УДК 004.318

Разработка протокола и контроллера сетевого взаимодействия для вычислительного кластера на микропроцессорах с архитектурой «Эльбрус»

И.В. Белянин, П.Ю. Петраков

ЗАО «МЦСТ»

Рассматриваемый кластер представляет собой группу вычислительных модулей, объединенную через высокоскоростные каналы связи, которая с точки зрения пользователя образует единый аппаратный ресурс. По сравнению с одиночным компьютером он позволяет значительно уменьшить время выполнения расчетов путем разделения задачи на некоторое количество параллельно выполняющихся потоков, каждый из которых обрабатывается отдельным вычислительным модулем.

Общая производительность кластера зависит не только от производительности отдельных вычислительных модулей — нодов, но и от качества и скорости связи между ними. На рынке присутствует большое количество решений по объединению отдельных машин в единую вычислительную сеть с использованием различных стандартных высокоскоростных сетевых интерфейсов, таких как 10G Ethernet, InfiniBand, RapidIO и других. После проведенного анализа был сделан вывод, что применение любого из них существенно увеличивает стоимость конечного устройства и требует разработки специального моста для подключения контроллера интерфейса к эльбрусовским процессорам. Кроме того, принималось во внимание требование разработчиков программного обеспечения по поддержке удаленных DMA-обращений, которому соответствуют не все стандартные интерфейсы. Исходя из этого, было решено разработать собственный сетевой протокол передачи данных и реализующий его контроллер.

Разработанный протокол NIP (Node Interconnect Protocol) предполагает реализацию сетевых функций при совместной работе устройств двух типов – абонентов и сетевых коммутаторов. Определено взаимодействие на трех уровнях (логическом, транспортном и физическом), при котором абоненты ведут обмен посредством установленных операций на логическом уровне, в то время как сетевые коммутаторы обеспечивают маршрутизацию на транспортном уровне. Для этого, в процессе инициализации сети, каждому абоненту присваивается уникальный идентификатор, а в каждом сетевом коммутаторе формируется таблица маршрутизации. Протокол допускает использование нескольких абонентов и/или сетевых коммутаторов в одном вычислительном модуле. В рамках одного кластера поддерживается до 255 абонентов. Дальнейшая масштабируемость сети обеспечивается

через отдельные устройства — межсетевые шлюзы. Топология сети может быть различной: кольцо, звезда или тор. Возможны и другие конфигурации.

Для реализации протокола было создано устройство NIC (Node Interconnect Controller), объединяющее функции логического уровня и сетевого коммутатора. Оно напрямую подключается к процессорам, причем в варианте многопроцессорной оконечной системы возможно подключение контроллера к нескольким процессорам одновременно для более эффективного использования внешних каналов связи. Возможна реализация 4-х внешних связей каждого процессора через его интерфейс с контроллером IO-link, обеспечивающий пропускную способность до 8 Гбит/с.

Сетевой коммутатор выполняет переключение пакетов между портами в соответствии с таблицей маршрутизации. Он поддерживает до 6-ти внешних портов с интерфейсом Wlink и один системный порт. . Интерфейс внешних портов имеет пропускную способность до 6 Гбит/с, причем при замене физического уровня она может быть увеличена до 8 Гбайт/с. Поддерживается два типа маршрутизации: детерминированная — однозначно заданный маршрут, и адаптивная — маршрут определяется исходя из нагрузки портов сетевого коммутатора.

С использованием описанной разработки предполагается создание первого кластера собственного производства, где в качестве вычислительного нода используется четырехпроцессорная плата с 4-х ядерными микропроцессорами Эльбрус-2S. Суммарное число нодов — 60. Для топологии сети выбран 3D-тор. Сетевое взаимодействие осуществляется контроллером NIC с тремя процессорными линками и шестью внешними портами, реализованный на FPGA Cyclone V.

Литература

1. RapidIO Interconnect Specification Version 3.0 / RapidIO Trade Corporation. - 2013. 674 с.
2. Quartus II Handbook Version 12.1 / Altera Corporation. - 2012. 1694 с.
3. Cyclone V Device Handbook / Altera Corporation. - 2014. 1077 с.