узлов для их привязки;
4. Количество итераций в БОЛЬШОМ цикле детектора;
5. Количество итераций в МАЛОМ цикле.

Надо отметить, что в данной работе не использовались конфигурации с 28 логическими ядрами на узел, поэтому в опытах, которые изначально проводились с работой агента системы мониторинга производительности на тех же ядрах что и детектор, требовалось вручную поставить их выполнение на одни и те же ядра.

В данной работе были проведены следующие опыты:

1. Эксперимент по обнаружению шума для проверки заданных критериев обнаружения шума;
2. Оценка влияния шума системы мониторинга на детектор;
3. Оценка влияния шума системы мониторинга на детектор при стандартной постановке задачи с разными режимами работы мониторинга (частота проверки 1, 5, 10 Гц);
4. Оценка влияния шума системы мониторинга на детектор с разными вариантами привязки процессов к логическим ядрам.

Графики, описывающие результаты воспроизведенных опытов для сравнения представлены вместе с изначальными результатами. На них показываются скрипичные диаграммы, составленные из нескольких повторных запусков. Это позволяет частично обойти временное ограничение тестового раздела в 15 минут и описать доверительные интервалы. На графиках по ординатам - шкала времени, которая указана в секундах; синим крестиком обозначено среднее значение времени запусков детектора шума. По координатам - конфигурации запусков, обозначена в виде [количество узлов]_[количество логических ядер на узел].

4. Критерий обнаружения шума системы мониторинга

«На запусках детектора продолжительностью работы от 10 до 15 минут, когда в цикле выполняется большое число коллективных операций, можно собрать статистику времени работы детектора. В таком случае будем считать, что детектор обнаруживает шум системы мониторинга, если 99% доверительные интервалы времени запусков детектора без и с шумом удалены друг от друга и не пересекаются. Расстояние между доверительными интервалами

называется *boundary gap*. Влияние детектора может быть обнаружено статистически, если *boundary gap* имеет положительное значение.» [2, стр. 4]

Для понимания, проведем эксперименты по обнаружению шума. Запустим детектор на 2, 4 и 8 узлах, используя 14 ядер на узел. Во всех опытах будем использовать конфигурацию с 100000 итераций. Собрав доверительные интервалы, рассмотрим полученный график. (Рис. 2) Результаты схожи с примером расположения доверительных интервалов, представленном в изначальном исследовании. (Рис. 1)

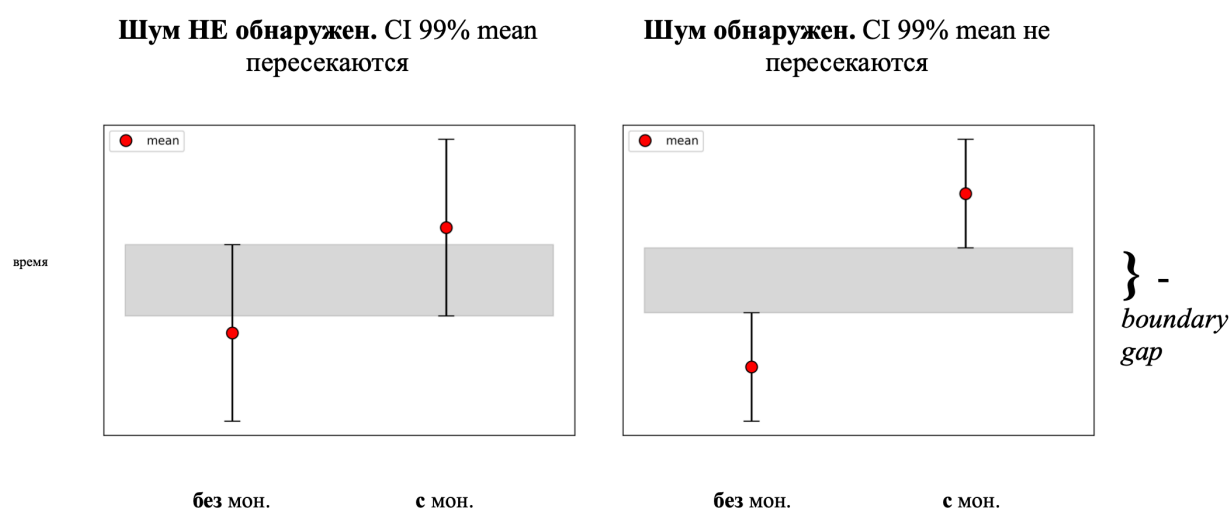


Рис. 1: Варианты расположения доверительных интервалов, представленные в исследуемой работе. [2, рис. 1]

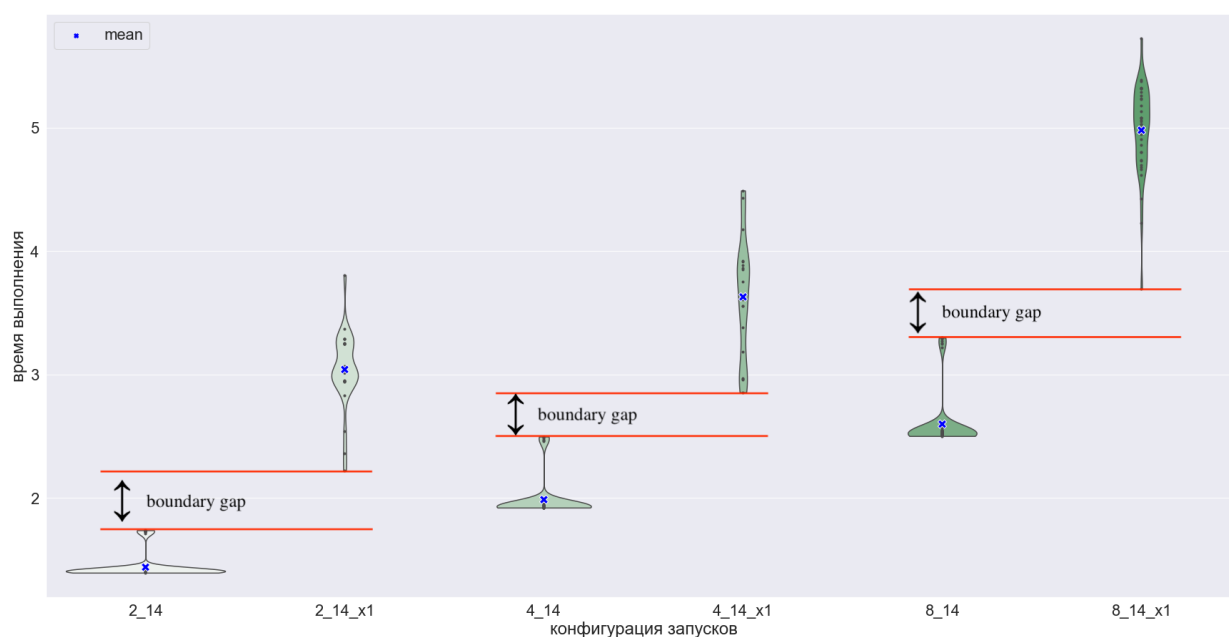


Рис. 2: Пример расположения доверительных интервалов при обнаружении шума. Детектор запускался на 2, 4 и 8 узлах, 14 логических ядер на узел.

5. Влияние шума на детектор

«Приведем результаты запусков детекторов с операцией All-to-All и Barrier на 4-х и 8-ми узлах СК. Детектор запущен на 28 ядрах каждого узла (на всех логических ядрах), мониторинг на одном ядре узла. Значение *boundary gap* во всех трех случаях положительное. Это означает, что детектор обнаруживает агент системы мониторинга, работающий в стандартном режиме с частотой 1 Гц, так как доверительные интервалы не пересекаются.

Обозначения на графиках:

1. *normalized mean gap* – разница между средними значениями времени работы детектора с и без мониторинга;
2. *normalized boundary gap* – *boundary gap*. Если значение положительное, значит детектор обнаруживает шум 1 Гц.» [2, стр. 5]

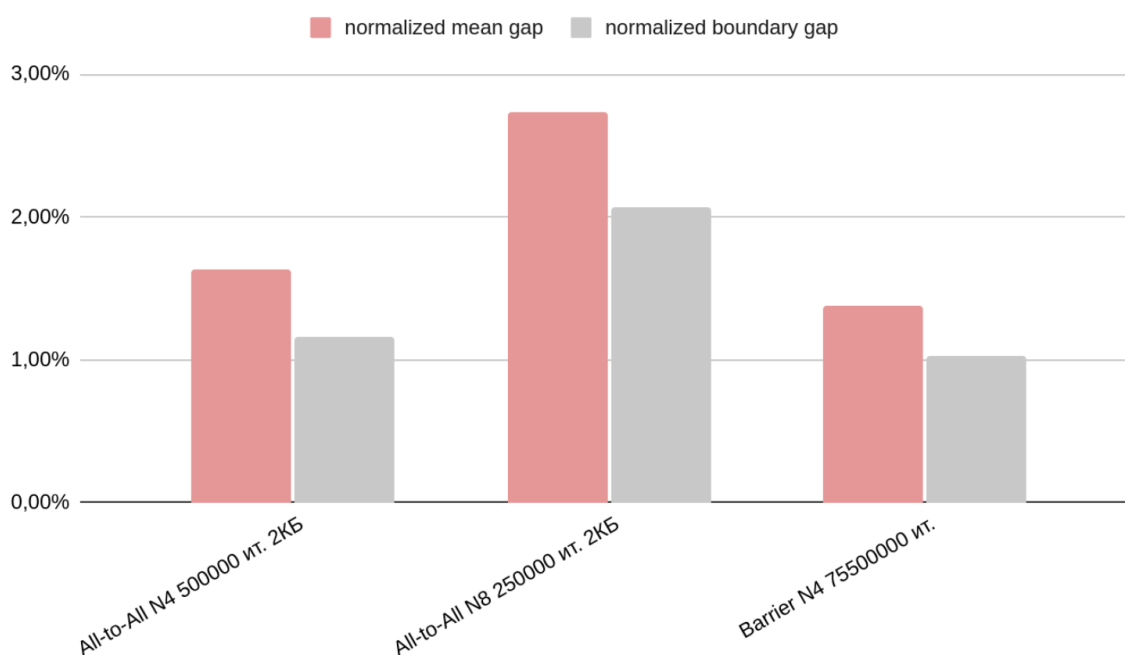


Рис. 3: Результаты запусков детектора. 28 ядер на узел. [2, рис. 2]

Теперь проведем запуски детектора на 2-х и 4-х узлах и 14 ядрах каждого узла. Во всех опытах будем использовать конфигурацию с 100000 итераций. Во всех четырех случаях значение *boundary gap* положительное. Это означает, что детектор обнаруживает агент системы мониторинга, работающий в стандартном режиме с частотой 1 Гц, что соответствует результатам изначальных исследований.

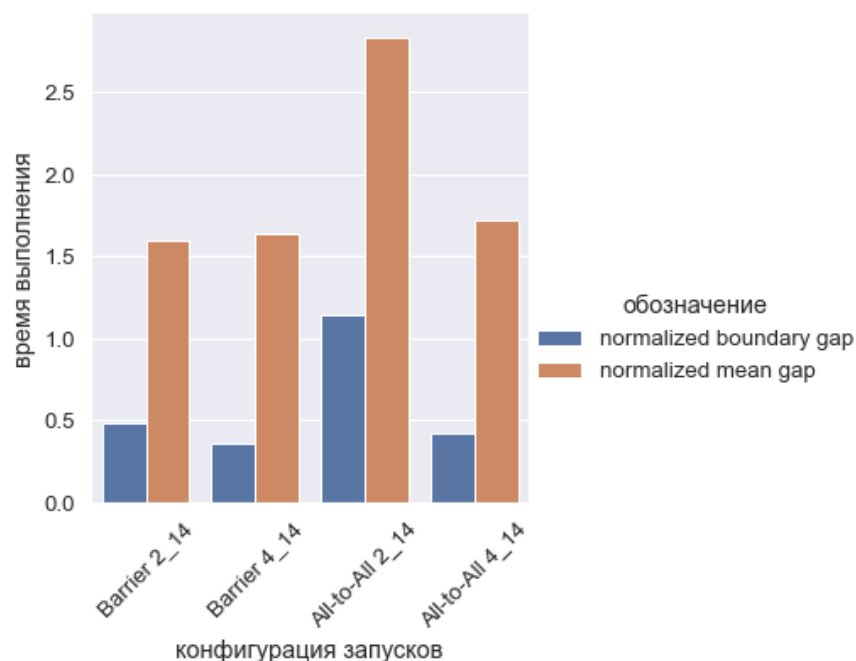


Рис. 4: Результаты запусков детектора. 14 ядер на узел.

6. Влияние шума на детектор, стандартная постановка задачи

«После того, как было показано, что время работы детектора, работающего на тех же узлах, что и задача, увеличивается от шума агента системы мониторинга, было решено проверить, можно ли обнаружить систему мониторинга, если детектор запущен стандартно, когда, каждый процесс работает на отдельном ядре. Так запускается большинство приложений на СК, это стандартный способ постановки задачи.» [2, стр. 5]

На рисунках 5 [2, рис. 3], 6 и 7 представлены результаты запусков. Во всех опытах используется конфигурация с 14 ядрами на узел. В воспроизводимых опытах 60000 итераций. Видно, что для каждого режима работы мониторинга - с частотой 1, 5 и 10 Гц, доверительные интервалы времени запусков детектора, работающего без системы мониторинга и с ней пересекаются для всех рассмотренных конфигураций детектора. Это означает, что в каждом случае влияние агента системы мониторинга статистически не обнаруживается. То есть результаты воспроизведенных опытов полностью совпадают с изначальными результатами.

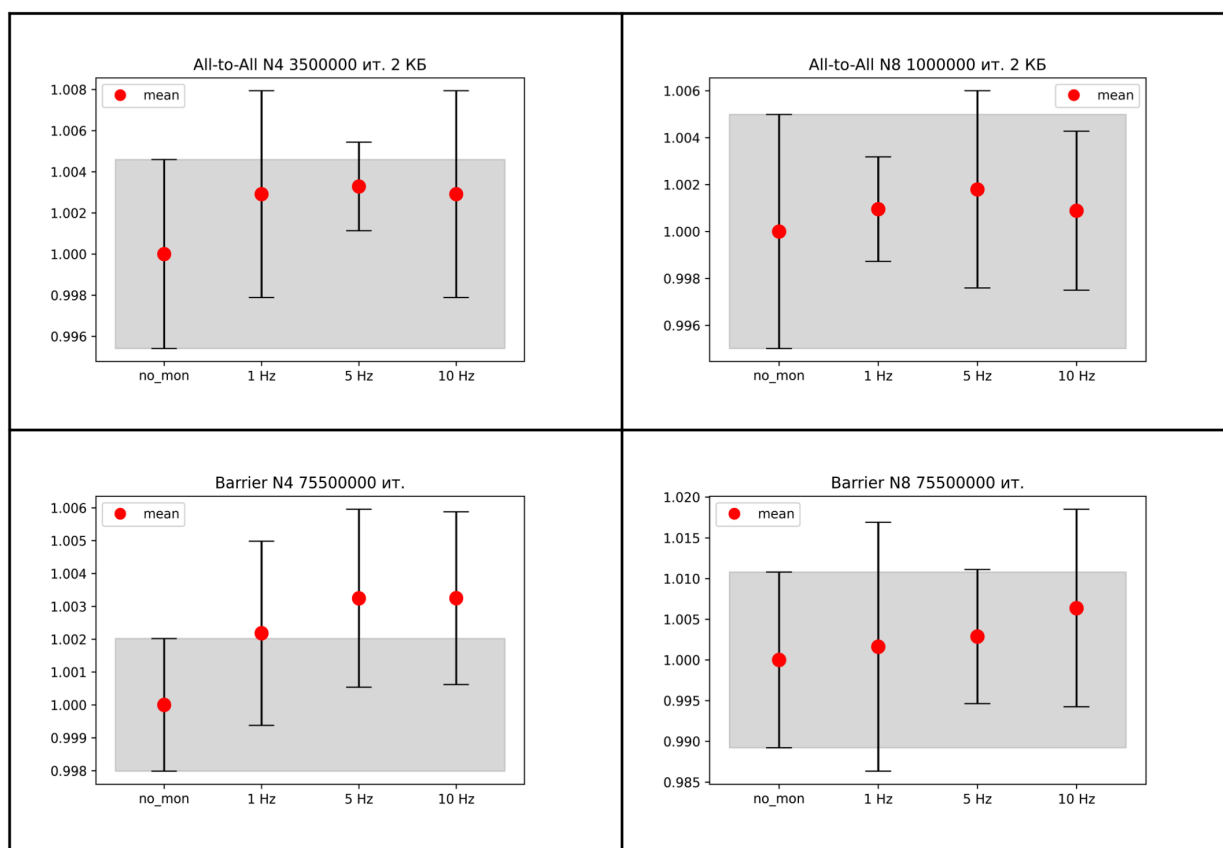


Рис. 5: Результаты запусков детектор. Детектор использует 14 ядер на узел. [2, рис. 3]

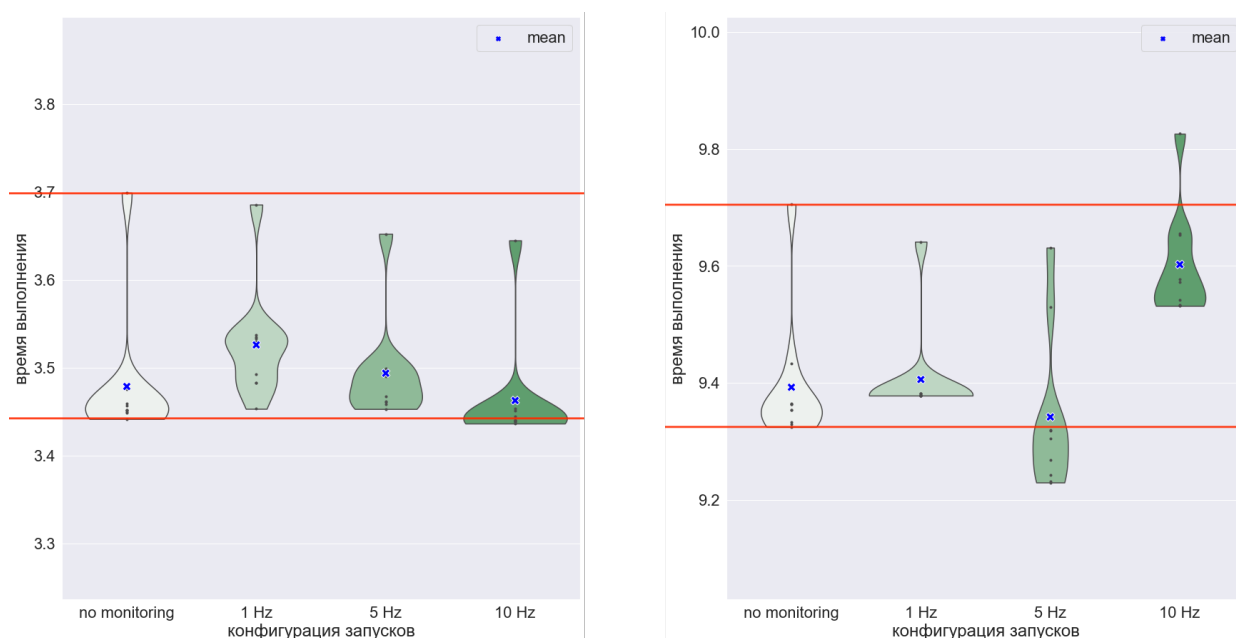


Рис. 6: Результаты запусков детектора. All-to-All, 4 и 8 узлов соответственно. Детектор использует 14 ядер на узел.

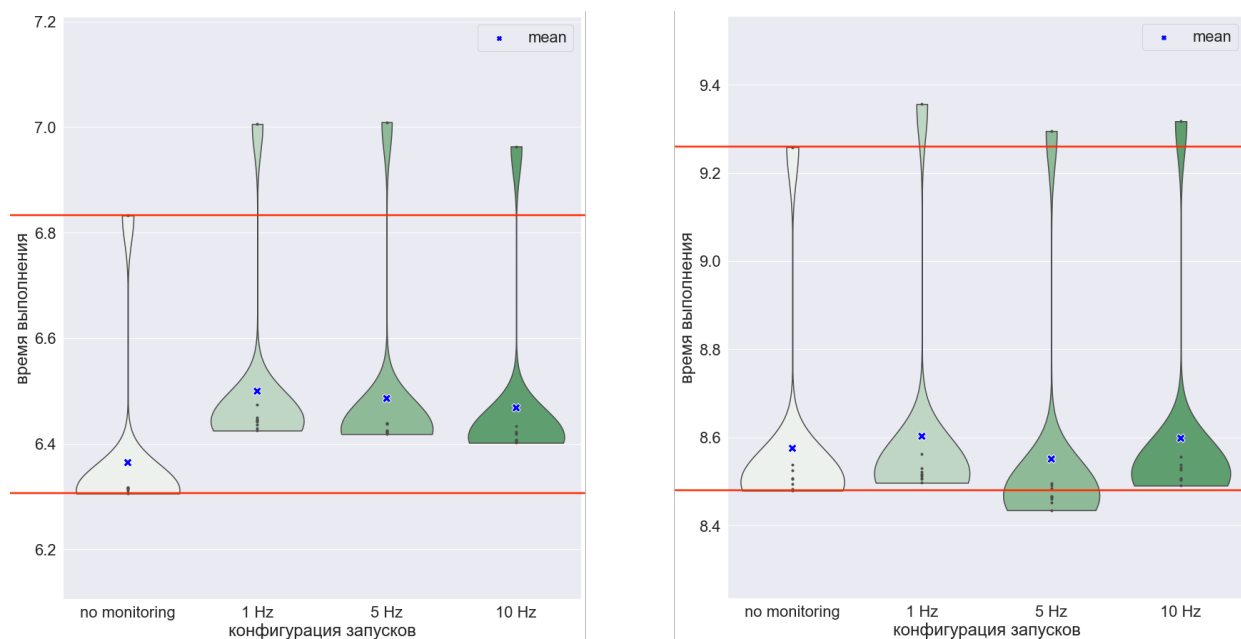


Рис. 7: Результаты запусков детектора. Barrier, 4 и 8 узлов соответственно. Детектор использует 14 ядер на узел.

7. Влияние шума на детектор, привязка к ядрам

«Для продолжений исследований, и для получения более точных и понятных результатов, попробуем установить привязку детектора и агента системы мониторинга к ядрам. Будем рассматривать три варианта работы детектора:

1. А - детектор работает без системы мониторинга, число процессов совпадает с числом физических ядер, каждый процесс работает на отдельном ядре;
2. В - детектор и мониторинг расположены на логических ядрах с разными номерами. Число процессов детектора совпадает с числом физических ядер;
3. С - агент привязан к логическому ядру, на котором выполняется детектор шума. Число процессов детектора совпадает с числом физических ядер.

В этом случае рассматриваются интервалы времени работы меньшего размера и тогда будем считать, что детектор обнаруживает систему мониторинга, если время работы детектора с мониторингом можно отнести к выбросу, относительно значений выборки по времени запуска детектора без мониторинга.» [2, стр. 6-7]

Повторим те же опыты на 2 и 4 узлах и 14 ядрах на каждый из узлов, 30000 итераций.

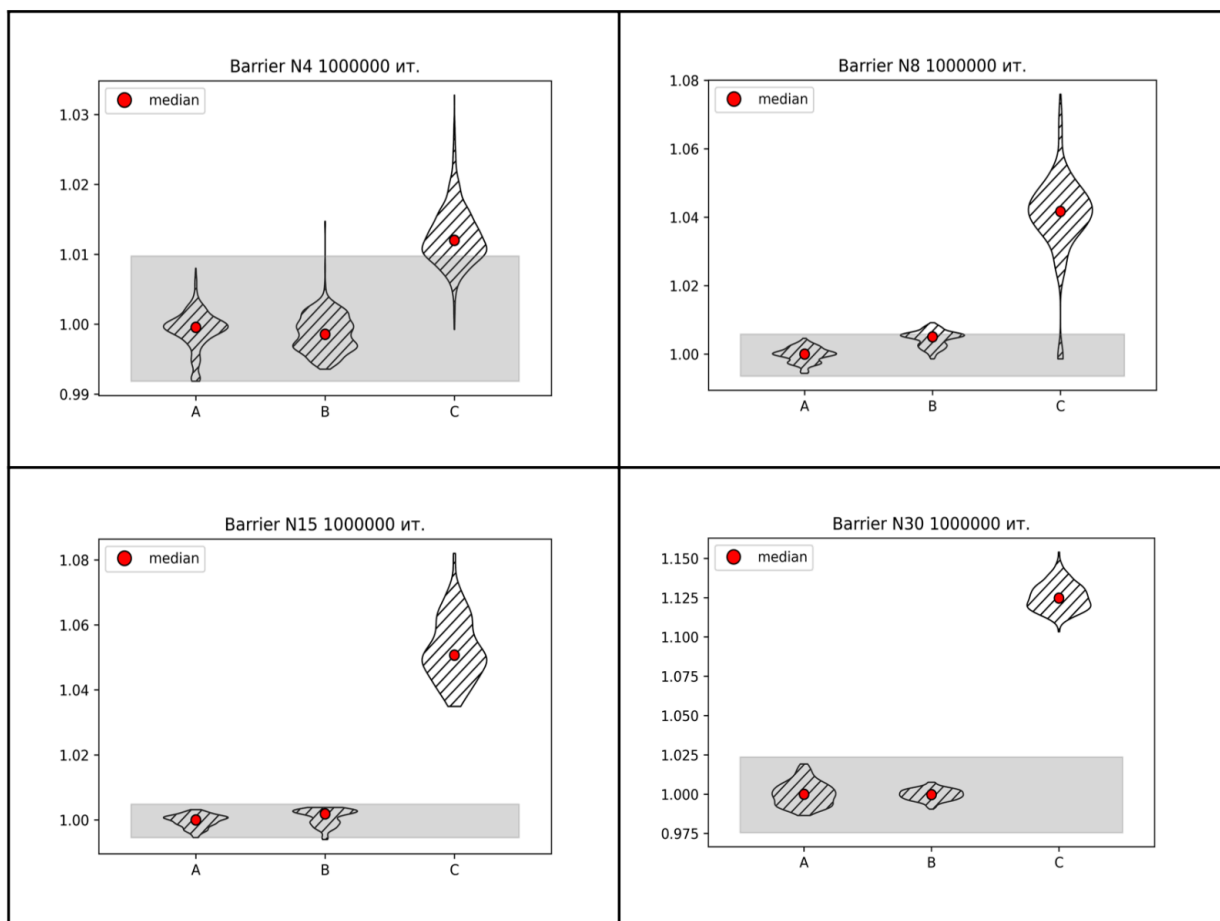


Рис. 8: Barrier 14 ядер на узел, представлены в исследуемой работе. [2, рис. 5]

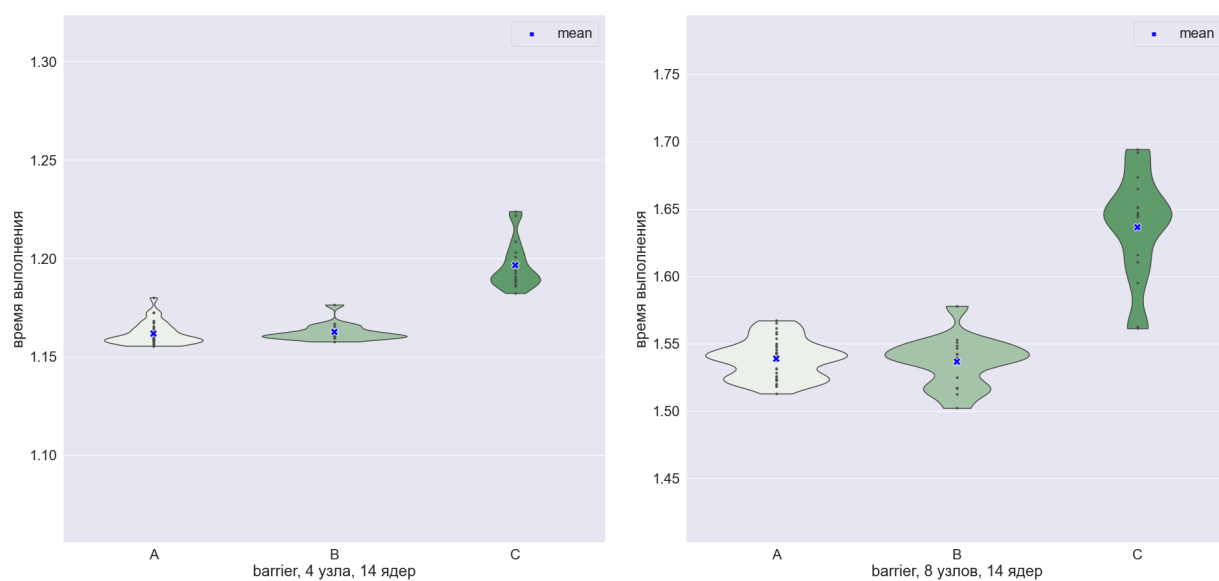


Рис. 9: Barrier 14 ядер на узел, 4 и 8 узлов соответственно.

На рисунке 9 видно, что в воспроизводимых опытах детектор с операцией Barrier обнаруживает запуски в случае С, когда мониторинг работает на том же логическом ядре, что и детектор, также, как и в изначальных исследованиях. (Рис. 8) Однако результаты немного отличаются: можно увидеть, что значения времени работы программы в воспроизведенных опытах имеют больший диапазон, чем в изначальных результатах. Такой разброс обусловлен малым количеством итераций: 30 тысяч в воспроизведенных опытах против 1 миллиона в изначальных. По той же причине исследования с детектором с операцией All-to-All воспроизвести не удалось, в результирующих данных не удалось выявить закономерность.

8. Заключение

В работе по воспроизведению результатов ВКР Худолеевой А.А. были рассмотрены два варианта запуска агента системы мониторинга производительности – когда выполнение процесса детектора происходит на том же логическом узле, что и агент системы мониторинга, и когда они выполняются на различных логических ядрах.

Результаты проведенных в данной работе опытов:

1. Эксперимент по обнаружению шума показал обоснованность заданных критериев обнаружения шума;
2. Воспроизведение опыта по оценке влияния шума системы мониторинга на детектор показало результаты идентичные с изначальными. Это означает, что детектор обнаруживает агент системы мониторинга, работающий в стандартном режиме с частотой 1 Гц;
3. Воспроизведение опыта по оценке влияния шума системы мониторинга на детектор при стандартной постановке задачи с разными режимами работы мониторинга с частотой проверки 1, 5, 10 Гц показало результаты идентичные с изначальными. Это означает, что влияние агента системы мониторинга статистически не обнаруживается при стандартной постановке задачи;
4. Воспроизведение опыта по оценке влияния шума системы мониторинга на детектор с разными вариантами привязки процессов к логическим ядрам не показало полное соответствие с изначальными результатами. Детектор с операцией Barrier обнаруживает запуски в случае, когда мониторинг работает на том же логическом ядре, но значения времени работы программы в воспроизведенных опытах имеют больший диапазон, чем в изначальных. Исследования с детектором с операцией All-to-All воспроизвести не удалось.

По итогам воспроизведения опытов можно подтвердить, что шум агента системы мониторинга в случае выполнения его на одном логическом ядре с процессом детектора, может замедлить выполнение задач, содержащих большое количество коллективных MPI операций типа All-to-All и Barrier.

Также можно подтвердить вывод о незначительности влияния агента системы мониторинга, работающего с задачами, содержащими большое количество коллективных MPI операций на разных логических ядрах. Это верно для разных режимов работы системы мониторинга с частотой 1, 5, 10 Гц.

9. Ссылки

1. <https://parallel.ru/cluster/lomonosov2.html>
2. Худолеева А.А. Исследование влияния системы мониторинга производительности на выполнение memory-bound инструкций и MPI операций.