DOI 10.37882/2223–2966.2022.04–2.26

ИННОВАЦИОННЫЙ ВАРИАНТ РАЗВИТИЯ ЗАЩИЩЕННЫХ СУПЕР-ЭВМ ДЛЯ РЕШЕНИЯ ВАЖНЫХ ЗАДАЧ ФУНДАМЕНТАЛЬНОЙ МЕДИЦИНЫ И ИНЖЕНЕРИИ В РОССИИ

INNOVATIVE VARIANT OF SECURED SUPERCOMPUTER DEVELOPMENT FOR SOLVING IMPORTANT PROBLEMS OF FUNDAMENTAL MEDICINE AND ENGINEERING IN RUSSIA

A. Molyakov

Summary. The article describes innovative variant of secured supercomputer development for solving important problems of fundamental medicine and engineering in Russia. Secured strategic supercomputer "Angara" is a set of nodes of different types, united by several communication networks. The service nodes are built on conventional superscalar microprocessors. Computing nodes are built on special multicore multithread-stream microprocessors (microprocessors of the J series), are combined into modules in the form of multi-socket boards and can work on a logically single addressable memory (globally addressable memory) formed by local memories of modules with computational nodes.

Keywords: J7, J10, project "Angara", secured supercomputer.

Моляков Андрей Сергеевич
К.т.н., доцент, ФГБОУ ВО «Российский государственный гуманитарный университет», г. Москва
andrei_molyakov@mail.ru

Аннотация. В статье приведено описание инновационного варианта развития защищенных Супер-ЭВМ для решения важных задач фундаментальной медицины и инженерии в России. Защищенный суперкомпьютер стратегического назначения (СКСН) "Ангара" представляет собой множество узлов разного типа, объединенных несколькими коммуникационными сетями. Сервисные узлы строятся на обычных суперскалярных микропроцессорах. Вычислительные узлы строятся на специальных многоядерных мультитредово-потоковых микропроцессорах (микропроцессоры серии J), объединяются в модули в виде многосокетных плат и могут работать над логически единой адресуемой памятью (глобально адресуемой памятью), образуемой локальными памятями модулей с вычислительными узлами.

Ключевые слова: J7, J10, проект "Ангара", защищенный суперкомпьютер.

Введение

СКСН "Ангара" представляет собой множество узлов разного типа, объединенных несколькими коммуникационными сетями, одна из которых обладает уникальным свойством передачи с высокой пропускной способностью больших потоков коротких пакетов. Эта сеть необходима для реализации работы с глобально адресуемой памятью, далее ее будем называть базовой рабочей сетью. Узлы могут быть вычислительными и сервисными, они подключаются к базовой рабочей сети [1, 2]. Вычислительные узлы строятся на специальных многоядерных мультитредово-потоковых микропроцессорах (микропроцессоры серии J), объединяются в модули в виде многосокетных плат и могут работать над логически единой адресуемой памятью (глобально адресуемой памятью), образуемой локальными памятями модулей с вычислительными узлами.

Сервисные узлы строятся на обычных суперскалярных микропроцессорах, выполняют функции ввода-вывода, подключения пользователей, интерфейса с глобальной сетью, а также могут выполнять и вычисления, если они хорошо локализуются и эффективно выполнимы на этих узлах. Вычислительные и сервисные узлы подключены еще к одной сети (RAS-сети), являющейся компонентом подсистемы обеспечения надежности, готовности и сервиса, RAS-подсистемы.

Методология исследования

Новые архитектурно-технологические подходы в области создания защищенных суперкомпьютеров "нового поколения"

Микропроцессор J7 имеет два мультитредовых ядра (MTcore0 и MTcore1), разработан в свое время с учетом особенностей 90 нм технологии ASIC фирмы Fujitsu (серия 101), рассчитан для работы на частоте 1 Ггц. В одном мультитредовом ядре имеются четыре конвейера команд, каждый из которых работает с 16 тредовыми устройствами, на каждом из которых может выполняться один процесс-тред [3].

Выделяются две модели микропроцессоров серии J: младшая — J7, старшая — J10. Каждое тредовое устройство содержит регистры управления выполнением потока команд загруженного на него треда (слово состояния треда TSW со счетчиком команд, признаками и масками исключительных ситуаций и другие регистры), а также достаточно большие наборы 64-разрядных архитектурных регистров для хранения чисел с фиксированной и плавающей точкой, адресуемых однобитовых регистров-признаков и регистров адресов передачи управления. Переключение с выполнения команд одного тредового устройства на выполнение команд с другого происходит без перезагрузки регистров, за один такт процессора.

Кроме того, несколько тредовых устройств могут за такт выдать одновременно команды на выполнение в функциональных устройствах. Информационная зависимость выполняемых команд с одного тредового устройства контролируется аппаратно посредством таблиц признаков занятости регистров.

Один конвейер может выдавать за такт одну или две команды. Две команды могут быть выданы в случае, если одна из них над общими регистрами, а другая — над регистрами с плавающей точкой (в настоящее время разработан, но не реализован, вариант, когда возможна выдача всегда двух команд, независимо от используемых ими регистров).

Готовность к выполнению команды определяется по готовности её регистровых операндов. Если они не ожидают записи результата ранее выданных команд, то команда считается готовой к использованию. Это отслеживается по таблицам резервирования, которые объединены в одном блоке с регистровыми файлами. Один тред конвейера может выдать одну (или две) команды на выполнения раз в четыре такта. Таким образом, всего в микропроцессоре J7 за такт может быть выдано от 8 до 16 команд.

В одном ядре микропроцессора может одновременно выполняться несколько задач. Каждой задаче ставится в соответствие один домен защиты, причем одна из задач ядра — обязательно операционная система. Выполняемая в микропроцессоре задача пользователя может одновременно выполняться в доменах защиты его разных ядер, информация о привязке задачи к доменам защиты хранится в специальной таблице микропроцессора. Микропроцессор J7 называется *"многоядерным мультитредово-потоковым микропроцессором с поддержкой операций над глобально-адресуемой памятью"*.

Основная идея сокрытия задержек, т.е. обеспечения его толерантности, нечувствительности к этим задержкам по развиваемой реальной производительности,— обеспечение высокого темпа выполнения операций с памятью и сетью.

Для этого требуется специальная организация процессора и выполняемых на нем приложений, специальная организация коммуникационной сети, специальная организация памяти. Для всех этих устройств требуется возможность одновременного выполнения большого количества операций и высокая конвейеризация. Естественно, от вычислительной модели приложения требуется возможность выдачи большого количества операций, для этого и нужна мультитредовость.

В реальности задержки памяти и сети могут оказаться до десяти раз больше, темп не всегда удается выдержать равным операция за такт, это особенно относится к коммуникационной сети — сказываются физические ограничения пропускной способности линков передачи данных между узлами внутри сети. По этим причинам выбирается большее количество тредовых устройств, до 64–128 на одно МТ-ядро, что позволяет иметь в одном ядре до 512–1024 одновременно выполняемых обращений к памяти или сети. Дополнительно используются команды обращений к памяти за короткими векторами.

Например, до 8 64-разрядных слов, что позволяет еще больше увеличить количество одновременно выполняемых обращений к памяти и сократить накладные расходы на организацию одного обращения к памяти.

Термин "потоковость" в большей степени применим к архитектуре микропроцессоров J10 и может быть использован в смысле обеспечения возможности обработки потоков данных с использованием моделей графов потоков данных. Поддерживаются две графовые потоковые модели — статические и динамические графы.

Эти модели используются для обеспечения большего параллелизма и асинхронности, а статические графы еще и для снижения количества обращений к памяти при передаче данных между узлами. Узел статического графа появляется вместе со всем графом, функционирует некоторое время и потом удаляется вместе с графом. Узел динамического графа может возникать и уничтожаться в процессе функционирования графа.

В китайском варианте [4, 5] эта возможность сохранена и усилена. Для узла динамического графа такая последовательность поступления данных в дуге может быть нарушена, поэтому данные поступают со специальными тегами, по совпадению которых они могут отыскать себе пару в потоке другой дуги. Выполнение операции может быть в порядке, лишь бы была пара, для которой можно выполнить операцию. Такой отбор соответствующих

друг другу данных в потоках дуг требует применения памяти с ассоциативным доступом, в данном случае ассоциативным адресом является тег данных.

Термин "глобально-адресуемая память" в названии микропроцессора следует понимать следующим образом. В микропроцессорах J7/J10 виртуальная память организована так, что при обращении к ней автоматически распознается при трансляции адреса то, в какой узел системы следует обращаться и производится это обращение, что происходит без участия пользователя.

Результаты исследований

Новые принципы модели безопасности доступа на основе исходящей сборки команд и мультидоменной защиты

С учетом развития методов поиска уязвимостей следует говорить о реализации реактивных методов защиты информации, наряду с превентивными.

Вместо классического понятия «локализованная задача» для суперкомпьютеров (СК) следует говорить о параллельных распределенных потоковых структурах, генерируемых и обрабатываемых на разных уровнях иерархии конвейера команд процессорными устройствами, объединенных высокоскоростными сетями. Атрибуты доступа к объекту и привилегии субъекта, связи между ними формализуются в виде набора (конъюнкции) предикатов. Отслеживать взаимодействие и контролировать доступ можно по набору характерных признаков (маркеров), представленных в виде кортежей логических переменных.

Информационная безопасность основана на контроле доступа к объектам управляющих и гостевых операционных систем (ОС), эти объекты можно отнести к разным уровням защиты.

Традиционный подход контроля доступа предполагает использование атрибутов (прав) доступа в запросах к этим объектам на выполнение некоторых операций над ними. Если проверка таких атрибутов оказывается успешной, то доступ к объекту на его уровне защиты разрешается, далее выполняется запрашиваемая операция над ним. При таком подходе оказывается технически возможным перехват запроса и использование его прав доступа в подменяющем его запросе, нацеленном на вредоносное воздействие.

В работе предложена новая математическая модель обеспечения безопасности, основанная на предложенной им за счет добавления одного уровня защиты 8-уровневой модели и новой логики контроля доступа запросов на выполнение действий с объектами ОС, что реализуется гипервизором с соответствующей организацией, в котором дополнительно используются расширенные автором архитектурные возможности предложенной модели суперкомпьютера.

Практическая значимость заключается в разработке технологии и программно-технических средств на основе предложенных моделей и методов, составляющих принципы реактивной защиты, с учетом внутрисистемных и внешних показателей защищенности и стабильности функционирования высокопроизводительных систем, а также определяется решением следующих прикладных задач:

1. Разработано с учетом специфики СК *принципиально новое технологическое решение*, создающее изолированную среду исполнения программ в виде 8-уровневой "песочницы с реализацией контролирующих механизмов как на уровне гипервизоров, так и на уровне контроллеров транзакционной памяти. Таким образом, решена важная научно-техническая проблема в области создания средств защиты информации для нового класса систем — стационарных и бортовых СК;
2. В ходе экспериментальных исследований получены *новые научные результаты*, подтверждающие эффективность и минимальную потерю производительности применения технологии аппаратной виртуализации в виде многоуровневой "песочницы" для перспективных СК по сравнению с использованием традиционных кластеров;
3. Поскольку невозможно контролировать работу всей аппаратуры, а только осуществлять контроль выполнения запросов на уровне компонентов гипервизора и контроллера транзакционной памяти, в ходе исследований был найден максимальный уровень функционирования агентов СЗИ — S8. При такой конфигурации, когда число уровней иерархии N = 8, выполнение контекстно-зависимых операций становится квазидетерминированным с доверительной вероятностью приблизительно 0.9, а потери производительности менее 6–7 процентов;
4. Предложенное технологическое решение позволяет снизить стоимость и продолжительность разработок для отрасли промышленности за счёт переноса большей части испытаний с опытных образцов на программное обеспечение.

В классических процессорах с архитектурой фон-Неймана совместно используются коды данных и программ, что препятствует эффективному ограничению доступа одного объекта к адресному пространству и данным другого. В них также не реализованы механизмы многоуровневой защиты от атак при выполнении системных вызовов в среде многоуровневого контекста вложенных

гостевых и управляющих операционных систем. Например, тегированная архитектура на примере процессора МЦСТ Эльбрус не поддерживает технологию аппаратной виртуализации. Однако отсутствие поддержки аппаратной виртуализации делает подобные архитектурные решения узкоспециализированными и не поддерживающим эмуляцию разного оборудования и поддержку широко используемых гипервизоров.

Кроме того, ряд особенностей технологии аппаратной виртуализации ускоряет работу виртуальных машин и повышает уровень безопасности.

Аппаратная поддержка виртуализации и механизм многоуровневой защиты может снизить накладные расходы на создание изолированной среды исполнения. Однако в любой операционной системе при сопряжении кода ядра с оборудованием на физическом уровне возникают запрещенные состояния — нулевой кортеж данных, к которым процессор запрещает обращаться даже программам в нулевом кольце защиты.

Кроме того, переключение контекста активных задач в защищенном режиме может выполнять только процессорный модуль, поскольку при теневом копировании данных исполняемого кода программист не может получить прямой доступ к информации. Это требует реализации многоуровневой иерархии обработки запросов.

Данный подход был неприменим к микропроцессорам предыдущего поколения из-за низкой производительности. Введение дополнительных уровней привилегий и уровней защиты сильно замедляло работу системы в целом [6]. Высокая производительность СК, наоборот, позволяет быстро анализировать дескрипторные таблицы, вычислять хеш-значения процессов.

Их отличает: самодиагностика, многоуровневая защита, привязка каждого тредового устройства к определенному домену, программирование с использованием нефункциональных, непроцедурных языков — цепочки вычислений в виде селекторов, подстановок справа, слева функциональных вычислений, многоуровневое распараллеливание алгоритмов и т.д.).

При выполнении программы загрузки тредовое устройство работает в особом режиме привилегий — IPL_LEVEL. В этом режиме используется физическая адресация при доступе к памяти команд и памяти данных. Затем загружаются диспетчер работы с виртуальной инфраструктурой и хостовые ОС с поддержкой гипервизоров. При этом для модулей ядра используется уровень KERNEL_LEVEL, для диспетчера работ с оборудованием –SUPERVISOR_LEVEL. На заключительном этапе осуществляется запуск гостевых ОС на уровне USER_LEVEL.

Единственным «камнем преткновения» в многоядерных многопроцесссорных системах являются задержки проблема эффективной реализации внутрикристальной сети и работы с памятью.

Мультитредовая организация позволяет одновременно выполнять не один, а несколько потоков команд, что дает возможность увеличить множество выполняемых команд, но важнее — усилить поток одновременно выполняемых операций с памятью.

Потоковая архитектура предполагает применение решающих полей элементарных процессоров в виде статических графов потоков данных.

Это позволяет сократить общее количество обращений к памяти, поскольку на решающем поле данные передаются с одного быстрого ресурса на другой без обращения в память [7]. Мы предлагаем метод реконфигурации среды исполнения с учетом требований мобильности и обеспечения удельных характеристик производительности программы для безопасного расширения функциональности системы или приложения.

Только аппаратная поддержка сегментированных стеков может снизить сложность компилятора и накладные расходы на этот механизм во время исполнения. В любой операционной системе при сопряжении кода ядра с оборудованием на физическом уровне возникают запрещенные состояния — нулевой кортеж данных, к которым процессор запрещает обращаться даже программам в нулевом кольце защиты.

Более того, переключение контекста активных задач в защищенном режиме может выполнять только процессорный модуль, поскольку при теневом копировании данных исполняемого кода программист не может получить прямой доступ к информации. Данные коллизии разрешает новый подход — маркерное сканирование, в котором используются генеративные таблицы [8, 9].

Помимо кодирования адресного разделения колец защиты памяти, в ОС разных классов реализован строго типизированный интерфейс сопряжения с аппаратным ядром процессора и управления Контекстом исполнения бинарного кода с учетом профиля компиляции и сборки — использование маркеров системных объектов.

Конвейер обработки команд следующий. После выборки и выдачи каким-либо тредовым устройством команды обращения к памяти, команда попадает в функциональный блок LSU выполнения команд обращения к памяти. Подготовленный в LSU исполнительный адрес обращения к памяти, затем передается в блок MMU,

в котором происходит трансляция виртуального адреса в физический адрес или глобальный виртуальный адрес.

В случае необходимости обращения к памяти удаленного узла, что определяется автоматически в блоке MMU, через блок управления сетевыми сообщениями MSU, внутрикристальную сеть, сетевой интерфейс с коммуникационной сетью происходит передача аппаратно сформированного системного короткого пакета-сообщения с командой обращения к памяти. Дополнительно заметим, что при продвижении по сети некоторые пакеты могут отклоняться от фактически предписанного таблицей маршрутизации пути, обходя, таким образом, всевозможные "пробки", которые могут быть им самостоятельно обнаружены. После такого отклонения пакет при определенных условиях может возвратиться на гарантированно бездедлоковое продвижение по сети.

Глобально-адресуемая память дает не только дополнительные удобства программирования, что, как ожидают специалисты, должно сказаться в повышении продуктивности параллельного программирования приблизительно в 10 раз. Также ожидается и повышение эффективности параллельных программ, поскольку двусторонние модели взаимодействий типа "send-receive", как правило, длинными сообщениями, заменяются односторонними взаимодействиями с использованием коротких сообщений. Реальность достижения большей эффективности при таком переходе к новой модели памяти и организации вычислений доказана уже во многих экспериментах [10].

Между тем, глобально-адресуемая память, используемая множеством параллельных процессов, требует наличия богатых средств синхронизации.

Обычно в качестве таких средств применяются атомарные операции с памятью по типу "считывание-операция-запись", выполняемые в неделимом (атомарном) режиме. В микропроцессорах J7/J10 кроме таких операций используется также аппарат теговых битов и битов управления выполнением обращений к памяти, имеющихся как непосредственно в ячейках памяти, так и в адресах-указателях на них. В программах возможно использование физической и виртуальной адресации. Управление выдачей физических или виртуальных адресов при обращении к памяти производится установкой специального бита в слове состояния треда. Физическая адресация разрешена только в привилегированных режимах ядра ОС и начальной загрузки. Адресация памяти команд и данных осуществляется по разным схемам.

Заключение

В России наиболее продвинутыми являются работы класса "общедоступного уровня", связанные с разработкой сети MBC-экспресс, в которой используется интерфейс PCI-express и коммуникационные микросхемы PLX. Имеются уже две установки, где этот подход реально используется — К100 (ИПМ им. М.В. Келдыша РАН) и ПТК (Санкт-Петербургский Государственный Политехнический Университет). Усовершенствование аппаратных и программных средств MBC-экспресс продолжается, также отрабатываются приемы эффективного параллельного программирования прикладных задач.

Похожие работы начаты в ОАО "НИЦЭВТ", но они ориентированы на использование интерфейса HyperTransport. Кроме этого, в ОАО "НИЦЭВТ" ведется работа класса "умеренного уровня" реализации GAS/PGAS — реализуется маршрутизатор сети N-тор и планируется его усиление мультитредовыми ядрами [11].

Работа такого типа ведется и ГК "Т-платформы" по сети Extoall, также разрабатывается свой вариант микропроцессора, но информации об этих разработках нет [12].

Наши изобретательские и рационализаторские предложения были учтены и успешно реализованы в рамках китайского проекта 863/ИТ по созданию защищенных суперкомпьютерных вычислительных комплексов серии СТ-2 (ОКР шифр «Удар грома»), в рамках японского проекта JST CREST "Разработки новых HPC-технологий" 5-го Базового плана развития науки и технологий по созданию суперкомпьютерных вычислительных комплексов (ОКР шифр «Стрела времени»). Главная доработка — были резко усилены вычислительные возможности посредством введения SIMD операций над короткими векторами и элементов графических процессоров типа синхронно выполняемых тредов в дополнение к асинхронным из J7.

Следует отметить, что работы этого направления в Китае ведутся в NUDT, Национальном университете оборонных технологий Китая. Они имеют серьезные перспективы создания в конечном итоге суперкомпьютера экзафлопсного уровня производительности не только для построения информационных систем, но и для решения научно-технических задач с высоким уровнем реальной производительности, т.е. не с пиковой, а реальной производительностью в экзафлопс.

Наиболее актуален средний уровень: на основе уже имеющихся схем реализации мультитредового ядра J7/J10 провести проектирование архитектуры и микроархитектуры этого ядра, учитывая китайский или японский опыт его доработки и близкие американские работы, провести доработки и приемочное тестирование на фрагментах интересующих специальных задач разных ведомств с целью последующего внедрения в промышленную эксплуатацию в России.

ЛИТЕРАТУРА

1. Based on Reconfiguring the Supercomputers Runtime Environment New Security Methods. A.S. Molyakov / Advances in Science Technology and Engineering Systems Journal, vol. 5, no. 3, pp. 291–298 (2020).
2. Main Scientific and Technological Problems in the Field of Architectural Solutions for Supercomputers. A.S. Molyakov / Computer and Information Science, Vol. 13, No. 3 (2020).— DOI: 10.5539/cis.v13n3p89.
3. Семенов А.С. Разработка и исследование архитектуры глобально адресуемой памяти мультитредово-потокового суперкомпьютера. Диссертация на соискание ученой степени кандидата технических наук. Специальность 05.13.15 — Вычислительные машины, комплексы и компьютерные сети. Научный руководитель — Эйсымонт Л.К. Москва 2010, 224 стр., защищена в декабре 2010 года, утверждена ВАК в 2011 году.
4. China Net: Military and Special Supercomputer Centers. A.S. Molyakov / Journal of Electrical and Electronic Engineering 7 (4), 95–100, 2019.— DOI: 10.11648/j.jeee.20190704.12.
5. Age of Great Chinese Dragon: Supercomputer Centers and High Performance Computing. A.S. Molyakov / Journal of Electrical and Electronic Engineering 7 (4), 87–94, 2019.— DOI: 10.11648/j.jeee.20190704.11.
6. New Multilevel Architecture of Secured Supercomputers. AS Molyakov / Current Trends in Computer Sciences & Applications 1(3) — 2019.— ISSN: 2643–6744. DOI:10.32474/CTCSA.2019.01.000112.
7. Моляков А.С. Супер-ЭВМ и операционные системы нового поколения / Информационная безопасность: вчера, сегодня, завтра. Международная научно-практическая конференция, Москва, 23 апреля 2019 г.: сборник статей РГГУ.— М., 2019.— С. 196–200.
8. Molyakov, A.S. A Prototype Computer with Non-von Neumann Architecture Based on Strategic Domestic J7 Microprocessor. Automatic Control and Computer Sciences.— 2016.— № 50(8).— PP. 682–686.
9. Molyakov, A.S. Token Scanning as a New Scientific Approach in the Creation of Protected Systems: A New Generation OS MICROTEK. Automatic Control and Computer Sciences.— 2016.— № 50(8).— PP. 687–692.
10. Горбунов В., Эйсымонт Л. Экзафлопсный барьер: проблемы и решения. «Открытые системы», № 6, 2010, с. 12–15.
11. Л. DARPA UHPC — дорога к экзафлопсам. «Открытые системы», № 12, 2010, http://www.osp.ru/.
12. Горбунов В., Елизаров Г. Эйсымонт Л. HPC: региональные новости. «Открытые системы», № 2, 2011, с. 12–16.

Журнал «Современная наука: актуальные проблемы теории и практики»

Российский Государственный Гуманитарный Университет