

О возможностях параллельной обработки данных в ГРИД-системе хранения медицинских изображений для повышения эффективности диагностики.

С.В. Баранник¹, А.Л. Головинский², А.В. Демин¹, А.Л. Маленко², Ю.А. Петровский¹

¹ Институт сцинтилляционных материалов НАН Украины

² Институт кибернетики им. В.М. Глушкова НАН Украины

svbarannik@gmail.com, golovinsky.andriy@gmail.com, demin@isma.kharkov.ua,
exipilis@yandex.ru, petrovsky@isma.kharkov.ua

Аннотация. В работе рассмотрена система хранения и обработки медицинских изображений с использованием Грид-технологий, создаваемая в настоящее время в Институте сцинтилляционных материалов в сотрудничестве с другими учреждениями НАНУ и АМН. Большой объем данных, структурированность и распределенность и требования к надежности и доступности требуют использования ГРИД технологий. В качестве стандарта записи медицинской информации используется DICOM. Данная система может использоваться не только для хранения и обработки изображений, но и для решения статистических и эпидемиологических задач. Особое внимание уделено возможностям параллельной обработки медицинских данных, накопленных в рассматриваемой системе.

Ключевые слова

Грид, Медицинская визуализация, DICOM, Медицинские изображения, Распределенное хранение, Распределенные вычисления

1 Введение

С 2010 года в Украине осуществляется Национальная научно-техническая программа внедрения и использования ГРИД-технологий. Наибольшее применение ГРИД-технологии нашли в области физики высоких энергий. Тем не менее, в рамках Программы поддерживаются грид-проекты и для других областей знаний, в том числе и медицины [1]. Одним из них стал проект «Создание системы хранения медицинских изображений с использованием Грид-технологий». Институт сцинтилляционных материалов на протяжении последних 17 лет проводит работы в области проектирования и производства сцинтилляционного диагностического оборудования для ядерной медицины. В настоящее время в разных городах Украины установлены более 40 систем радионуклидной диагностики разработанных и произведенных в ИСМА. В последние годы в ИСМА активно развивается кластер ГРИД. В настоящее время он занимает второе место в Украине по вычислительным возможностям. Эти 2 фактора явились предпосылкой для создания системы хранения медицинских изображений с использованием грид- технологий.

Для решения задач, связанных с хранением и обработкой медицинских данных была создана виртуальная организация MedGrid. Организация создана в рамках проекта накопления и популяционной обработки кардиологических данных и данного проекта. В настоящее время ее поддерживают 5 кластеров Украинского Национального Грида.

При постановке диагноза и проведения лечения врачи все больше полагаются на медицинские изображения. Медицинские изображения дают основной объем информации о пациенте, однако, сами по себе они недостаточны, поскольку нужно их анализ и интерпретация в контексте истории болезни пациента. Хотя современное оборудование позволяет получать медицинские изображения в цифровом виде, долговременное хранение в Украине не производится. Медицинские данные о пациенте собираются в различных лечебных учреждениях. Врачи часто не имеют доступа ко всем историям болезней всех своих пациентов. Часто изображений одной модальности оказывается недостаточно, поскольку на получение изображения влияет

множество параметров, и дополнительная информация накапливается различными системами сбора физических данных. Эти недостатки должна преодолеть ГРИД-система.

2 Построение системы хранения медицинских изображений

Грид-технологии применяют для проектов, требующих больших вычислительных мощностей и хранилищ данных большого объема, работающих с географически распределенной информацией или требующих быстрого гарантированного доступа. Для создаваемой системы основным ресурсом является объем распределенной информации и скорость доступа. Грид-инфраструктура позволяет достичь географической распределенности, гарантированности и скорости доступа. В распределенной системе обеспечивается заданный уровень избыточности, что позволяет восстанавливать данные при падении любого из узлов и увеличивать скорость доступа к данным [2].

Медицинские изображения представляются огромными объемами данных: одно изображение может занимать от нескольких мегабайт до одного гигабайта и более. Объемы такой информации ежегодно возрастают. Медицинская информация требует также компьютерной обработки, масштабирования и создания структурированных архивов. На сегодня снимки, как часть медицинского документа, не следуют за пациентом в процессе обследования или лечения из клиники в клинику, медицинские данные рассеяны по местам проведения обследований. При лечении, как правило, нет возможности получить полную медицинскую информацию о пациенте в историческом разрезе. После утверждения стандартов хранения и передачи медицинской информации и внедрения электронных карт пациентов появится возможность перейти от бумажного к электронному учету медицинских документов, который можно будет легко передавать из клиники в клинику вслед за пациентом. Доступ к этой информации должен быть ограничен, данные должны быть закодированы для гарантии их безопасности, но форма стандартизованных клинических протоколов должна быть понятна для врачей независимо от места и страны пребывания пациента. Согласно действующему законодательству, объективная медицинская информация должна храниться в архиве в течение 20 лет после последнего обращения пациента в лечебное учреждение. Применение этой нормы к цифровым медицинским изображениям, которые безусловно являются объективной медицинской информацией, означает, что хранить эти изображения следует в течение всей жизни пациента и еще 20 лет после его смерти.

Решение задачи распределения данных предполагает построение огромной, многоцентровой интегрированной базы данных. Для реализации алгоритмов обработки и для хранения медицинских данных сегодня используются автоматизированные системы хранения данных — PACS (Picture Archival and Communications System). PACS-системы определяют как специализированные информационные системы, которые рассчитаны для работы с большими объемами данных радиологических, генетических исследований и медицинской графики [3].

В настоящее время создана распределенная база данных и хранилище изображений (Рис. 1). Происходит накопление медицинских изображений и разработка алгоритмов параллельной обработки [4].

Одна из основных проблем при использовании Грид-технологий в медицине состоит в том, что практически ни у кого из врачей в настоящее время нет пользовательских Грид-сертификатов. В данной системе эта проблема решается за счет использования WEB-портала с сертификатом сервиса. Каждый врач авторизуется на портале, что позволяет ему работать с системой, загружать исследования и просматривать публичные данные других пациентов. Врач имеет доступ только к исследованиям, которые он загрузил, или к тем, ссылки на которые ему сообщает другой пользователь системы, или пациент. Врач может работать с базой данных с помощью обычного Интернет браузера, однако удобнее использовать специализированные клиенты доступа.

От медицинского оборудования данные поступают в хранилища в формате DICOM [5]. Формируется уникальный случайный идентификатор и QR-код анализа, предоставляемые пациенту для доступа к собственным анализам в будущем. Данные обрабатываются, отделяется изображения и мета-информация. Изображения сохраняются в хранилище в виде файлов, а мета-информация поступает в базу данных. Полученные данные сервер DICOM пересылает на парсер, который выделяет поля для занесения в базу данных. После этого данные проходят сквозь модуль деперсонификации. Модуль фильтрует поля DICOM, и удаляет из них персональную информацию. Такие поля не сохраняются в базе данных. Кроме того модуль деперсонификации уничтожает поля с персональными данными из структуры DICOM-файла. После сохранения новый файл добавляется в хранилище исследований. В качестве каталога этого хранилища выступает разработанная в ЦЕРН служба LFC. Это позволяет эффективно управлять хранением данных, их репликацией и балансировкой нагрузки.

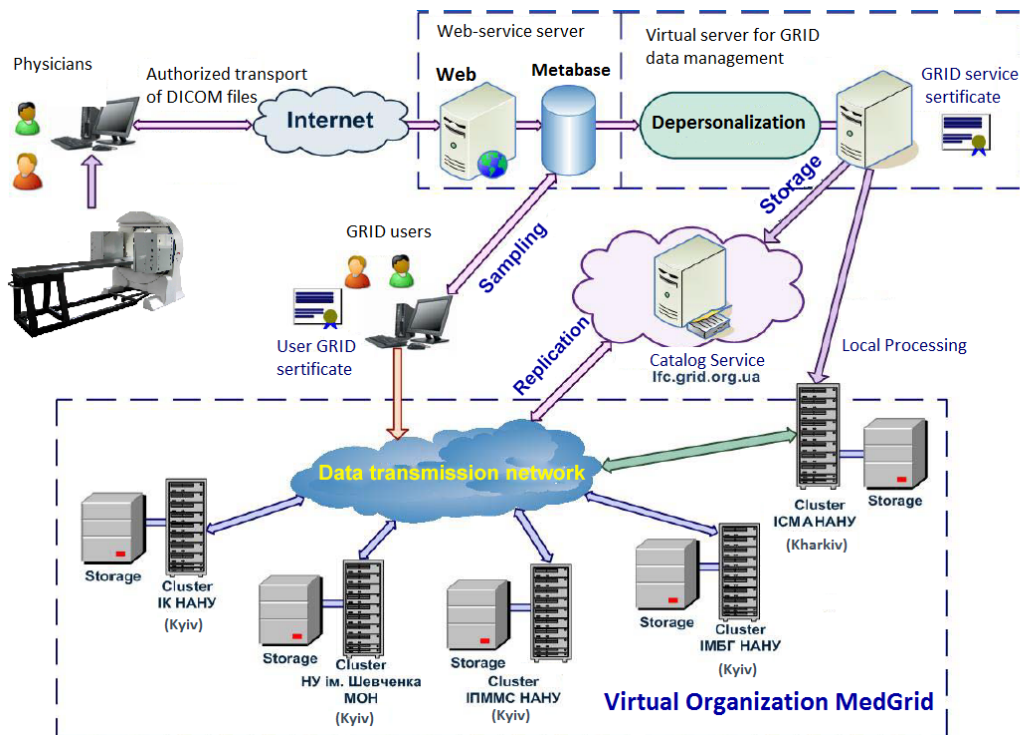


Рис. 1 Схема передачи данных в системе хранения медицинских изображений с использованием грид-технологий

3 Параллельная обработка медицинских изображений

В настоящее время в большинстве программ для медицинской томографической реконструкции используется интегральные методы и, прежде всего, метод свертки и обратной проекции. В то же время, алгебраические методы дают ощутимо лучшее качество изображения, меньшее количество артефактов и лучшее соотношение сигнал/шум. Их меньшая распространенность объясняется большой ресурсоемкостью алгебраических алгоритмов. Нахождение трехмерного томографического распределения в алгебраических методах сводится к решению системы большого количества линейных уравнений. Число уравнений имеет порядок N^3 , где N — размер матрицы проекций. Так для $N=128$, что характерно для радиоизотопных исследований, число уравнений составляет 2 М, а для $N=1024$ (рентгеновская томография) достигает 1 G. Облегчает задачу то, что томографическая система уравнений сильно разрежена и распадается на отдельные независимые блоки. Это также способствует хорошему распараллеливанию данных алгоритмов.

Для решения таких систем уравнений применяются итерационные алгоритмы. Нами был выбран алгоритм OSEM (Ordered Subset Expectation Maximization). На каждом шаге итерации увеличивается ожидаемое правдоподобие того, что вычисленное трехмерное распределение будет соответствовать наблюдаемым двумерным проекциям. Алгоритм был предложен Хадсоном и Ларкиным еще в 1994 [6] и в настоящее время рассматривается как самый перспективный с точки зрения качества реконструкции. Реализация алгоритма для последовательного вычисления показала характерное время порядка 20 минут на процессоре Intel Core i5 3 GHz. После распараллеливания и адаптации к Грид, время расчета снизилось до 5-10 секунд, что соответствует увеличению производительности в 100-200 раз.

Главным недостатком Грид вычислений для медицинской визуализации является пакетная организация вычислений в Грид и, соответственно, плохая приспособленность к интерактивным приложениям. Так в данном случае, время ожидания задачи в Грид очереди может в десятки раз превышать собственно время вычислений. Таким образом, результат томографической реконструкции может быть доступен лишь через несколько минут, вместо нескольких секунд. Для преодоления этой проблемы было использовано следующее решение. Задача томографической реконструкции запускается Грид-системой автоматически, сразу после добавления в хранилище DICOM-файла, который содержит томографические проекции. Если такой файл поступает с томографа непосредственно после завершения сканирования, то за время, пока пациент встанет с установки и дойдет до кабинета врача, результат томографической реконструкции будет уже готов. Кроме того, он может быть передан в кэширующую PACS госпитальной медицинской информационной системы (МИС) для еще более оперативного доступа. Однако вопросы взаимодействия Грид-системы хранения медицинских изображений и МИС еще не рассматривались в рамках текущего этапа.

Еще одним способом Грид обработки изображений в хранилищах является нахождение сходных изображений. В общем виде данная задача тесно связана с распознаванием изображений и автоматическими методами постановки диагноза и далеко выходит за рамки проекта.

В рамках данного проекта был предложен метод Грид обработки для нужд дифференциальной диагностики. Одним из первых этапов такой диагностики является составление списка заболеваний, возможных в случае данного пациента, для последующего дифференцирования. В этом и может помочь Грид система, которая разрабатывается. Алгоритм начинается с того, что врач обводит аномальную область на медицинском изображении. Изображение с очерченной аномалией передается в Грид систему. Следующий шаг — поиск по базе данных. При этом отбираются изображения, аналогичные данному. Т.е. имеющие ту же модальность, исследующие тот же орган, с теми же препаратами и т.д. Таким образом из всего накопленного массива отсекаются те изображения, на которых аномалии, которая интересует, не может быть в принципе. Выборка производится по метаданным, которые хранятся в базе, следовательно, производится быстро. Оставшиеся в выборке изображения исследуются на предмет нахождения в них указанной аномалии. Обработка происходит на уровне матриц данных, что требует больших вычислительных ресурсов. Именно на этом этапе Грид вычисления задействуются с максимальной эффективностью. То, что медицинские изображения хранятся на разных кластерах позволяет разбросать обработку разных изображений по разным кластерам. Избыточность хранения данных также способствует распределению вычислений. В результате мы получаем набор изображений в которых присутствует заданная аномалия, либо структура схожая с этой аномалией.

В том случае, когда этот набор будет состоять из малого количества изображений (ориентировочно до десятка) все изображения из него могут быть предоставлены врачу, как представляющие интерес. Однако такие случаи будут редкими, как правило система будет отбирать тысячи, или даже десятки тысяч, изображений. В этом случае из DICOM файлов будут выбраны диагнозы. После удаления из списка диагнозов всех повторений в нем останется совсем небольшое количество элементов. Это будут те диагнозы, которые ранее ставились пациентам с аналогичными изменениями на изображениях. Таким образом, мы получаем список диагнозов для последующей дифференциальной диагностики.

4 Выводы

Новое решение позволяет строить гибкие и масштабируемые системы хранения данных как для мелких, так и для крупных учреждений здравоохранения и сетей медицинских учреждений. Медицинские учреждения получают возможность через свои PACS-системы и средства визуализации обращаться к распределенным системам хранения медицинских изображений, используя открытые системы и технологии Грид в сочетании с мощными возможностями обеспечения целостности данных. Таким образом, медицинские учреждения могут совместно использовать распределенные ресурсы хранения данных, оптимизируя загрузки оборудования, упрощая управление и сокращая затраты на хранение данных.

Накопление медицинских изображений в Грид позволяет организовать их эффективную параллельную обработку. В настоящее время реализованы алгоритмы томографической реконструкции и предложен метод поиска похожих изображений для нужд дифференциальной диагностики.

Литература

- [1] В. Авраменко, А. Загородний, Е. Мартынов Особенности применения ГРИД-технологий в медицине. *Вестник НАН Украины*. N10: 5-15, 2008.
- [2] Hasting S., Oster S., Langella S., Kurs T., Pan T., Catalyurek U.V., Saltz J.H. A grid based image archival and analysis system. *Am. Med. Inform. Assoc.*, 12(3):286-295, 2005.
- [3] Ходжибаев А.М., Адылова Ф.Т. Новейшие информационные ГРИД-технологии в электронной медицине. *Укр.ж.телемед.мед.телемат.*.3(1):23-24, 2005
- [4] Barannik S., Dyomin A. GRID as a Storage for Medical Information. *International Conference Nuclear Medicine: Physics, Engineering and Practice 2011, Kharkov, Ukraine, September 19-21 2011*
- [5] Kovalenko A. Application of the standard DICOM in exchange medical data systems. *International Conference Nuclear Medicine: Physics, Engineering and Practice 2011, Kharkov, Ukraine, September 19-21 2011*
- [6] H.M. Hudson, R.S. Larkin Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Medical Imaging*, 13 (4): 601–609, 1994