

УДК 681.3

ЗАЩИТА ДАННЫХ КАРТОГРАФИИ НА КЛАСТЕРНОЙ ПЛАТФОРМЕ С ИСПОЛЬЗОВАНИЕМ АППАРАТА МАСКИРОВАНИЯ

В.А. Райхлин, И.С. Вершинин, Р.Ф. Гибадуллин

Рассматривается эффективный по критерию временных затрат и стойкости метод стегозащиты точечных объектов картографии. Метод основан на представлении десятичных кодов объектов и координат стилизованными бинарными матрицами, их маскировании и погружении в контейнеры псевдослучайных последовательностей. Приводятся результаты теоретических исследований и разработки специализированной СУБД защищенной картографии на кластерной основе.

Ключевые слова: маскирование, защита объектов картографии, вычислительный кластер.

ВВЕДЕНИЕ

Картографическая продукция отличается высокой себестоимостью работ по ее созданию и часто – конфиденциальностью сведений. То и другое обуславливает необходимость ее защиты, которая связывается с разработкой специальных методов формирования и анализа защищенных сцен. Известные универсальные СУБД со встроенными механизмами защиты в данном случае мало подходят. Организация приемлемой защиты БД картографии значительных объемов (десятки – сотни GB) «в реальном времени» связывается с построением специализированных СУБД кластерного типа.

В наиболее быстрых потоковых системах ключом защиты является псевдослучайная последовательность (ПСП) [1]. Во избежание присущих им трудностей, ПСП (гамму) целесообразно использовать как «не ключевой» носитель информации, в который внедряется цифровое сообщение. Для обеспечения нужной стойкости защиты сообщение должно занимать незначительную часть ПСП [2]. Возникает задача такого сжатия сообщения, чтобы санкционированный пользователь мог его распознать на фоне неизвестной гаммы.

Эту задачу можно решить, используя аппарат двумерного ассоциативного поиска [3,4], применяемый для распознавания замаскированных бинарных изображений. Защита данных с использованием маскирования рассматривается как частный случай трафаретного способа стеганографии, когда скрываемое сообщение внедряется по трафарету в ПСП – контейнер, не несущий информационной нагрузки. Трафарет, ассоциированный с содержимым окон, играет роль ключа.

Главными вопросами теории являются:

- 1) структуризация данных;
- 2) алгоритмизация процесса формирования ключа;
- 3) генерация требуемой гаммы.

СТРУКТУРИЗАЦИЯ КАРТОГРАФИЧЕСКИХ ДАННЫХ

Рассмотрим текст в алфавите почтовых индексов 1,2,3,4,5,6,7,8,9,0

1. Представим этот текст в виде таблицы, где в естественном порядке следования по тексту показываються символы и соответствующие им координаты.

	1	2	3	4	5	6	7	8	9	10	11 j
i	3	9	5	2	9	0	3	5	7	2	9

Тогда i-строка текста отобразится следующим фрагментом таблицы (табл.1).

Таблица 1. Фрагмент текстовой таблицы

Символ (код)	Координата	Символ (код)	Координата
3	(i, 1)	3	(i, 7)
9	(i, 2)	5	(i, 8)
5	(i, 3)	7	(i, 9)
2	(i, 4)	2	(i, 10)
9	(i, 5)	9	(i, 11)
0	(i, 6)		

2. Кластеризуем исходный текст в виде набора подтаблиц (карт) размерами $\epsilon \times \eta$ ($\epsilon < i_{\max}$, $\eta < j_{\max}$):

а) Случайным образом выбираем некоторую строку полной таблицы. Отмечаем эту строку. Позиционируем выделяемый кластер (определяем его глобальные координаты – см. далее рис.5). Сама же запись преобразуется к виду:

Символ (код)	(x, y)
--------------	--------

Здесь (x, y) – локальные координаты выделенного объекта в данном кластере.

б) Повторяем п.3 ($i_{\max} \times j_{\max}$) раз на множестве неотмеченных строк. При этом всякий раз устанавливаем принадлежность вновь выделенной строки к одному из ранее введенных кластеров и преобразуем координаты (из глобальных по тексту в локальные по кластеру). Если такового кластера не оказывается, инициуем новый кластер. В итоге исходная таблица заменяется совокупностью подтаблиц – кластеров с экстерриториальным порядком следования.

Пусть, например, размер кластера 3×3 (рис.1). По условию кластер 1 основан раньше кластера 2. Тогда кластер 1 содержит 9 элементов, а кластер 2 – всего лишь 5 элементов, ибо 4 элемента, которые принадлежат ему территориально, будут отнесены к кластеру 1 по критерию приоритета во времени его образования.

с) Во избежание детерминированности расположения в кластере его «родителя», содержимое каждого кластера перемешивается. Для выравнивания числа элементов во всех кластерах некоторые из них случайно дополняются так называемыми «пустыми» объектами, которые скрываются вместе с основными.

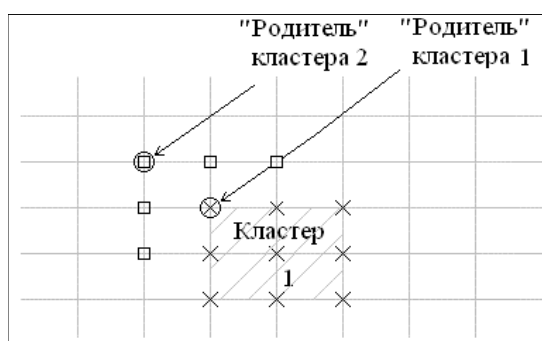


Рис.1. Иллюстрация к формированию кластеров

3. Стилизуем используемые почтовые символы как двоичные матрицы-эталон фиксированных размеров $m \times n$, $n=2m-1$ (рис.2–пример представления символа ☞ для $m=5$).

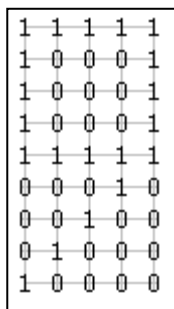


Рис.2. Эталон цифры ☞

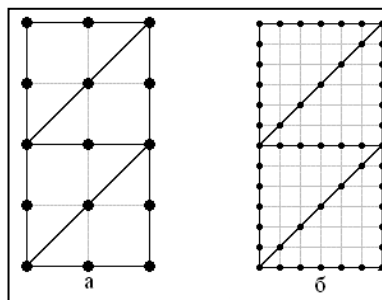


Рис.3. Примеры размещения существенных бит

4. Случайным образом генерируем маски для множества эталонов. Набор композиционных элементов (<эталон><маска>) – ключ защиты. Существенные биты символов располагаются по внешнему контуру и внутреннему «зигзагу» матриц (рис.3; $a - m=3$, $b - m=7$). Сканируя матрицу сначала – по контуру, затем – по «зигзагу», получаем ее информационный эквивалент в виде строки длиной $L = (9m-12)$ бит. Из них маскирование выделяет всего лишь несколько значимых бит, случайно распределенных по строке.

5. Цифровые коды объектов карты и их координат имеют разрядность k . По рассматриваемому далее условию защиты (выполнимому при введении «пустых» объектов) множество кодов для данного k является полным. Мощность этого множества $\Gamma = 10^k$. Число градаций координат $\Gamma_{x,y} = \Gamma$. На рис.4: $Y = X = A$ – максимальные значения координат ‘ y ’ и ‘ x ’; ε – погрешность определения координат объектов; 2ε – шаг локальной координатной сетки. Глобальная координатная единица $A/\Gamma_{x,y}$. Линейный размер кластера $C = (2\varepsilon)\Gamma_{x,y}$. Число градаций координат $\Gamma_{x,y} > [A/(2\varepsilon)]^{1/2}$.

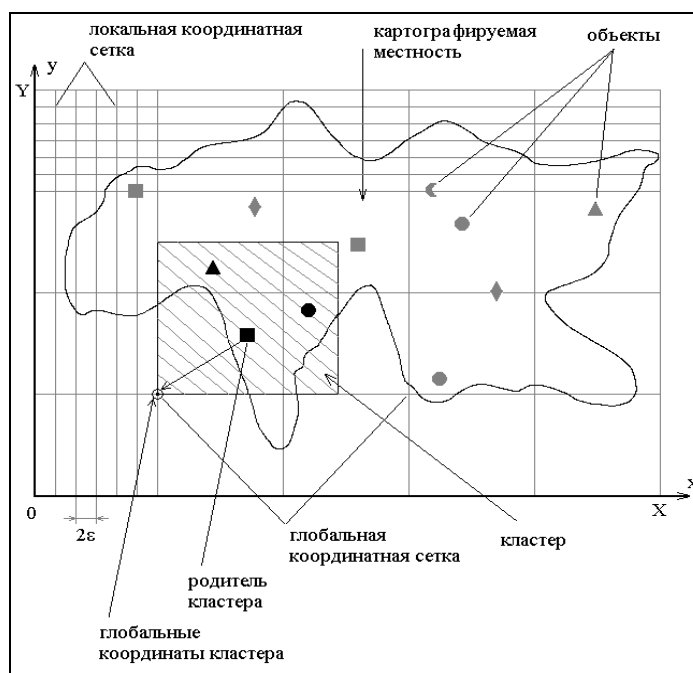


Рис.4. Иллюстрация градационных характеристик карты

Пример: $A = 0,5 \times 10^6$ м, $\varepsilon = 1$ м. Тогда $\Gamma_{x,y} > 500$. Ближайший выбор $\Gamma_{x,y} = 1000$, $k = 3$. $A/\Gamma_{x,y} = 500$ м. Размер кластера $C = 2000$ м.

ФОРМУЛИРОВКА ЗАДАЧИ СТЕГОЗАЩИТЫ

Рассматриваются двумерные картографические изображения с точечными объектами. Имена объектов и их координаты кодируются единообразно словом длины k в алфавите почтовых индексов.. Таблица кодов в терминах «объекты – координаты» кластеризуется. Итогом является некоторая совокупность закодированных тематических карт. Решается следующая задача защиты.

Имеем блок закодированных тематических карт. Заголовок блока: <ИМЯ_ТЕМЫ>. Имя карты: <ГЛОБАЛЬНЫЕ_КООРДИНАТЫ_КАДРА>. Кадр отображает участок определенных размеров на некоторой карте местности. Тематическая карта – отношение с кортежами: <ИМЯ_ОБЪЕКТА><ЛОКАЛЬНЫЕ_КООРДИНАТЫ_ОБЪЕКТА>.

Любая из t -цифр, $t \in \{1, 10\}$, представлена двоичными матрицами эталона и маски размерами $m \times n$ бит, m и n – число столбцов и строк. Эти матрицы формируют троичные эталоны X^t . Не маскируемые элементы исходного эталона определены единичными компонентами его инверсной матрицы масок

$$M^t = |m_{pq}^t|; p = 1 \dots m, q = 1 \dots n; m_{pq}^t = \begin{cases} 1, & x_{pq}^t \in \{0, 1\}; \\ 0, & x_{pq}^t \in \{-\}. \end{cases}$$

Набор [$\langle \text{эталон}^t \rangle \langle M^t \rangle \rightarrow X^t$] выполняет роль стежоключа. В матрице M^t значение $m_{pq}^t = 1$ позиционирует окно в трафарете. При погружении кодов по маскам M^t в свои контейнеры получаем множество тематических стежокарт, которое составляет защищенную картографическую базу данных (ЗКБД).

Требуется построить метод стегозащиты такой, чтобы уровень его стойкости был не ниже доказуемого.

ПОДХОД К РЕШЕНИЮ ЗАДАЧИ

Санкционированное распознавание. На каждое местоположение цифры в контейнере последовательно накладываются трафареты (маски) из набора, пока в окнах не проявятся элементы одного из эталонов.

УТВЕРЖДЕНИЕ 1. Если в любой паре матрицы X^{t1} и X^{t2} , $t1 \neq t2$, различаются хотя бы одним значащим элементом $x_{pq}^t \in \{0, 1\}$, то решение задачи санкционированного распознавания единственно.

При разработке базового алгоритма маскирования [5] (АЛГОРИТМа) было дополнительно принято: 1) генерируемый набор масок случаен; 2) число единиц матрицы M^t любого t -эталона близко к минимально возможному. Оно определяется условием дихотомизации любой пары троичных эталонов в сгенерированном наборе по одному существенному биту.

Один из вариантов маскирования, полученного по АЛГОРИТМу при $m \times n = 3 \times 5$, показан на рис.5.

позиций знания координат некоторых объектов и на передаваемую или хранимую гамму, выбор подходящей гаммы обеспечивает уровень стойкости стегозащиты не ниже доказуемого.

Пояснение 1. В нашем случае не зависящая от m оценка максимального числа единиц M^t : $(q1)_{\max} = 9$ [6]. Реально $1 \leq q1 \leq 8$ при матожидании $M_{q1} = 5$ и среднем объеме вкраплений $g = 3M_{q1} = 15$. Длина гаммы $L = 3(9m - 12)$. Поэтому при $k=3$, $m=60$ имеем: $L = 1584$, $g / L \approx 0,01$.

Гарантированно обнаружить факт включений можно, только если ПСП непрерывно генерируется на множестве контейнеров и объем вкраплений растет быстрее \sqrt{L} от суммарной длины ПСП [7]. У нас это условие не выполняется, ибо каждый гамма-контейнер уникален, и отношение $g / \sqrt{L} = \text{const}$.

Пояснение 2. АЛГОРИТМ стохастичен и существенно нелинеен. Его начальное состояние определяется случайной перестановкой эталонов. Используются операции поэлементного суммирования по mod2 и конъюнкции бинарных матриц (нелинейные), случайного выбора их единичных элементов. Тем не менее, справедлива

ТЕОРЕМА 2. Если гаммы некоторого подмножества контейнеров выявлены, то решение задачи определения истинного ключа с учетом дихотомальных особенностей алгоритма существует.

Мощность этого подмножества при $m = 40$ – более 30. Но каковы реальные возможности выявления такого числа контейнеров-носителей? Необходимо учитывать, что 1) формирование разных контейнеров происходит из различных (случайных) начальных состояний ГПСП и прекращается по заполнении контейнера; 2) ПСП-носитель искажена случайными вкраплениями значимых бит матриц M^t . Поэтому отыскать начальное состояние даже линейных ГПСП (например, «вихрь МЕРСЕННА» [8]) проблематично. И все же, чтобы исключить любую возможность позитивного исхода поиска, целесообразно использование нелинейных ГПСП (например, ГОСТ Р 34.10-2001 [9]).

Пояснение 3. Вероятность P повторения гаммы длиной $L = 1044$ на периоде ГПСП – $P = 2^{-1044}$ – ничтожно мала. Проведение операций сравнения над содержимым контейнера и побитно сдвигаемого «окна» длины L на полном периоде современных ГПСП практически не реализуемо. Столь же несостоятельна и «лобовая» атака полным перебором ключей. Верхняя оценка для числа ключей отвечает табл.3. [10].

Таблица 3. Оценки числа ключей для разных m

m	3	18	30	40
Число ключей	10^{13}	10^{23}	10^{25}	10^{27}

Если время интерактивного испытания одного ключа на неединичном подмножестве кластеров принять равным 1мкс. (реально, с учетом необходимой визуализации, оно много больше), то даже при $m=18$ (число ключей 10^{23}) полный перебор займет не менее $3 \cdot 10^9$ лет, и *условие доказуемой стойкости выполняется*.

РАЗРАБОТКА ЗАЩИЩЕННОЙ КАРТОГРАФИЧЕСКОЙ СУБД

Организация работ с защищенными картографическими базами данных (ЗКБД) связана с решением задач формирования ЗКБД (кластеризация, позиционирование, генерация ключей, поиск подходящей гаммы и др.), обработки запросов к ЗКБД, распознавания бинарных изображений, визуализации. Выполнение соответствующих процедур на одном ПК связано с чрезмерными временными затратами. Значительное ускорение процессов происходит при использовании вычислительных кластеров.

Результаты исследований приняты за основу построения системы **Security Map-Point Cluster** [11]. В настоящее время разработан исследовательский прототип этой системы. Развитый подход обобщен на случаи линейных и площадных объектов картографии [12].

На рис. 6 показана структура системы Security Map-Point Cluster. Система построена на базе СУБД MySQL с интегрированным в нее двумерно-ассоциативным методом защиты. Параллелизм в системе реализован с использованием интерфейса передачи сообщений MPI. Исследовательский прототип системы прошел успешные испытания на множестве репрезентативных запросов.

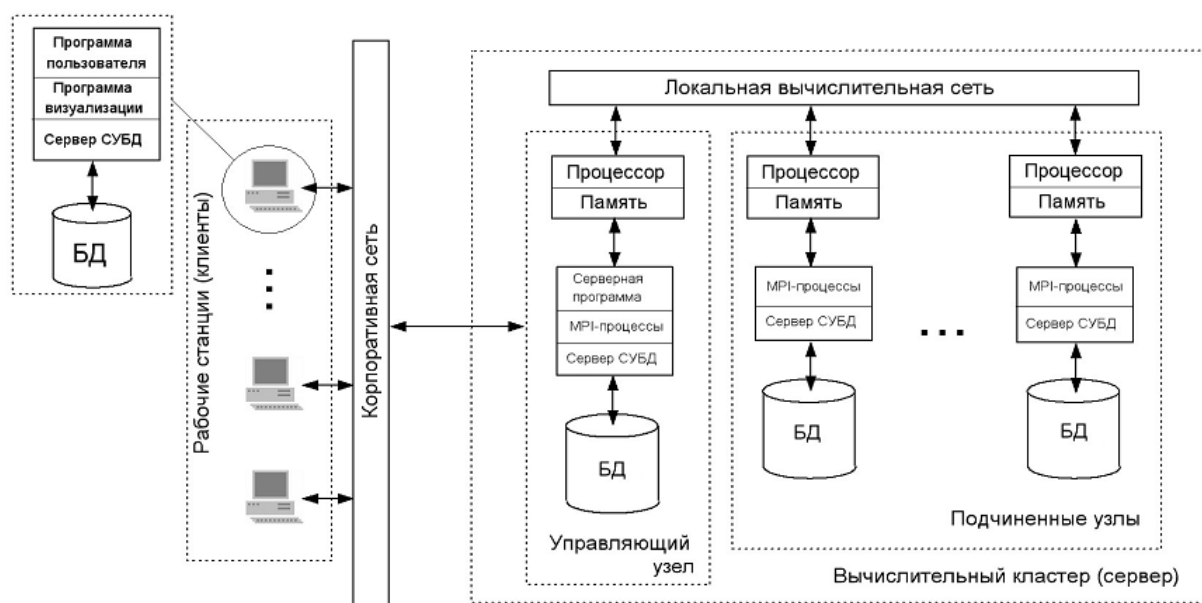


Рис. 6. Структура Security Map-Point Cluster.

Для тестирования комплекса программ, составляющих Security MapPoint Cluster, выбраны:

Аппаратная платформа. 12 вычислительных 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 GHz, оперативную память DDR2 3 GB, дисковый накопитель Western Digital 150 GB (с интерфейсом SATA).

Программное обеспечение: ОС семейства Microsoft Windows XP Professional, СУБД MySQL версии 5.1.45-win32, ГИС MapInfo Professional 10, интегрированная среда разработки MS Visual Studio 2008, библиотеки расширения языка C++: MPICH 1 (MPI), Boost 1.43, ГПСЧ «вихрь Мерсенна».

Тестовая карта: размером 300 x 300 км² участка местности республики Чувашии (рис.7), предоставленная ООО «Геодезическая компания «Зенит», г. Казань. Карта содержит один тематический слой и 1035 точечных объектов 4-х различных типов.

Градационные параметры: шаг глобальной координатной сетки – 300 м., шаг локальной координатной сетки – 0,3 м., число градаций координат в локальной и глобальной областях карты – 1000.



Рис.7. Тестовая карта

Результаты тестирования системы при $m=40$ показаны в табл.4.

Таблица 4. Результаты тестирования

<i>Процедура</i>	<i>Время выполнения (сек.)</i>
Формирование ЗКБД (10МВ)	41 (на одном вычислит. ядре – 912)
Добавление объекта	От 6 до 297
Модификация объекта	От 5 до 339
Удаление объекта	5
Визуализация селекции карты в целом	7
Выборка объекта	1,5

Результаты визуализации тестового картографического слоя на истинном и одном из ложных ключей представлены на рис.8.

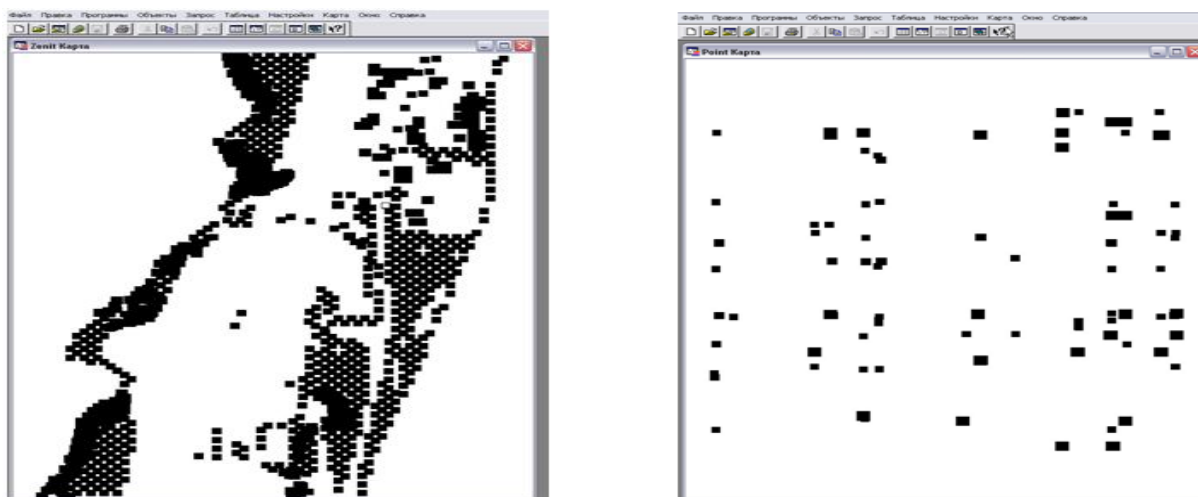


Рис. 8. Результаты истинного (слева) и ложного распознаваний.

Заметим, что время санкционированного распознавания в системе – вдвое меньше, чем при использовании шифра ГОСТ 28147-89.

Случай линейных и площадных объектов. На базе СУБД MySQL разработаны два основных модуля системы: 1) модуль формирования ЗКБД; 2) модуль обработки селективного запроса к ЗКБД. Проведены тестовые испытания второго модуля на двух конфигурациях системы, показанных на рис. 9 и 10. Характеристики узлов в обеих конфигурациях: CPU – Intel(R)Core (TM)2 1,87GHz; RAM – DDR2 3GB; SATA – West. Dig. 150GB.

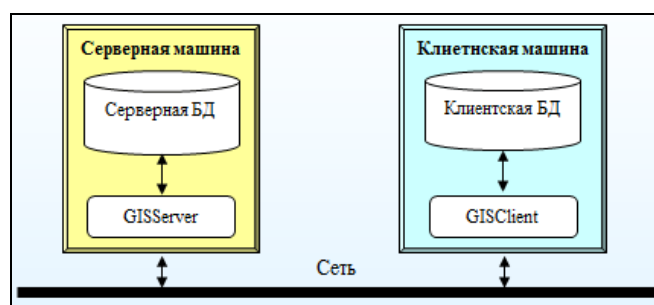


Рис. 9. ПК-сервер без «помощников».

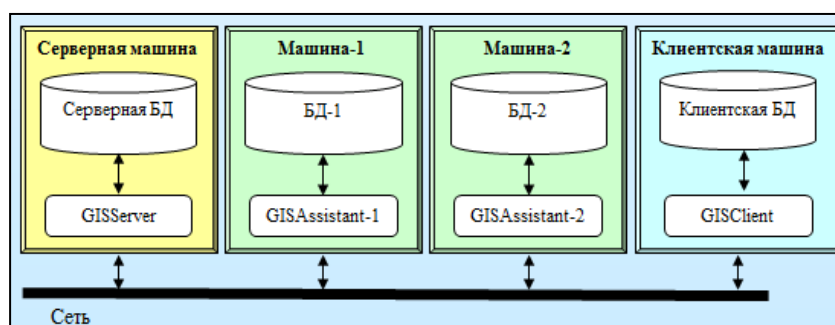


Рис. 10. ПК-сервер с двумя «помощниками».

Картографический слой для тестирования представлен четырьмя субъектами Российской Федерации: Кемеровской, Новосибирской областями, Алтайским краем и Республикой Алтай [13]. Он содержит 753 объектов при суммарном числе узлов по их контурам – 10485 (рис.11).

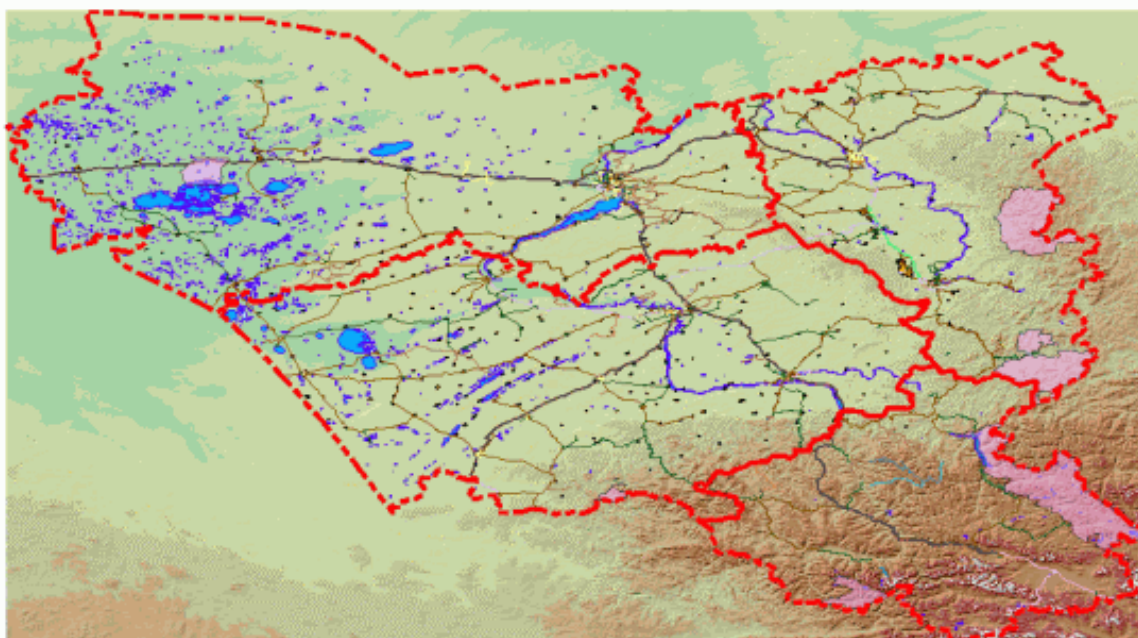


Рис.11. Картографический слой для тестирования в случае линейных и площадных объектов.

Среднее время селекции всего картографического слоя на первой конфигурации системы – 265 сек., на второй – 163 сек. (уменьшается на 38%). Это подтверждает необходимость использования кластерных технологий.

ЗАКЛЮЧЕНИЕ

В статье рассмотрено решение задачи построения метода стегозащиты данных картографии, основанного на использовании механизма двумерно-ассоциативного маскирования. Главным теоретическим результатом проведенных исследований является вывод о доказуемой стойкости предлагаемого метода. Доказуемая стойкость – достаточно «сильное» свойство, которое строго установлено только для шифров с применением гаммирования [1].

В прикладном плане, выполнена разработка исследовательского прототипа защищенной картографической СУБД Security Map-Point Cluster, предназначенного для сокрытия точечных объектов. Частично реализовано его расширение на случаи линейных и площадных объектов. Это является предпосылкой создания столь же эффективной, но более универсальной системы защиты, анонсируемой как Security Map Cluster.

СПИСОК ЛИТЕРАТУРЫ

1. Schneier B. Applied Cryptography, 2nd Edition. – John Wiley & Sons., 1996.
2. Агибалов Г.П. Sibecrypt'10. Обзор докладов //Прикладная дискретная математика. 2010. №4(10). С.109-124.
3. Райхлин В.А. Об использовании аппарата двумерного ассоциативного поиска в процессе распознавания //Проблемно-ориентированные средства повышения эффективности вычислительных систем. – Казань: КАИ им. А.Н. Туполева. 1991. С.38-54.
4. Райхлин В.А. Анализ производительности процессорных матриц при распознавании двоичных образов //Автоматрия. 1996. №5. С.97-103.
5. 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.
6. Райхлин В.А., Вершинин И.С. Элементы криптоанализа двумерного картографического шифра //Вестник КГТУ им. А.Н. Туполева. 2002. №4. С.48-54.
7. Ker D.A. A capacity result for batch steganography //IEEE Signal Processing Letters. 2007. V. 14(8). P. 525 -528.
8. What is Mersenne Twister (MT)? Интернет-адрес: <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/ewhat-is-mt.html>

9. ГОСТ Р 34.10-2001. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи. – М.: Государственный Стандарт Российской Федерации, 2001.
10. Вершинин И.С.. Верхняя оценка числа ключей двумерно-ассоциативной защиты объектов картографии //Методы моделирования. Труды Республиканского научного семинара АН РТ «Методы моделирования». Вып.4. – Казань: Изд-во «Фэн» («Наука»), 2010. С.96-100.
11. Вершинин И.С., Гибадуллин Р.Ф., Пыстогов С.В. Использование кластерных технологий при решении задач защиты картографических данных //Труды Межд. конф. НРС-2009. – Владимир: Изд-во ВГУ, 2009. С. 68-72.
12. Гибадуллин Р.Ф. Развитие единообразного формализма защиты точечных, линейных и площадных объектов картографии //Вестник КГТУ им. А.Н. Туполева. 2010. №2. С. 102–107.
13. <http://www.gis-lab.info/qa/geosample.html>

СВЕДЕНИЯ ОБ АВТОРАХ

1. Фамилия, имя, отчество: Райхлин Вадим Абрамович.

Место работы и должность: ФГБОУ ВПО Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, профессор кафедры компьютерных систем.

Год окончания учебного заведения и его полное название: 1962, Казанский авиационный институт им. А.Н. Туполева.

Ученая степень и звание: докт. физ.-мат. наук, профессор.

Количество печатных работ и монографий: 100.

Область научных интересов: конструктивное моделирование систем, высокопроизводительные вычисления.

Адрес электронной почты и контактный телефон: no-form@evm.kstu-kai.ru , +7(843)2310055.

2. Фамилия, имя, отчество: Вершинин Игорь Сергеевич.

Место работы и должность: ФГБОУ ВПО Казанский исследовательский национальный технический университет им. А.Н. Туполева-КАИ, доцент кафедры компьютерных систем.

Год окончания учебного заведения и его полное название: 2001, Казанский государственный технический университет им. А.Н. Туполева.

Ученая степень и звание: канд. техн. наук, доцент.

Количество печатных работ и монографий: 20.

Область научных интересов: защита объектов картографии, ГИС, высокопроизводительные вычисления.

Адрес электронной почты и контактный телефон: Vershinin_Igor@rambler.ru , +79172734641.

3. Фамилия, имя, отчество: Гибадуллин Руслан Фаршатович.

Место работы и должность: ФГБОУ ВПО Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, ассистент кафедры компьютерных систем.

Год окончания учебного заведения и его полное название: 2007, Казанский государственный технический университет им. А.Н. Туполева.

Ученая степень и звание: канд. техн. наук.

Количество печатных работ и монографий: 16.

Область научных интересов: защита объектов картографии, ГИС, высокопроизводительные вычисления.

Адрес электронной почты и контактный телефон: landwatersun@mail.ru , +79172262386.

In English:

PROTECTION OF CARTOGRAPHICAL OBJECTS IN CLUSTER PLATFORM USING MASKING APPARATUS

V.A. Raikhlin, I.S. Vershinin, R.F. Gibadullin

The method stegoprotection point cartographical objects is considered effective by criterion of temporal expenses and firmness. The method is based on representation of decimal codes of objects and coordinates by the stylized binary matrixes, their masking and immersing in containers of pseudorandom sequences. Results of theoretical researches and development of a specialized DBMS of the protected cartography on a cluster basis are resulted.

Keywords: masking, protection of cartographical objects, a computing cluster.

INFORMATION ABOUT THE AUTHORS

1. Surname, name, patronymic: Raikhlin Vadim Abramovich.

Place of operation and post: The Federal Government budgetary institution of higher education Kazan National Research Technical University named after A.N. Tupolev, professor to chair of computer systems.

Year of graduation and his full name: 1962, Kazan Aviation Institute named after A.N. Tupolev.

Academic degree and title: PhD, professor.

Number of scientific publications and monographs: 100.

Research interests: constructive simulation systems, high performance computing.

E-mail address and telephone number: no-form@evm.kstu-kai.ru , +7(843)2310055.

2. Surname, name, patronymic: Vershinin Igor Sergeevich.

Place of operation and post: The Federal Government budgetary institution of higher education Kazan National Research Technical University named after A.N. Tupolev, associate professor to chair of computer systems.

Year of graduation and his full name: 2001, Kazan State Technical University named after A.N. Tupolev.

Academic degree and title: PhD, associate professor.

Number of scientific publications and monographs: 20.

Research interests: protection of cartographical objects, GIS, high performance computing.

E-mail address and telephone number: Vershinin_Igor@rambler.ru , +79172734641.

3. Surname, name, patronymic: Gibadullin Ruslan Farshatovich.

Place of operation and post: The Federal Government budgetary institution of higher education Kazan National Research Technical University named after A.N. Tupolev, assistant to chair of computer systems.

Year of graduation and his full name: 2007, Kazan State Technical University named after A.N. Tupolev.

Academic degree and title: PhD.

Number of scientific publications and monographs: 16.

Research interests: protection of cartographical objects, GIS, high performance computing.

E-mail address and telephone number: landwatersun@mail.ru , + 79172262386.