

ТЕХНОЛОГИЧЕСКИЕ НАПРАВЛЕНИЯ ОБЕСПЕЧЕНИЯ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ В ГЕОИНФОРМАЦИОННЫХ ИССЛЕДОВАНИЯХ

И. В. БЫЧКОВ, А. Е. ХМЕЛЬНОВ

*Институт динамики систем и теории управления СО РАН,
Иркутск, Россия*

А. Д. КИТОВ

Институт географии СО РАН, Иркутск, Россия
e-mail: bychkov@icc.ru, kitov@irigs.irk.ru

Principles of organization of the distributed computing structures such as GRID which are capable to perform parallel calculations for processing the geographical information, including GRID-data, are discussed. Such structures and the technologies of their processing allow increasing the efficiency of computer networks, organizing large databases, providing access and organizing distributed processing of large volumes of cartographical and remote data (geoimages) as well as submitting data in the form, which is convenient for analysis and visualization.

Развитая компьютерная сеть, большие базы данных, в том числе картографических, требование к повышению уровня интеллектуализации и интеграции обработки пространственных данных, наличие распределенных вычислительных ресурсов делают актуальным решение фундаментальной проблемы организации выполнения взаимосвязанных программ с соответствующим информационным обеспечением, расположенных в хранилищах обобществленного знания и данных.

Одним из возможных путей ее решения (достаточно популярным в настоящее время) является реализация виртуальной организации междисциплинарного взаимодействия в сфере использования высокопроизводительных информационно-вычислительных систем и коммуникаций для решения научных, производственных и образовательных проблем. В сообществе исследователей в данной области введен термин “GRID”, который в отношении рассматриваемой проблемы с учетом ГИС-анализа имеет несколько значений. С одной стороны, это GRID-технология, обеспечивающая распределенные вычисления (это аббревиатура от Global Resource Information Distribution), с другой — GRID-информационное поле (grid — ячейка), представляющее ячеечную структуру данных [1]. Заметим, что в программировании этот термин относится к методам обработки и представления электронных таблиц.

Таким образом, GRID-технологии в контексте первого определения рассматриваются как методы и средства, обеспечивающие пользователя вычислительными мощностями, распределенными в рамках корпоративной информационной сети (Intranet), с возможностью использования программных компонентов и/или информационных баз для решения заданного класса задач. Одна из главных задач GRID-технологий состоит в предоставлении компьютерному сообществу новых средств взаимодействия и формирования виртуальных альянсов специалистов с целью решения практических задач. Это достигается на основе распределенного управления ресурсами и их интеграции, что в итоге формирует метакомпьютерную среду из временно свободных ресурсов рабочих станций, суперкомпьютеров, серверов и персональных компьютеров при выполнении разнообразных задач. Такая консолидация позволяет повысить степень загрузки вычислительных компонентов корпоративной сети с 20 до 90%. При этом значительную роль играет не только пространственно-сетевое покрытие компьютерных ресурсов, но и новая эффективная методология программирования, учитывающая распределенность информационных ресурсов.

Другое содержание термина GRID связано с геоинформатикой, в которой введены GRID-структуры для представления данных о территории. В первую очередь это результаты дистанционного зондирования Земли в виде цифровых аэрокосмических снимков и цифровой картографический материал, полученный в результате проведения серии измерений по заданному набору пространственно распределенных точек. В настоящее время это направление развивается как международная программа GRID (GLOBE Project) [2] на основе картографической базы данных GTOPO30 (шаг сетки 30 с), соответствующей топокарте масштаба 1:1 000 000 и спутниковым данным типа получаемых с помощью спутников серии NOAA с разрешением 1 км. Кроме того, аналогичным путем развивается Глобальная геохимическая база данных (GGD) на основе Глобальной геохимической опорной сети (GGRN) с шагом сетки порядка 166×166 км (A Global..., 1996) [3]. Развиваются также методы организации интеллектуальных данных, прототипом которых является пирамидальная система организации растровой информации [4]. Такой организации данных не противоречит использование векторных карт ГИС как упрощенных модельных структур, что позволяет говорить о некотором инвариантном представлении информации в пространственно распределенных базах данных.

Все это позволяет решать прикладные задачи, большинство которых связано с анализом пространственно-временных связей на Земле, а также минимизировать затраты на создание высокопроизводительной вычислительной среды путем объединения программных средств в рамках корпоративной сети, реализующих заданную совокупность как альтернативных, так и взаимно дополняющих методов решения определенных задач. Также GRID-технология (за счет интеллектуализации конечного программного продукта) позволяет решить фундаментальную проблему обеспечения доступности больших программных систем "рядовому" пользователю (в "обычном" состоянии эффективное использование которых возможно только автором или при его тесном участии как эксперта-консультанта) и консультативного решения задачи коллективом экспертов на стыке дисциплин.

Для решения названной проблемы предполагается применять уже созданные коллективом авторов средства, методы и технологии:

1. Средства автоматизации процессов накопления, отладки, использования, модификации, комплексирования и визуализации пакетных знаний, а также автоматизации постановок исследовательских задач, процессов планирования, конкретизации и исполнения схем решения задач; алгоритмы автоматического синтеза схем решения задач, когда в ка-

честве модели проблемной области используются большие разреженные системы булевых уравнений; базовые инструментальные средства организации распределенных вычислений в сети Интернет.

2. Язык и исчисления позитивно-образованных формул (ПО-формул); стратегии для поиска логического вывода в этих исчислениях на базе программного комплекса КВАНТ/2 с целью интеллектуализации информационно-управляющих систем; логико-эвристические методы сетевого планирования, имеющие практическое применение в задачах составления расписаний.

3. Язык описания форматов двоичных данных FlexT, не имеющий полноценных аналогов в мире и спецификации, позволяющей описывать структуры данных, генерируемых современными программными системами (с различной степенью детализации описано порядка 50 известных форматов, в том числе графические и растровые данные, например форматы ГИС, форматы исполняемых файлов, программных библиотек и т. п.). В настоящее время язык FlexT позволяет формулировать спецификации интерпретаций, при помощи заданных описаний производить преобразование содержащейся в бинарных данных информации к пригодному для восприятия человеком виду. Использование спецификаций форматов бинарных данных позволяет сделать эти данные прозрачными и обеспечить надежный информационный обмен между различными программными системами без необходимости преобразования информации в неэффективные, но пригодные для восприятия человеком представления, такие как XML. Другим важным преимуществом предлагаемого подхода является возможность описывать уже существующие форматы данных, не требуя конвертации информации в какой-либо “универсальный” обменный формат.

4. Информационно-справочные системы, предоставляющие доступ к базам данных (БД) в технологии Интернет/Интранет и при необходимости использующие в своем составе ГИС. Накопленный опыт их применения позволяет разработать достаточно универсальные технологии реализации таких систем с использованием метаинформации о структуре БД.

5. Средства преобразования картографической информации и публикации электронных карт в Интернет/Интранет. Накопленный опыт создает возможности разработки технологии хранилища пространственных данных. Разработана система анализа структуры векторных и растровых картографических данных, в качестве практического приложения которой, например, решена задача автоматического выделения дорожной сети города на векторной карте. В этой системе используется встроенный интерпретатор Пролога в качестве языка пользовательских сценариев, позволяющего выполнять настройку системы на особенности конкретных распознаваемых объектов. В направлении развития этой системы предполагается создать ряд алгоритмов обработки векторной картографической информации.

Для развития GRID-технологий и создания на этой базе программных средств для решения задач в области геоинформационных исследований требуется дополнительная разработка методов интеллектуализации, средств спецификации и технологий их использования для различных видов информационно-вычислительных ресурсов (баз данных, ГИС, библиотек программных модулей базы знаний и метаданных, мультиагентных технологий, планирования действий агентов и т. д.). Более детально планируется:

1. Разработать концепцию модели и методы децентрализованного планирования и управления распределенными ресурсами и связями в сети Интернет на основе системы взаимодействующих интеллектуальных агентов, многомерных булевых моделей и эффективных логико-эвристических алгоритмов их решения. Разработать имитационную модель

распределенного вычислителя с целью оптимизации значений его параметров избыточности по заданному критерию отказоустойчивости.

2. Разработать (на текущем этапе развития языка FlexT) высокоуровневые средства описания распределения памяти в бинарных данных, что позволит существенно уточнить спецификации многих форматов файлов, а также разработать модуль автоматического синтеза подпрограмм чтения информации из файлов для различных языков программирования на основе спецификаций форматов файлов.

3. Разработать и формализовать универсальную схему доступа к данным БД с использованием Интернет/Интранет-технологий. Разработать средства автоматизации описания структуры БД для реализации пользовательского интерфейса формирования произвольных запросов и транслятора пользовательских запросов на универсальный язык запросов.

4. Разработать технологию интеграции разнородных пространственных данных на основе хранилища пространственных данных, алгоритмы сравнения и автоматической генерализации векторных электронных карт, систему автоматической векторизации растровых

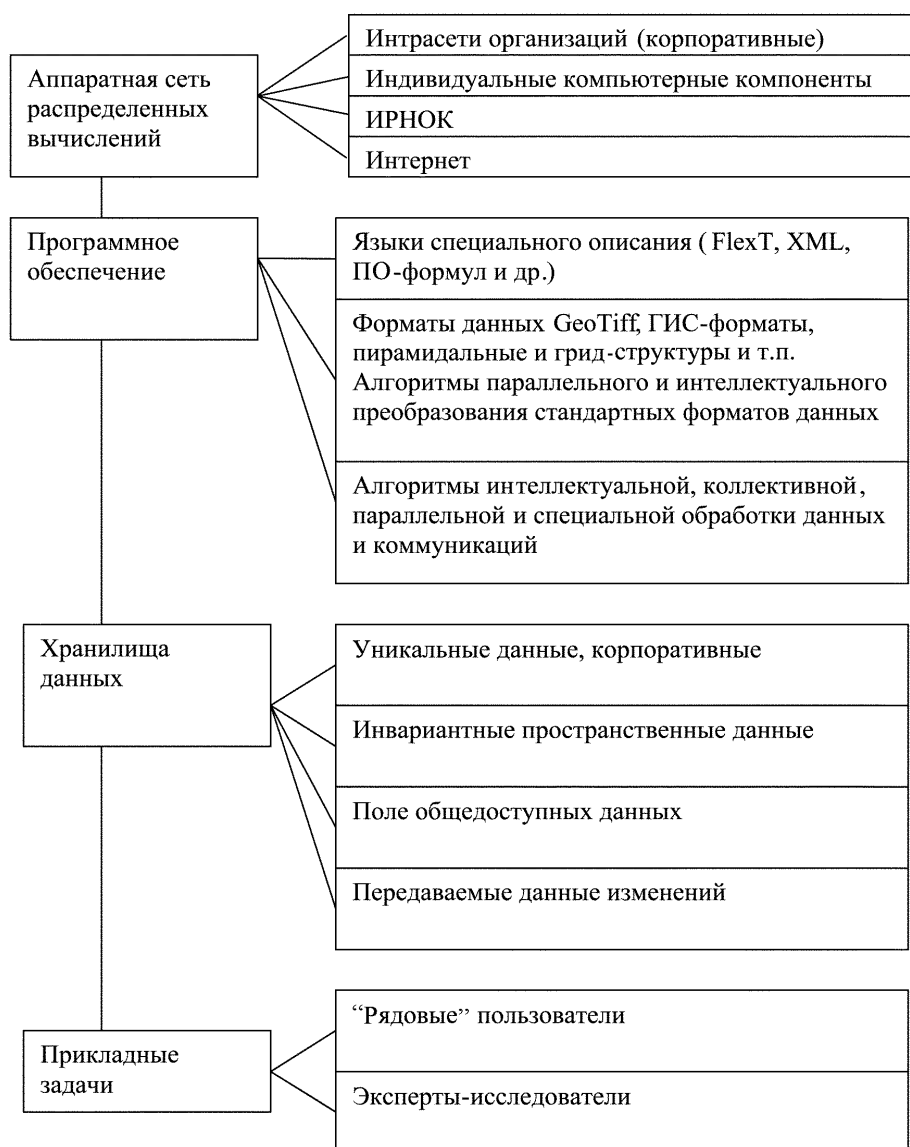


Рис. 1. Схема компонентов распределенного решения задач.

карт, использующих в своей работе логические методы описания характеристик классов объектов карты. Поскольку пространственные данные ГИС с учетом использования данных дистанционного зондирования имеют колоссальные объемы, то необходимо создание единого картографического пространства и баз таких данных на основных узлах с возможностью передачи оперативных изменений. То есть необходимо разработать методы передачи и интеграции изменений информации на уровне инвариантных форматов.

5. Для построения первого уровня grid системы внедрить многопроцессорную ЭВМ и вычислительные кластеры в интегрированную информационно-вычислительную сеть Иркутского регионального научно-образовательного комплекса (ИРНОК).

Таким образом, схема компонентов, включающих аппаратные, программные и информационные подсистемы внедрения GRID-технологий в задачах, возникающих при геоинформационном исследовании, представлена на рис. 1.

Традиционный путь передачи данных — использование конверторов и универсальных обменных форматов. В этом случае не всегда удастся сохранить все особенности, в первую очередь семантические, исходной информации. Эта проблема решается с помощью специального языка FlexT, позволяющего на полученных описаниях данных обойтись без обменного формата. С другой стороны, эта проблема решается накоплением данных и знаний в хранилище инвариантной структуры, а также представлением информации в виде “интеллектуальных” данных, хранящих в себе информацию о преобразованиях и подмножества данных для сокращения объема передаваемой и визуализируемой информации (рис. 2). Целесообразно также в распределенной сети на базовых узлах иметь копии общих данных большого объема. Передача этой базовой информации не требует оперативности. Такой информацией может быть набор тематических и топографических цифровых карт или покрытие какой-либо территории серией космоснимков. После обработки специальной информации компаньоном (оператором этого узла) в общее хранилище или другому

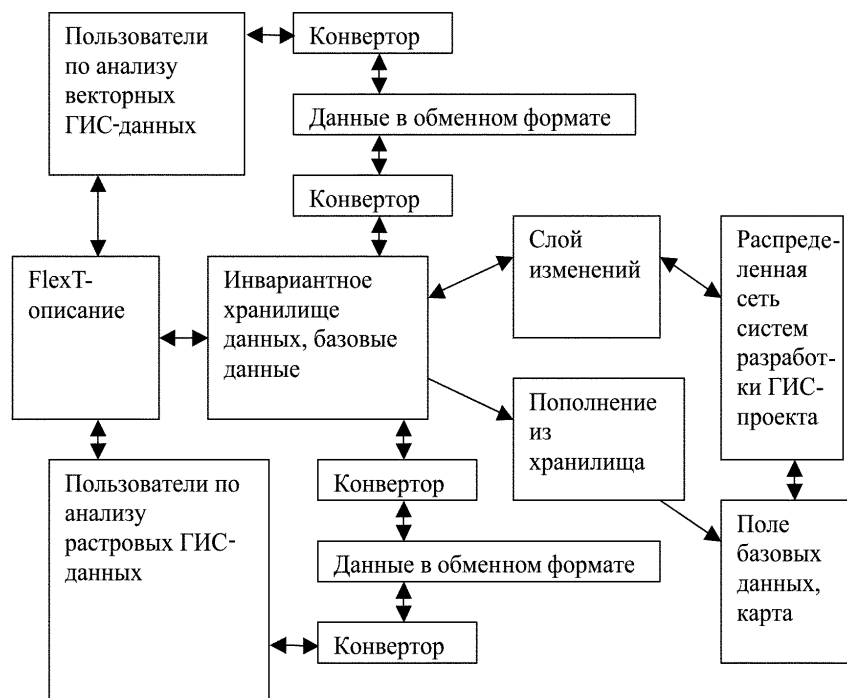


Рис. 2. Схема вариантов передачи и преобразования данных.

компаньону ГИС-проекта, имеющего базовую информацию, передается графический слой изменений для дальнейшей специальной обработки или фрагмент для параллельного процесса.

Другим преимуществом инвариантных GRID-структур является возможность распараллеливания обработки. Например, в случае попиксельного преобразования растрового изображения одна и та же операция выполняется над каждым элементом независимо. Другим примером фрагментации и параллельной обработки является сегментация изображения или отслеживания контуров. Один из эффективных методов распознавания образов — метод комитетов или голосования — также может быть распараллелен. В этом случае исходное изображение на каждом отдельном процессоре параллельной структуры обрабатывается своим методом распознавания, а на процессор принятия решений передается результат анализа, где методом голосования выбирается окончательное решение.

Разработка алгоритмов распараллеливания процессов обработки данных на базе ГИС-технологии целесообразна в следующих случаях.

1. Иерархический синтез и анализ данных. Например, при мультиспектральном анализе геоизображений и обработке множества данных, характеризующих исследуемую область территории, разбиение множества данных на подгруппы и попарная их обработка внутри подгруппы позволяют построить эффективные процедуры для распараллеливания процессов синтеза и анализа (рис. 3, а).



Рис. 3. Примеры параллельных процессов при ГИС-анализе данных.

2. Формирование тематических карт, например ландшафтно-оценочной карты территории. По заданным критериям (рис. 3, б), таким как высота над уровнем моря, экспозиция и крутизна рельефа, плотность распределения вида, яркость элемента изображения (космоснимка), попадание определенного векторного объекта цифровой топокарты в этом месте и т. п., формируется совокупность параметров и производится распараллеливание процедуры определения (классифицирования, сегментации и т. п.) элемента (выдела) из множества исходных материалов.

3. Выбор вариантов анализа по ГИС-данным, например сети дорог. Варианты трассы анализируются одновременно, результаты параллельного анализа сравниваются.

4. Обновление карт (атрибутивных данных — изменение свойств и графических элементов — изменение местоположения). Одновременное обновление синтаксических и семантических атрибутов данных, например изменение координат объекта при смещении (рис. 3, в) или пересчет координат в связи с изменением проекции карты.

5. Визуализация данных (сборка изображения из векторных сегментов или растровых элементов). Представление изображения на экране монитора осуществляется с помощью синтеза растрового изображения (экран — растр). При масштабировании параллельно можно производить генерацию нескольких одинаковых пикселей (увеличение) или формирование одного пикселя из множества (уменьшение) (рис. 3, г). Также параллельно можно осуществлять осреднение элементов по некоторому фрагменту, при этом процесс сокращается на число фрагментов в изображении. Эффективно применять процедуру распараллеливания при создании пирамидальных слоев растровых данных.

6. GRID-анализ — обработка данных с учетом заданной сети: оценивание расстояния от объекта до узлов сетки с целью выявления ранга объекта, общности объектов, пространственной смежности и т. п.

Предполагаемые методы и средства планируется использовать при проведении крупномасштабных мультидисциплинарных исследований на основе уникальных (оригинальных) программно-информационных ресурсов институтов ИНЦ СО РАН и распределенного вычислителя для решения геопространственных задач большой размерности с целью разработки и реализации системы поддержки принятия управленческих решений.

Список литературы

- [1] Китов А.Д. Компьютерный анализ и синтез геоизображений. Новосибирск: Изд-во СО РАН, 2000. 220 с.
- [2] EUROPEAN Collaborative programme. Remote sensing in landscape ecological mapping / В. Delbaere, Н. Gulinck. Luxembourg: Office for Official Publ. of the Europ. Communities, 1995. Vol. IV. 164 p.
- [3] A GLOBAL Geochemical Database for Environmental and Resource Management. Recommendations for International Geochemical Mapping. Final Report of IGCP Project 259 / by ed. A.G. Darnley, A. Bjorklund, B. Bolviken et al. France: UNESCO Publ., 1996. 122 p.
- [4] КАРТОГРАФИЯ. Геоинформационные системы / Под ред. А.М. Берлянта, В.С. Тикунова. М.: Картгеоцентр-Геодезиздат, 1994. Вып. 4. 350 с.