

Использование кластерных технологий при решении задач защиты картографических данных

И.С. Вершинин, Р.Ф. Гибадуллин, С.В. Пыстогов

Казанский государственный технический университет им. А.Н. Туполева

Введение

Цифровые топографические основы отличает высокая себестоимость работ по их созданию. Например, поиск новых месторождений - это рискованные, высокозатратные в финансовом отношении проекты, со сроками реализации порядка 10-15 лет. Следовательно, возникает вопрос об авторских правах на цифровые карты. Некоторые сведения государственного земельного кадастра носят закрытый характер [1]. Содержание цифровых карт, которые используются в военных целях, также не является информацией общего пользования [2]. Картографические данные, несущие определенную коммерческую информацию, имеющую ценность для коммерческих организаций использующих ГИС, требуют ограничения доступа к ним.

Таким образом, выработка тактики и стратегии защиты данных геоинформационных систем (ГИС) является актуальной задачей и требует серьезного внимания.

Организация работ с защищенными картографическими базами данных (ЗКБД) неотъемлемо связана с решением таких задач, как шифрование, расшифрование и обработка запросов к ЗКБД. Выполнение процедур по решению данных задач целесообразно организовать на вычислительном кластере, так как их выполнение на одной ПЭВМ связано с чрезмерными временными затратами. Считается, что применение параллельных вычислительных систем при решении задач защиты картографических данных даст существенный рост в быстродействии, чем решение этих задач на обычной последовательной ЭВМ.

В тезисах подчеркивается актуальность применения кластерных технологий при решении задач защиты картографических данных; оценивается сложность поиска полного множества ключей шифра, используемого для шифрования картографических данных; рассматриваются принципы формирования распределенной защищенной картографической базы данных (ЗКБД); даются временные оценки работы процедур по управлению ЗКБД; рассматривается проект системы Security Map-Point Cluster.

Формирование ЗКБД

В качестве модели хранения картографических данных выбрана векторная модель. Векторные модели предпочтительны, так как требуют на порядки меньше памяти для хранения и меньше затрат времени на обработку и представление, чем растровые. Основой для задания пространственных характеристик объектов картографии служит прямоугольная система координат.

Шифрование данных геоинформатики происходит с помощью стохастического алгоритма шифрования, основанного на двумерно-ассоциативном механизме маскирования [3]. Для оценки сложности поиска полного множества ключей (наборов масок) шифра решается задача для размеров эталонов $m \times n = 18 \times 35$ следующим образом.

По АЛГОРИТМУ [4] случайным образом генерируется некоторый набор масок, который сравнивается со множеством ранее сгенерированных уникальных наборов масок (для первого ключа такое сравнение не проводится). Если вновь сгенерированный ключ не совпадает ни с одним ключом данного множества, он добавляется в это множество, иначе добавление не происходит. Далее происходит генерация следующего ключа, и процедура повторяется.

Приведем некоторые временные оценки. Время получения набора из 6 000 новых уникальных ключей при наличии ранее сгенерированного множества из 300 000 клю-

чей составляет примерно 10 часов. Эксперименты проводились на ЭВМ с процессором AMD Turion™ 64x2, тактовая частота 2 ГГц, ОЗУ 2 Гб. Очевидно, что с увеличением множества время поиска новых уникальных ключей будет только увеличиваться.

Возможность повышения быстродействия поиска заключается в распараллеливании процесса. Для этого необходимо проводить разбиение генерируемых ключей на подмножества по количеству сохраняемых по маскам битов эталонов, которое варьируется в интервале от 34 до 48. Каждый из 15 процессоров будет работать со своим подмножеством. Подмножества ключей (по количеству сохраняемых битов) заносятся в соответствующие файлы.

Каждый процессор осуществляет генерацию набора масок до тех пор, пока количество единиц в этом наборе не будет совпадать с назначенным данному процессору числом сохраняемых битов. Далее проводится проверка на совпадение данного ключа с ключами соответствующего подмножества. Если совпадения нет, ключ добавляется в данное подмножество. Далее процедура повторяется.

Для формирования ЗКБД пространственные характеристики объектов тематических слоев карты, содержащие исключительно точечные объекты, подвергаются процедурам позиционирования и кластеризации (разбиению на кластеры (фрагменты)) [4].

После применения к исходным картографическим данным указанных процедур, строится база данных со следующей схемой (рис. 1).

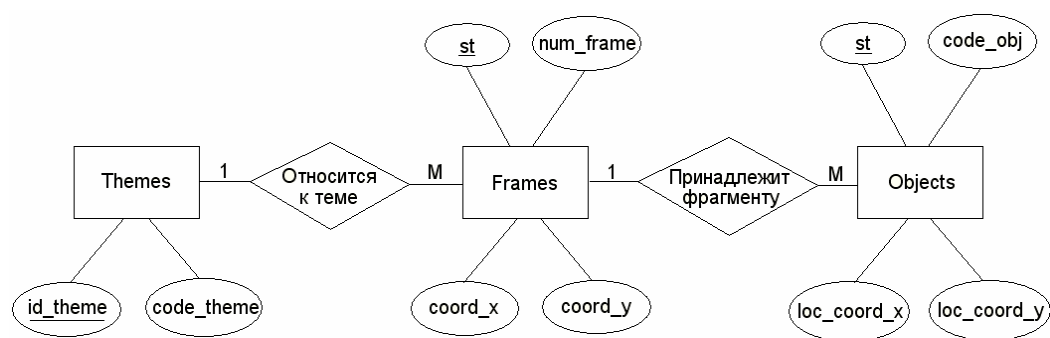


Рисунок 1.

Фрагменты, полученные после кластеризации, дополняются пустыми (реально не существующими) объектами, которые шифруются вместе с остальными. Код пустого объекта на определенном тематическом слое выбирается случайным образом из множества тех кодов объектов, которые не были задействованы на данном тематическом слое. Места локализации пустых объектов определяются случайно. После добавления пустых объектов должно соблюдаться требование: сумма пустых и непустых объектов в каждом фрагменте любого слоя карты равно числу N, где N не превышает максимально возможное число объектов во фрагменте;

Кластеризация и размещение на каждом тематическом слое ложных объектов является эффективным методом борьбы против криптографических атак, связанных с ассоциациями с картой местности.

В базе данных шифрованию подвергается содержимое столбцов, по которым хранятся коды тематических слоев, коды объектов, а также глобальные и локальные координаты фрагментов и объектов соответственно.

Шифрование ведется параллельно на узлах вычислительного кластера. Для распараллеливания программ по узлам кластера применяется библиотека передачи сообщений MPICH-1.

Вычислительными экспериментами установлено, что шифрация базы данных любой размерности на одной ПЭВМ происходит приблизительно в n раз дольше, чем на вычислительном кластере, состоящем из n таких ПЭВМ [6].

Для организации работ с защищенными картографическими данными на вычислительном кластере, ЗКБД равномерно распределяется по узлам вычислительного кластера [7].

Процедуры управления ЗКБД

Для организации работ с ЗКБД выбраны и реализованы следующие процедуры: *добавление, удаление и изменение* точечного объекта картографии; *выборка* точечных объектов картографии. Алгоритмы процедур *добавления, удаления и выборки* даны здесь [5]. Алгоритм процедуры *изменения* приводится ниже.

Алгоритм процедуры *изменения* объекта.

1. Определить фрагмент Q , в котором находится изменяемый объект q . Если такого фрагмента не существует, то процедуру изменения следует прервать.
2. Проверить наличие изменяемого объекта q во фрагменте Q . Если такого объекта не существует, то процедуру изменения следует прервать.
3. Определить фрагмент Q' , в который может переместиться изменяемый объект q . Если фрагмент Q' по новым координатам объекта q уже имеется, проверить на совпадение нового кода и координат изменяемого объекта q с кодом и координатами какого-либо объекта во фрагменте Q' . Если имеет место совпадение, то процедуру изменения следует прервать.
4. Если новые координаты объекта q не выходят за пределы фрагмента Q , изменить атрибуты объекта, завершить алгоритм.
5. Проверить, является ли объект q единственным непустым объектом во фрагменте Q . Если является, то удаляется весь фрагмент Q . Иначе объект q заменяется пустым объектом.
6. Если фрагмент Q' не существует, то его требуется создать, заполнить необходимым числом пустых объектов и непустым объектом q и завершить алгоритм.
7. Переместить объект q во фрагмент Q' .
8. Проверить, имеется ли во фрагменте Q' хотя бы один пустой объект. Если имеется, то удалить в Q' любой пустой объект. Иначе дополнить каждый фрагмент, кроме Q' , одним пустым объектом. Данный шаг необходим для соблюдения равного числа объектов в каждом фрагменте.

Для оценки времени выполнения процедур был взят участок карты 700×700 км² и по ней сгенерирована ЗКБД, содержащая один тематический слой с числом фрагментов 71, $N = 28$.

Работа процедур с тестовой ЗКБД проводилась на вычислительном кластере при различном числе узлов, объединенных сетью Gigabit Ethernet посредством коммутатора D-LINK DGS-1016D. Каждый из узлов имеет двухъядерный процессор Intel(R) Core(TM)2 CPU частотой 1,87 ГГц и оперативную память 3 Гб.

Результаты тестирования процедур представлены в таблице 4. Отметим, что по каждой процедуре проведено 50 испытаний.

При тестировании процедуры *добавления* рассмотрены 3 случая:

1. Добавление объекта сопровождается удалением пустого объекта;
2. Добавление объекта требует создание нового фрагмента;
3. Добавление объекта требует увеличение числа N на 1.

При тестировании процедуры *изменения* рассмотрены 4 случая:

1. Объект не выходит за пределы своего фрагмента в процессе изменения;
2. Объект перемещается во фрагмент с хотя бы одним пустым объектом;

3. Объект перемещается во фрагмент без пустых объектов. Изменение объекта требует увеличение числа N на 1;
4. Добавление объекта требует создание нового фрагмента.

Таблица 1.

Тип процедуры		Среднеарифметическое время выполнения процедуры, сек.			
		На одной машине	На различных конфигурациях кластера		
			4-х узловая	8-ми узловая	12-ти узловая
Добавление объекта	Случай 1	53,2	36,5	41,4	34,6
	Случай 2	1414,5	374,6	211	158,7
	Случай 3	3230,2	858,6	476,9	358,6
Изменение объекта	Случай 1	30,2	27,5	25,4	37,6
	Случай 2	123,5	90,6	43	54,7
	Случай 3	1027,2	224,6	164,6	147,9
	Случай 4	1545	298	184	147
Удаление объекта		56,5	32,1	37,6	34,3
Выборка объектов		3,6	2,4	1,7	1,4

По таблице видно, что выполнение процедур на вычислительном кластере происходит сравнительно быстрее, чем на одной машине. Что доказывает необходимость использования кластерных технологий при решении задач защиты картографических данных. Так как при выполнении процедур удаления, добавления для 1-го случая и изменения для 1-го и 2-го случаев число шифруемых кодов невелико, то масштабирование кластера не увеличивает скорость их работы. При выполнении процедур для всех оставшихся случаев кластер демонстрирует хорошую масштабируемость.

Проект системы Security Map-Point Cluster

Система Security Map-Point Cluster разрабатывается для обеспечения удаленной работы множества пользователей с ЗКБД. Система базируется на трезвенной системе вычислений (рис. 2).

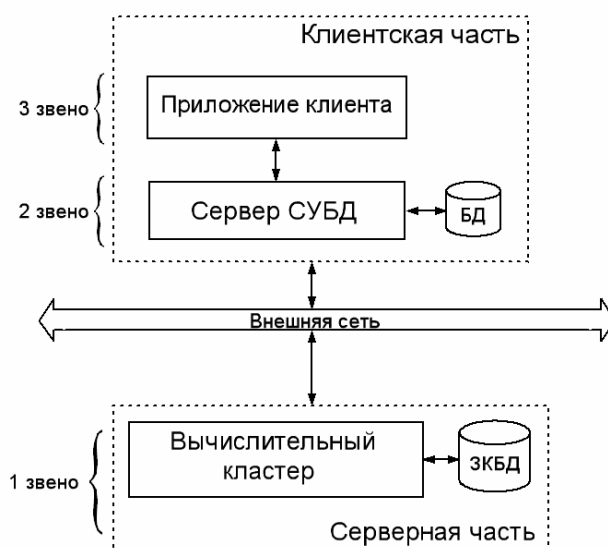


Рисунок 2.

В первом звене происходит обработка SQL-запроса селективного типа. Во втором звене происходит обработка SQL-запросов любого типа над результирующими отношениями, которые были сформированы первым звеном. Третье звено является SQL-клиентом для второго звена.

В системе предлагается:

- решать проблему анализа TCP/IP пакетов с помощью их шифрования;
- решать задачу защиты информации на SQL-сервере путем хранения картографических данных в зашифрованном виде.
- обеспечивать безопасность доступа пользователей к серверу за счет средств идентификации и аутентификации СУБД, предустановленной на Host-узле вычислительного кластера.
- организовать работы с защищенными картографическими базами данных без их предварительного расшифрования.

Цикл работы клиента с сервером состоит из следующих этапов:

1. Клиент посылает на сервер SQL-запрос селективного типа.
2. Сервер проводит синтаксический анализ, определяет соответствующие действия для обработки SQL-запроса и выполняет их.
3. По окончании обработки SQL-запроса сервером, клиент получает результирующие отношения и сохраняет их в своей локальной БД.
4. Клиент посылает SQL-запросы любого типа на локальный сервер СУБД и анализирует результаты обработки, проводя их визуализацию с помощью прикладной ГИС.

Для снижения вероятности нарушения целостности данных, хранимых на сервере, управление модулями модификации ЗКБД делается доступным только администратору системы.

Литература

1. Федеральный закон РФ "О государственном кадастре недвижимости".
2. Федеральный закон РФ "О навигационной деятельности".
3. Райхлин В.А., Вершинин И.С., Глебов Е.Е. К решению задачи маскирования стилизованных двоичных изображений //Вестник КГТУ им. А.Н. Туполева. 2001. №1. С. 42-47.
4. Райхлин В.А. Конструктивное моделирование систем – Казань: Изд-во ФЭН (Наука), 2005. С.237-272.
5. 8-я Международная Конференция. Высокопроизводительные параллельные вычисления на кластерных системах (НРС-2008). Казань, ноябрь 17-19, 2008. Труды конференции – Казань: Изд. КГТУ, 2008. С. 216-221.
6. И.С. Вершинин, Р.Ф. Гибадуллин, П.Е. Земцов. Параллельные алгоритмы защиты бинарных объектов картографии //Моделирование процессов /Под. ред. В.А. Райхлина. Труды Казанского научного семинара «Методы моделирования». Вып. 3. – Казань: Изд-во КГТУ, 2007. С.96-108.
7. Абрамов Е.В., Вершинин И.С., Гибадуллин Р.Ф., Шагеев Д.О. Практикум по параллельным вычислениям /Под ред. В.А. Райхлина – Казань: Изд-во Казан. гос. техн. ун-та, 2008, С.107-127.