# Отказоустойчивая коммуникационная среда для системы фрагментированного программирования LuNA

Н.А. Беляев

Новосибирский Государственный Технический Университет

С ростом производительности суперкомпьютеров растет и сложность их организации. Вместе с тем повышается вероятность и количество отказов узлов суперкомпьютера. Необходимо дать возможность параллельным программам преодолевать отказы оборудования. При этом разработчик параллельных программ должен быть абстрагирован от задачи системного программирования при обработке параллельной программой отказов оборудования. В данной работе предлагается прототип отказоустойчивой коммуникационной среды для системы фрагментированного программирования LuNA.

Система LuNA состоит из нескольких основных компонент, для каждой необходимо разработать алгоритмы обеспечения отказоустойчивости. Одной из таких компонент является очередь задач, в которой должна быть предусмотрена возможность дублирования задач и их восстановления в случае сбоя оборудования. Для решения проблемы отказоустойчивости в системе LuNA также необходим и низкоуровневый коммуникационный слой, предоставляющий возможность компонентам системы LuNA реагировать на отказы оборудования.

Разработанный коммуникационный слой FTPA подобно MPI организует связи между процессами параллельной программы и позволяет параллельной программе получать уведомления о сбоях узлов вовлеченных в вычисления или поступлении от системы управления прохождением задач новых узлов суперкомпьютера. Система следит за целостностью системы взаимодействующих процессов параллельной программы помощью распределенного сервиса, который проверяет узлы, вовлеченные в вычисления посредством обмена короткими сообщениями и в случае отказа уведомляет параллельную программу о наступившем событии. Особенностью FTPA является локальная адресация процессов. Локальная адресация не накладывает ограничений на масштабируемость и удобна для реализации коммуникационной среды для системы LuNA.

С использованием библиотеки FTPA разработан прототип отказоустойчивой очереди задач для системы LuNA. Очередь производит дублирование поступивших задач и при сбое на одном из узлов суперкомпьютера восстанавливает задачу, используя ее копию на других узлах. Такая реализация обеспечивает отказоустойчивость при выходе из строя одного или нескольких узлов суперкомпьютера при условии, что сбой произошел не в момент восстановительной операции.

## Литература

1. Bland, W. "User Level Failure Mitigation in MPI," Euro-Par 2012: Parallel Processing Workshops, Caragiannis, I., Alexander, M., Badia, R., Cannataro, M., Costan, A., Danelutto, M., Desprez, F., Krammer, B., Sahuquillo, J., Scott, S., and Weidendorfer, J. eds. Springer Berlin Heidelberg, Rhodes Island, Greece, 7640, 499-504, August, 2012.