Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.75

Грачёв
Ян Юрьевич

Использование мобильных устройств в качестве сопроцессоров для распараллеливания вычислений

АВТОРЕФЕРАТ
на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Петровский Н.А.
к.т.н., доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

В настоящее время исследования и обучение зависит от высокопроизводительных вычислений. Применение в самых разных научных областях требует огромного масштаба вычислительных систем с поддержкой параллельных вычислений на огромном количестве вычислительных узлов в течении длительного времени. Невероятные высокопроизводительные вычислительные системы сегодня, уже завтра могут стать настольными и повседневными, как было с суперкомпьютерами 1980ых годов, которые фактически сейчас в руках каждого: в мобильных телефонах, автомобилях и просто компьютерах. Новые подходы к вычислительным системам позволили в 4-5 раз, а то и более, ускорить появление научных открытий и достижений.

Современные технологии предъявляют возрастающие требования к вычислительным мощностям. Большинство решений использует технологии CUDA на GPU, сопроцессоры Intel Xeon Phi, но все больше внимание уделяется ARM процессорам. Их выпускается огромное количество, что делает их весьма доступными для построение неоднородных распределенных вычислительных систем. Появляются вопросы финансовой и производительной эффективности их применения. Наиболее распространены процессоры x86 и ARM архитектур, которые имеют принципиальные различия в производительности и энергоэффективности. Построение вычислительных систем с использованием мобильных устройств на процессорах ARM позволит сэкономить, не теряя производительность, а иногда и превосходя x86, в задачах, где количество параллельных вычислений чрезвычайно высоко и разнообразно. С помощью сетевых и облачных технологий подобная система получает большую масштабируемость.

Обычно вычислительные системы на архитектуре x86 создаются с использованием одинаковых по характеристикам процессоров, что для ARM систем практически невозможно из-за разнообразных типов и поколений. Таким образом, требующими ответ вопросами являются вычислительная эффективность ARM процессоров в сравнении с x86 для разных типов задач, применение полученных вычислительных мощностей для решения практических задач, а также вопросы организации неоднородных распределенных вычислительных систем на основе мобильных устройств.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является исследование возможностей и эффективности использования мобильных устройств при распараллеливании задач криптографии.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать эффективность параллельного применения ARM процессоров для криптографических вычислений.
2. Разработать программное обеспечения для использования процессоров мобильных устройств в качестве неоднородной распределенной вычислительной системы.

Объектом исследования являются процессоры мобильных устройств.

Предметом исследования является распараллеливание криптографических вычислений на мобильные процессоры для выявления эффективности и создания неоднородных распределенных вычислительных систем.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования мобильных устройств на ARM процессорах в качестве сопроцессоров для распараллеливания вычислений.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н. А. Петровского, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 54-й научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2018); 55-й юбилейной научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2019).

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатные работы в сборниках материалов конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования. Вторая

глава посвящена анализу существующих решений их анализе. В третьей главе предложена архитектура неоднородной распределенной вычислительной системы на основе мобильных устройств, рассмотрены метрики ее производительности, осуществлены и представлены результаты экспериментальных исследований.

Общий объем работы составляет 63 страницы, из которых основного текста 53 страницы, 22 рисунка на 18 страницах, список использованных источников из 60 наименований на 5 страницах и 5 приложений на 10 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** рассмотрены вопросы энергоэффективности и производительности вычислительных систем.

При наличии увеличивающейся необходимости достигать высокой вычислительной производительности и решать комплексные вычислительные проблемы, распределенные вычислительные системы становятся все более значимыми в виду высокой эффективности и возможности закрывать практически любые запросы, динамически распределяя имеющиеся мощности.

Одним из направления развития вычислительных систем и мобильных устройств является парадигма переноса высокопроизводительных вычислений с мобильных устройств на удаленные вычислительные центры, когда мобильное устройство само по себе лишь отображает результат и является связующим звеном ввода-вывода между пользователем и удаленной распределенной вычислительной системой. Это позволяет пользователю мобильных устройств осуществлять выполнение задач, которые в ином случае были бы полностью недоступны из-за необходимости выполнять все вычисления на самом устройстве в виду ограниченности ресурсов.

ARM процессоры RISC архитектуры преимущественно применяются на мобильных платформах, включая смартфоны, планшеты и устройства «интернета вещей», обычно работая с 32 битным набором инструкций. Процессоры AMD и Intel применяют CISC архитектуру с поддержкой аппаратной виртуализации и 64 битным набором инструкций. Традиционно ARM разрабатывался с целью обеспечения энергоэффективности, в то время как серверные решения были ориентированы на производительность.

ARM процессоры являются крайне неоднородными и имеют огромное множество различных реализаций, также применяется технология big.LITTLE – разделение ядер на энергосберегающие малые и высокопроизводительные большие в пределах одного процессора.

Сравнивая производительность между серверными x86 и ARM устройствами, ARM является в 3-4 раза более эффективным с точки зрения производительности к потребляемой энергии.

Результаты исследований, проведенных в этих направлениях, отражены в работах В. Дирлевангера (W. Dirlewanger), Н. Фернандо (N. Fernando), Д. Майр (J. Mair), Л. Гвеннапа (L. Gwennap), К. и Э. Атигхехчи (K. & E. Atighehchi), Д. Калуанасундарам (J. Kalyanasundaram), А. Лин (I. Lin), С. Миттал (S. Mittal) и др.

Вторая глава посвящена рассмотрению и оценке существующих решений для создания вычислительных систем, а также их архитектур.

GPU применяются для обработки графики и включает в себя две важные особенности: множество параллельных вычислительных ядер для увеличения производительности и быстрая память достаточного объема. После попыток использования этих мощностей для решения различных задач возникло новое направление универсальных вычислений на GPU. Для этих целей и взаимодействия с вычислительными мощностями используются технологии OpenCL и CUDA.

Использование GPU для криптографии целесообразно из-за возможности выполнения сложных параллельны вычислений. В 2007 году была реализован алгоритм AES для работы с технологией CUDA, в 2008 был реализован алгоритм MD5 также с использованием этой же технологии для генераций белого шума.

Xeon Phi от Intel представляет из себя многоядерную распределенную вычислительную систему с возможностью выполнять одну инструкцию на множестве данных, а также с поддержкой одновременной многопоточности. Популярность обусловила наличие множества методов и готовых отлаженных библиотек для применения на этом сопроцессоре в целях ускорения достижения научных открытий.

Мультипроцессорные самодостаточные вычислительные системы x86 и x86-64 с объединением нескольких процессоров, использующих архитектуру NUMA для решения конфликтов и оптимизации доступа к памяти.

FPGA является удобной платформой для разработки и применения, так как отсутствует необходимость физически переделывать плату и достаточно лишь изменить программное ядро, в отличии от ASIC устройств. С аппаратной точки зрения сама сущность FPGA позволяет организовать из них параллельные вычислительные системы. При этом можно использовать все преимущества: аппаратную производительность, возможность модификации арифметической логики на уровне битов, высокое соотношение производительности к энергопотреблению.

Основным недостатком использования FPGA являются довольно низкие частоты работы и трудности поддержки плавающей точки в сравнении с процессорами или GPU. Для решения этой проблемы обычно выстраиваются схемы работы FPGA в связке с процессором. Однако наиболее рационально рассматривать работу в связке с GPU.

Из всех рассмотренных решений наиболее оптимальными с точки зрения организации распределенных вычислительных систем являются GPU, мультипроцессорные системы на x86, а также FPGFA. Общим и ключевым недостатком всех рассмотренных решений является необходимость организации и поддержания сложной и дорогостоящей инфраструктуры с множеством потенциальных слабых узлов.

В **третьей главе** предложена архитектура неоднородной распределенной вычислительной системы с использованием мобильных ARM устройств в качестве вычислительных узлов, проанализирована производительность мобильных устройств для решения криптографических задач, дана оценка энергоэффективности вычислений на мобильных устройствах.

При построении неоднородной вычислительной системы вопрос коммуникации между ее вычислительными единицами или коммуникации между устройством управления и вычислительными стоит очень остро в виду потенциально огромного разнообразия, как аппаратного, так и программного обеспечения устройств.

Основное достоинство неоднородной распределенной вычислительной системы одновременно является и недостатком, так как использование разных мобильных устройств накладывает ограничения не столько на вычислительные мощности, сколько на количество необходимых для вычислений данных. Для неоднородной распределенной вычислительной системы, которая для коммуникации использует беспроводное соединение временные задержки являются ключевым вопросом в возможности сравнения и выполнения вычислений для реальных задач, а не исключительно отдельных тестов. Главное ограничение накладывается на объем данных передаваемых между узлами или к хосту, несмотря на потенциальную возможность работы сложных и требовательных к производительности алгоритмов обработки графики, машинного обучения, работы с базами данных и других на мобильном ARM устройстве. Несмотря на это, возможно применение для отдельных задач в криптографии, где не требуется передача больших данных.

К параметрам, определяющим неоднородную распределенную вычислительную систему, можно отнести: количество ядер процессора, объем оперативной памяти. В виду того, что необходим тест для оценки общей производительности, рационально использовать криптографию – подсчет хэшей.

Введен показатель энергоэффективности, позволяющий сравнивать эффективность вычислений между мобильными ARM устройствами с разной архитектурой. Для сравнения энергоэффективности устройств с разными архитектурами требуется ввести формулу (1) ее расчета, которая отражает зависимость потребления ресурсов процессора и количество высчитанных за это время хэшей MD5

$$E = {}^{c}/_{u}\,, \tag{1}$$

где c – [количество вычислений MD5]/c,

u – процент использование ресурсов CPU.

Ограничивающим производительность фактором являются различные аппаратные и программные реализации функций энергосбережения, но они также способствуют повышению эффективности вычислений в виду меньшего тепловыделения и меньшего потребления мощности. Применяются механизмы IPA и EAS для умного распределения задач между доступными вычислительными

мощностями с учетом максимального энергосбережения и ограниченного тепловыделения в пределах всей системы. Исходя из этого можно выделить три основных режима работы устройства в зависимости от состояния батареи:

– устройство работает от батареи, режим энергосбережения выключен;

– устройство работает от батареи, режим энергосбережения включен;

– устройство работает от сети, режим энергосбережения выключен.

Ключевым требованием к реализации является оценка вычислительной производительности каждого устройства на высоком уровне при условии работы вместе с диспетчером задач операционной системы, так как используются устройства потребительского класса и принципиально не рассматривается внесение каких-либо дополнительных изменений в стандартную работу системы.

Диспетчеризация вычислительных задач производится на двух уровнях: главного узла и конкретного устройства. На уровне главного узла основополагающей задачей является выбор мобильного устройства из числа неоднородных по производительности, которое решит поставленную задачу за необходимое ожидаемое время с минимальным потреблением энергии. Также на этом уровне должно осуществляется деление задачи на атомарные подзадачи.

Для реализации алгоритма оценки производительности применяется среда разработки Android Studio и язык программирования Java. Выбор Java продиктован тем, что виртуальная машина Android работает именно с ее использованием, так и потому что большинство приложений разрабатывается именно с ее использованием.

Разница в производительности протестированных мобильных ARM устройств для однопоточного режима составляет от 2 до 20 раз, рисунок 1.

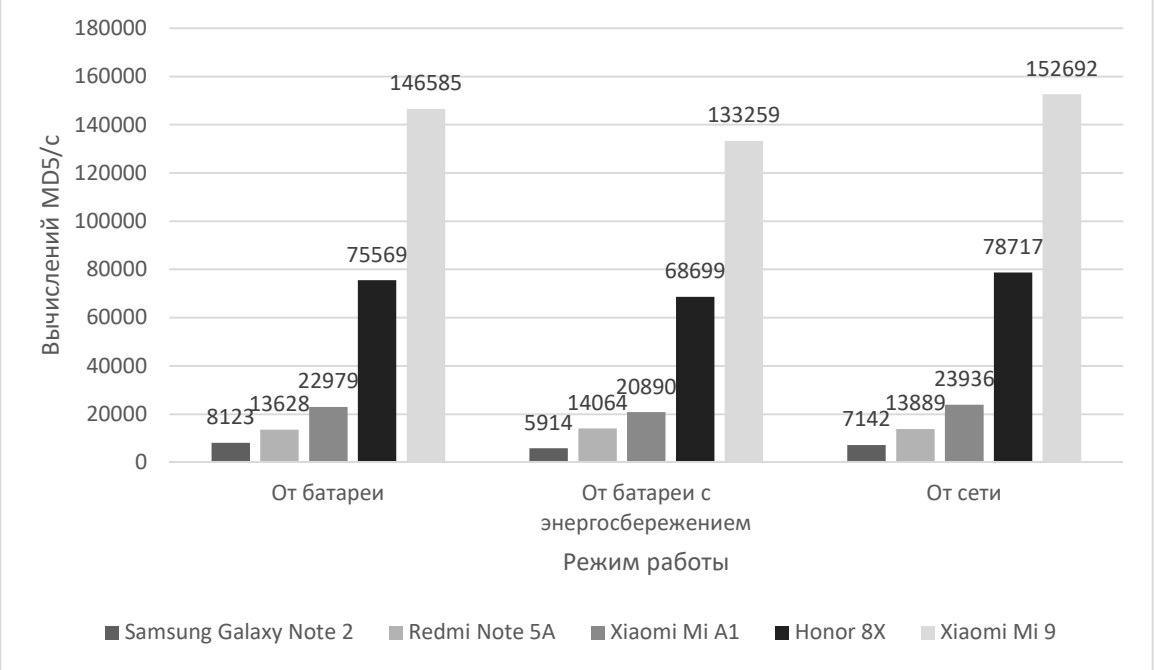

Рисунок 1 – Результаты тестирование производительности мобильных устройств в однопоточном режиме

Сравнивая производительность между мобильными устройствами на Android с ARM и эмулятором на x86 наиболее интересным представляется сравнение с результатами работы аналогичного алгоритма оценки производительности на обычном реальном компьютере, рисунок 2.

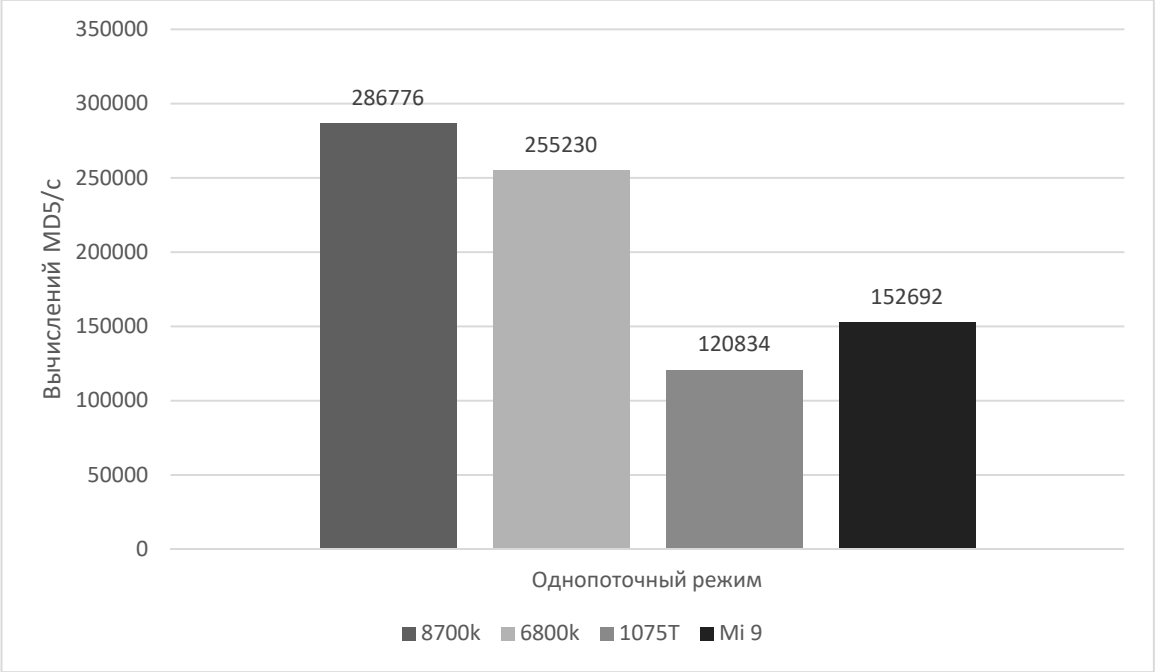

Рисунок 2 – Сравнение производительности x86-64 и ARM

Достаточно сравнения Intel 8700k с наиболее высокопроизводительным и энергоэффективным мобильным устройством Xiaomi Mi 9, так разница в пользу полноценного процессора настольного компьютера в однопоточном режиме составляет всего 88%, при значительно большем энергопотреблении из-за особенностей самой архитектуры x86-64. Но если рассмотреть более старый шестиядерный процессор 2011 года от AMD Phenom II x6 1075T архитектуры x86-64, его производительность будет уступать при решении аналогичной задачи на мобильном ARM устройстве на 26%.

Рассмотренными результатами подтверждается возможность и практический смысл применения мобильных устройств для распараллеливания вычислений в составе неоднородной распределенной вычислительной системы, с целью осуществления криптографических вычислений.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Предложена и рассмотрена архитектура неоднородной распределенной вычислительной системы с использованием мобильных устройств в качестве вычислительных узлов для решения задач не требующих передачи больших объемов данных.

2. Предложен универсальный метод для тестирования производительности

мобильных ARM устройств, позволяющий сравнивать производительность устройств в зависимости от производительности процессора и работы операционной системы. Применение устройств на самых современных процессорах ARM позволяет приблизиться к производительности x86, однако использование наиболее высокопроизводительных мобильных устройств является неэффективным, так как при сходных характеристиках преимущества x86 неоспоримы. Обратная ситуация наблюдается для устройств на ARM процессорах из нижнего и средних сегментов, чьи процессоры в вычислительной мощности в одиночку, безусловно, не могут составить конкуренцию с x86, но при объединении в неоднородные распределенные вычислительные сети производительность будет сходной с более лучшей энергоэффективностью вычислений.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют практическую базу для разработки программного обеспечения неоднородных распределенных вычислительных систем с использованием мобильных устройств в качестве вычислительных узлов.
2. Областью применения неоднородной распределенной вычислительной системы в первую очередь являются образовательные и научные проекты, которым требуется возможность организации каких-либо превосходящих стандартный компьютер на x86 решений в плане производительности. Рационально рассматривать данное решение для задач, где часто требуется выполнение длительных нагрузок без четкой привязки ко времени исполнения.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Грачев, Я. Ю. Неоднородная распределенная вычислительная система / Я. Ю. Грачев // Компьютерные системы и сети: материалы 54-й научной конференции аспирантов, магистрантов и студентов, Минск, 23 – 27 апреля 2018 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2018. – С. 58 - 59.

3. Грачев, Я. Ю. Применение мобильных устройств для криптографических вычислений / Я. Ю. Грачев // Компьютерные системы и сети: 55-я юбилейная научная конференция аспирантов, магистрантов и студентов, Минск, 22-26 апреля 2019 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2019. – С. 91 - 92.