На правах рукописи

ХНЫКИН Иван Геннадьевич

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ МИНИМИЗАЦИИ ФУНКЦИОНАЛОВ, АССОЦИИРОВАННЫХ С ЗАДАЧЕЙ ВЫПОЛНИМОСТЬ

Специальность 05.13.17 – Теоретические основы информатики

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Омск - 2009

Работа выполнена в государственном образовательном учреждении высшего профессионального образования «Омский государственный университет им. Ф.М. Достоевского» на кафедре безопасности информационных систем.

Научный руководитель — доктор технических наук, профессор
Файзуллин Рашит Тагирович

Официальные оппоненты: доктор физико-математических наук
Чернов Владимир Михайлович
кандидат физико-математических наук, доцент
Цветов Виктор Петрович

Ведущая организация: Институт математики и механики
Уральского отделения Российской академии наук

Защита состоится 23 декабря 2009 г. в 12 часов на заседании диссертационного совета Д 212.215.07, созданном при государственном образовательном учреждении высшего профессионального образования «Самарский государственный аэрокосмический университет имени академика С.П. Королева», по адресу: 443086, г. Самара, Московское шоссе, 34.

С диссертацией можно ознакомиться в библиотеке Государственного образовательного учреждения высшего профессионального образования «Самарский государственный аэрокосмический университет имени академика С.П. Королева».

Автореферат разослан 20 ноября 2009 г.

Ученый секретарь диссертационного совета
доктор технических наук, профессор Белоконов И.В.

Общая характеристика работы

Целью диссертационной работы является разработка информационной технологии решения задачи ВЫПОЛНИМОСТЬ посредством итерационных алгоритмов минимизации функционалов, ассоциированных с непрерывной моделью исследуемой дискретной задачи.

Актуальность работы. Задача ВЫПОЛНИМОСТЬ — одна из наиболее важных задач информатики в целом. На практике к решению задачи ВЫПОЛНИМОСТЬ сводятся многие проблемы, возникающие при синтезе систем искусственного интеллекта, при проектировании компьютерных систем, в задачах роботостроения, криптографии и других. Сведению задач из различных областей прикладной математики к задаче ВЫПОЛНИМОСТЬ посвящены работы Ю.И.Журавлева, А.А.Семенова Е.В.Буранова, Ю.Г.Сметанина, А.А.Ушакова, F. Massacci, L.Marraro и других. Одна из причин этого – потенциальная возможность решать задачи, возникающие в самых различных областях, единым алгоритмом решения задачи ВЫПОЛНИМОСТЬ. Следует отметить, что к решению указанной задачи сводятся многие проблемы дискретной математики, для которых доказано отсутствие алгоритмов полиномиальной сложности, то есть, их «труднорешаемость».

Однако, доказанный факт этой труднорешаемости, справедливый для всего (бесконечного) множества параметров, характеризующих ту или иную конкретную задачу, доказанный часто построением единственного экстремального примера, не снимает как саму проблему построения эффективных алгоритмов решения задачи в ограниченном, «пользовательском» диапазоне параметров, так и безусловную практическую значимость проблемы синтеза таких эффективных алгоритмов. В частности, возможность сведения задач факторизации больших целых чисел, дискретного логарифмирования, дискретного логарифмирования на эллиптической кривой к задаче ВЫПОЛНИМОСТЬ делает разработку моделей и алгоритмов её решения особенно актуальной, так как перечисленные задачи являются теоретической основой для разработки систем криптографической обработки данных, повышения надежности современных систем телекоммуникаций, систем финансовых взаиморасчетов и других.

Как показал опыт, применение переборных алгоритмов решения задачи ВЫПОЛНИМОСТЬ сталкивается с принципиальными трудностями: задачи больших размерностей оказываются недоступными для переборных алгоритмов. Естественно, возникает идея перехода к непрерывным моделям, когда поиск выполняющего набора для соответствующей конъюнктивной нормальной формы (КНФ) осуществляется как поиск экстремума, ассоциированного с КНФ непрерывного функционала. Впервые эта идея была реализована в работах С.Ю. Маслова, В.Я. Крейновича и получила дальнейшее развитие в работах R. Sosic, J. Gu, A. Torn, A. Zilinskas. Следует отметить, что, имеется принципиальное отличие непрерывных методов от переборных алгоритмов

локального поиска — сдвиг по антиградиенту происходит по всем переменным сразу, что дает определенный вычислительный выигрыш. Кроме того, для многих задач априори известно, что глобальный минимум функционала единственен и в случае, когда локальных минимумов и других особых точек нет, минимизация происходит эффективно.

Тем не менее, известные автору непрерывные методы не удовлетворяют современным потребностям, возникающим при решении задач ВЫПОЛНИМОСТЬ. Одни из них достаточно эффективно находят решение только для задач небольших размерностей, другие показывают хорошие результаты только на задачах для КНФ специальной структуры. Наконец, указанные методы показывают низкую эффективность при решении задач ВЫПОЛНИМОСТЬ, ассоциированных с задачами факторизации больших целых чисел, дискретного логарифмирования, дискретного логарифмирования на эллиптической кривой.

В диссертационном исследовании за основу взят непрерывный метод минимизации функционалов, т.н. метод последовательных приближений с «инерцией», описанный в работах Р.Т. Файзуллина. В оригинальном виде указанный метод оказывается недостаточно эффективным в применении к функционалам, полученным при различных способах моделирования КНФ. Предлагается улучшить характеристики метода последовательных приближений с «инерцией» путем его интеграции с оригинальными и известными подходами, способствующими увеличению эффективности.

Для последовательной реализации указанной идеи необходимо разработать оригинальные и адаптировать известные подходы (как непрерывные, так и дискретные), используемые в методах решения задачи ВЫПОЛНИМОСТЬ. При этом следует максимально обобщить данные подходы для возможности их применения не только в методе последовательных приближений с «инерцией». Кроме того, эффективность данных подходов должна проявляться в применении к КНФ с различной структурой. Хотя предпочтение все же отдается задачам ВЫПОЛНИМОСТЬ, ассоциированным с задачами факторизации больших целых чисел, дискретного логарифмирования, дискретного логарифмирования на эллиптической кривой. Вышеуказанные положения определяют структуру диссертационной работы и содержание отдельных глав.

Научная новизна работы. В диссертационной работе получены следующие научные результаты:

Разработана стратегия применения правил резолюции для преобразования КНФ и исследована ее эффективность.

Разработаны и обоснованы подходы, улучшающие сходимость градиентных методов поиска решения задачи ВЫПОЛНИМОСТЬ для КНФ различной структуры.

Разработан способ построения множества булевых векторов (битовой записи чисел), в среднем на 68% совпадающих с решением задачи факторизации целых чисел больших размерностей (до 3072 бит включительно).

Разработана система дополнительных тестов, позволяющая с высокой долей вероятности определять конкретные биты сомножителей в задаче факторизации размерности 512 бит.

На защиту выносятся следующие результаты:

Стратегия применения правил резолюции для преобразования КНФ различной структуры как препроцессор для методов поиска решения задачи ВЫПОЛНИМОСТЬ.

Подходы, улучшающие сходимость градиентных методов поиска решения задачи ВЫПОЛНИМОСТЬ для КНФ различной структуры.

Гибридный метод последовательных приближений с «инерцией» как способ нахождения областей, близких к решению задачи факторизации больших целых чисел.

Апробация работы. Результаты работы опубликованы, в том числе и в рецензируемых журналах и прошли апробацию на научных конференциях и семинарах: Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям (Красноярск, 2006), 10-й Московской международной телекоммуникационной конференции студентов и молодых ученых (2007), Региональной молодежной конференции «Проблемы теоретической и прикладной математики» (Екатеринбург, 2007), Межрегиональной конференции Современные математические методы и информационные технологии (Тюмень, 2007), «Параллельные вычислительные технологии» (Нижний Новгород, 2009).

Публикации. Основные результаты диссертации опубликованы в 10 работах, из них 4 в рецензируемых изданиях и журналах, рекомендованных ВАК.

Личный вклад. Автором разработаны оригинальные и адаптированы известные подходы, увеличивающие эффективность градиентных методов, в том числе стратегия применения правил резолюции, сдвиг по антиградиенту, метод смены траектории и др. Исследована их эффективность. Предложена система дополнительных тестов, позволяющая с высокой долей вероятности определять конкретные биты сомножителей в задаче факторизации целых чисел размерности 512 бит. Указанные подходы интегрированы с методом последовательных приближений с «инерцией» и исследована эффективность полученного гибридного метода в применении к поиску решения задач ВЫПОЛНИМОСТЬ для КНФ различной структуры, в особенности для КНФ, ассоциированных с задачами факторизации больших целых чисел, дискретного логарифмирования, дискретного логарифмирования на эллиптической кривой. Гибридный метод адаптирован для поиска областей (множества двоичных чисел), близких к решению задач факторизации больших размерностей (до 3072 бит включительно).

В диссертационную работу включены только результаты, полученные лично соискателем. В совместных публикациях вклад соискателя заключается в разработке методик улучшения сходимости, формулировке и доказательстве утверждений и в конфликт с соавторами не вступает.

Структура диссертации. Диссертационная работа состоит из введения, трех глав, заключения, библиографии.

Краткое содержание работы

Во введении обоснована актуальность темы диссертационной работы, сформулирована ее цель и задачи, дан краткий обзор научных работ по рассматриваемым вопросам, показана научная новизна работы и приводятся основные положения, выносимые на защиту.

В **первой главе** представлены необходимые сведения из булевой алгебры, приводится постановка задачи. Описаны различные математические модели задачи ВЫПОЛНИМОСТЬ. Рассматриваются наиболее популярные алгоритмы и общие схемы методов поиска решающего набора задачи ВЫПОЛНИМОСТЬ для КНФ. Основное внимание уделяется вопросу сведения задачи ВЫПОЛНИМОСТЬ к задаче непрерывной минимизации и применения градиентных методов поиска точек экстремума. Глава не содержит новых результатов и включена в работу для замкнутости изложения и удобства ссылок.

Задача ВЫПОЛНИМОСТЬ заключается в том, чтобы определить, выполнима ли данная формула в КНФ и найти набор(ы) значений булевых переменных, при которых пропозициональная формула выполнима.

Пусть дана КНФ, на множестве булевых переменных $y = \{y_1, \dots y_N\} \in B^N\{0,1\}$:

$$L(y) = \bigwedge_{i=1}^{M} G_i(y), \text{ где}$$

G_i – дизъюнкты вида $\bigvee_{j \in \{1..N\}} q_{i,j}(y)$

$$q_{i,j}(y) = \begin{cases} y_j, \text{если } y_j \text{ входит в } G_i \\ \bar{y}_j, \text{если } \bar{y}_j \text{ входит в } G_i \end{cases}, \qquad (1)$$

N - число переменных, M - число дизъюнктов

В работах Колоколова А.А., Адельшина А.В., Ягофаровой Д.И. предлагается использовать модель линейного целочисленного программирования (ЦЛП) для возможности применения методов оптимизации при поиске решения задачи ВЫПОЛНИМОСТЬ:

$$\min_{x \in R^N[0,1]} F(x) = \min_{x \in R^N[0,1]} x_1$$

$$\sum_{G_i^+} x_i - \sum_{G_i^-} x_i \geq 1 - \left| G_i^- \right|$$

$$x_i \in Z[0, 1]$$

Здесь $\left|G_i^-\right|$ – число литералов, входящих в *i*-й дизъюнкт с отрицанием.

В работах Крейновича В.Я. предложен ряд непрерывных моделей, сводящих задачу ВЫПОЛНИМОСТЬ к задаче безусловной глобальной минимизации функционала, например:

$$\min_{x\in R^N[0,1]} F(x) = \min_{x\in R^N[0,1]} T(L(y))$$

1. $T(0) = 0$
   $T(y_i) = x_i$
   $T(y_i \vee y_j) = \max(x_i, x_j),$
2. $T(1) = \infty$
   $T(\bar{y}_i) = 1/x_i$
   $T(y_i \wedge y_j) = (1/x_i + 1/x_j)^{-1};$

$y_i \in B\{0,1\}, \quad x_i \in R[0,\infty), \quad i = 1\ldots N$

Здесь и далее по тексту $T(y)$ – литерная функция, определяющая соответствие между булевым пространством и множеством, в котором предполагается осуществлять поиск решения.

В работах Опарина Г.А., Новопашина А.П. предлагается еще ряд вариантов сведения задачи ВЫПОЛНИМОСТЬ к поиску точек глобальных минимумов специальным образом построенного функционала, например:

$$\min_{x\in R^N[0,1]} F(x) = \min_{x\in R^N[0,1]} T(L(y))$$

1. $T(0) = 0$
   $T(y_i) = x_i$
   $T(y_i \vee y_j) = x_i + x_j,$
2. $T(1) = 1$
   $T(\bar{y}_i) = 1 - x_i$
   $T(y_i \wedge y_j) = \min(x_i, x_j).$

$y_i \in B\{0,1\}, \quad x_i \in R[0,\infty), \quad i = 1\ldots N$

В обзоре J.Gu представлены как дискретные, так и непрерывные модели задачи ВЫПОЛНИМОСТЬ. В данной диссертационной работе используется непрерывная формулировка задачи ВЫПОЛНИМОСТЬ для ДНФ:

$$\min_{x\in R^N[0,1]} F(x) = \min_{x\in R^N[0,1]} \sum_{i=1}^{M} c_i(x)$$

$$c_i(x) = \prod_{j=1}^{N} p_{i,j}(x) \qquad (2)$$

$$p_{i,j}(x) = \begin{cases} (x_j - 1)^2, \text{если } y_j \text{ входит в } G_i \\ (x_j)^2, \quad \text{если } \bar{y}_j \text{ входит в } G_i \\ 1, \quad \text{иначе} \end{cases}$$

Преимущества данной модели задачи ВЫПОЛНИМОСТЬ по сравнению с рассмотренными в данной работе моделями в том, что целевая функция (полином) имеет наименьшую возможную (квадратичную) степень по переменным среди дифференцируемых положительно определенных функций в R^N.

Выбор модели определяет способ построения метода решения задачи. Программная реализация метода решения называется решателем. В зависимости от пространства, в котором ведется поиск решатели можно разделить на дискретные и непрерывные. В зависимости от полноты решения алгоритмы решателей можно разделить на полные и неполные.

Одним из самых первых полных (дискретных) переборных алгоритмов следует считать процедуру DP, разработанную в 1960 году американскими математиками M.Davis и H.Putnam. Данный алгоритм основан на классическом методе резолюций математической логики.

В процедуре DP выбирается одна из переменных и вычисляются все резольвенты по этой переменной. Вычисленные резольвенты конъюнкцией добавляются к исходной КНФ. Все дизъюнкты исходной КНФ, участвовавшие в формировании резольвент, удаляются. При этом полученная формула эквивалентна исходной в том смысле, что она выполнима тогда и только тогда, когда выполнима исходная КНФ. Процесс повторяется для всех переменных. Процедура завершает работу, если все резольвенты вычислены, либо если получен пустой дизъюнкт. В последнем случае исходная КНФ не имеет решения.

Развитием процедуры DP стал полный алгоритм поиска DPLL, представленный американскими математиками M.Davis, H.Putnam, G.Logemann, D.Loveland в 1962 году. Данный алгоритм основан на т.н. методе разделения с возвратом (бэктрекинга). На основе некоторой эвристики выбирается переменная, которой присваивается значение 1. КНФ разрешается относительно данной переменной, затем рекурсивно проверяется выполнимость полученной КНФ. Если формула невыполнима, алгоритм возвращается на шаг назад и выбранной на этом шаге переменной присваивается противоположное значение 0. По сути, на каждом шаге рекурсии КНФ разбивается на две формулы, полученные из данной КНФ путем установки выбранной переменной в значение 1 и 0 соответственно. В итоге получаем дерево формул, корень которого — исходная КНФ, узлы – формулы, получаемые из предыдущих формул путем разрешения по выбранной переменной. Дуги (ветви) могут быть обозначены по именам литералов, относительно которых происходит разрешение. Данное дерево называется деревом поиска решения.

На каждом шаге формула дополнительно упрощается по правилу поглощения, путем удаления тавтологий, заблокированных скобок и разрешения КНФ по уникальным и чистым переменным.

Другим подходом в решении задачи ВЫПОЛНИМОСТЬ является метод локального поиска. Примером алгоритма локального поиска является следующий: исходная КНФ представляется в виде целевой функции, представляющей число невыполненных дизъюнктов на данном наборе значений переменных, затем некоторым образом решается задача минимизации целевой функции.

Итерационная процедура локального поиска основана на поиске, так называемого соседнего вектора такого, что на каждом шаге итераций функционал не увеличивается.

Локальный поиск эффективен по двум причинам. Во-первых, метод не перебирает все точки пространства поиска, а фокусируется на одной траектории поиска вектора решения в зависимости от стартового вектора. Во-вторых, каждая итерация состоит из поиска соседнего вектора для улучшения текущего вектора приближений. А так как размерность целевой функции полиномиально зависит от размерности входных данных (КНФ), то итерации происходят относительно быстро.

Основным недостатком методов локального поиска является их тенденция к уменьшению сходимости вблизи точек локальных минимумов целевой функции. Для преодоления локальных минимумов вводят дополнительные методики: эвристика минимизации конфликтов, эвристика наилучших соседей, эвристики случайного выбора, случайный шум, туннелирование. Обзор методов преодоления локальных минимумов представлен в работах J.Gu.

Традиционные алгоритмы локального поиска выбирают следующее приближение, только исходя из условия, чтобы целевая функция не увеличивалась. При этом среди всех возможных приближений (соседних векторов) не выбирают наилучшее (в смысле наискорейшего спуска целевой функции). Среди методов непрерывной оптимизации существуют градиентные методы (например, метод Ньютона), способные эффективно находить направление наискорейшего спуска функционала. Таким образом, мы естественным образом приходим к идее глобальной оптимизации.

Методы глобальной оптимизации направлены на поиск глобального экстремума задачи. При этом подходы, которые рассматриваются в рамках данной группы методов, отличаются поиском в некотором роде оптимальной траектории поиска экстремума. Методы глобальной оптимизации могут использовать подходы, реализованные в методах локального поиска (например, градиентный локальный поиск), а также подходы непрерывной оптимизации и подходы из линейной алгебры. Одним из ярких примеров среди непрерывных методов глобальной оптимизации является классический непрерывный метод Лагранжа.

Во **второй главе** обосновывается выбор функционала специального вида, ассоциированного с задачей ВЫПОЛНИМОСТЬ. Предлагается применить к решению системы нелинейных алгебраических уравнений, определяющих стационарные точки функционала, модифицированный метод последовательных приближений (с «инерцией»).

Общая схема метода последовательных приближений с «инерцией» в применении к решению задачи ВЫПОЛНИМОСТЬ следующая.

1. Осуществляется переход от задачи ВЫПОЛНИМОСТЬ к эквивалентной задаче поиска экстремума в пространстве $R^N[0,1]$. Переход основан на

построении функционала специального вида, глобальный минимум которого соответствует решению исходной задачи.

2. Для поиска точки глобального минимума построенного функционала применяется модифицированный метод последовательных приближений.

3. Найденная точка глобального минимума однозначно преобразуется в решающий набор исходной задачи.

Переход от задачи ВЫПОЛНИМОСТЬ к задаче поиска глобального минимума функционала происходит по формуле (2). При этом справедливо следующее утверждение 1:

Утверждение 1. Задача ВЫПОЛНИМОСТЬ для КНФ имеет решение, тогда и только тогда, когда <=> $\exists\ x^* \in R^N : \min\limits_{x \in R^N} F(x) = F(x^*) = 0$.

Дифференцируя функционал по всем переменным, и приравнивая к 0, получаем систему нелинейных уравнений для поиска стационарных точек:

$$\left(\sum_{\xi \in \{\Xi(x_i)\}} \prod_{j \neq i}^{N} p_{i,j}(x_j) \right) \cdot x_i - \sum_{\xi \in \{\Lambda(x_i)\}} \prod_{j \neq i}^{N} p_{i,j}(x_j) = 0$$

$$i = 1 \ldots N \qquad (3)$$

$$\{\Xi(x_i)\} \underset{def}{=} \Xi \underset{def}{=} \{\text{множество индексов } k : x_i \text{ или } \bar{x}_i \text{ входит в } G_k\}$$

$$\{\Lambda(x_i)\} \underset{def}{=} \Lambda \underset{def}{=} \{\text{множество индексов } k : x_i \text{ входит в } G_k\}$$

Для поиска решения системы уравнений (3) предлагается применить модифицированный метод последовательных приближений, так называемый метод последовательных приближений с «инерцией», описанный в работах Файзуллина Р.Т., где он хорошо зарекомендовал себя при решении систем нелинейных уравнений гидравлики:

$$\left[\sum_{k=0}^{K-1} \alpha_k \left(\sum_{\xi \in \Xi} \prod_{j \neq i}^{N} p_{i,j}(x_j(t-k)) \right) \right] \cdot x_i(t+1) = \sum_{\xi \in \Lambda} \prod_{j \neq i}^{N} p_{i,j}(x_j(t)) \underset{def}{=}$$

$$A_i(t) \cdot x_i(t+1) = B_i(t) \qquad (4)$$

$$\sum_{k=0}^{K-1} \alpha_k = 1,\ \alpha_k \in R,\ \alpha_k \geq 0,\ K \geq 1$$

Отметим, что классический метод Ньютона в применении к решению системы уравнений (3) неустойчив, что подтверждается теоремой 1.

Теорема 1. Решение системы нелинейных уравнений (3) входит в ядро производного оператора.

Далее в работе разрабатываются методики улучшения сходимости метода (4).

Переход к задаче 3-ВЫПОЛНИМОСТЬ. Любая процедура решения системы (3) при произвольной длине дизъюнктов будет естественным образом приводить к большим ошибкам округления. Другая проблема, возникающая при

решении задач ВЫПОЛНИМОСТЬ, заключается в том, что градиентные методы могут не сходиться вообще. Частично это подтверждается теоремой 2.

Теорема 2. Существует класс КНФ, для которого существует траектория, на которой скорость сходимости метода Ньютона становится линейной и бесконечно малой при произвольной длине дизъюнктов и выборе модели представления КНФ (2).

В данном случае проблема заключается в том, что большое количество литералов в дизъюнкте влечет высокий порядок касания функционала в точке минимума.

Одним из способов преодоления этой технической трудности является представление исходной КНФ в виде эквивалентной 3-КНФ и применение метода поиска решения уже к 3-КНФ.

Однако при переходе к 3-КНФ количество переменных и дизъюнктов увеличивается, что также может негативно сказываться на сходимости методов решения задачи ВЫПОЛНИМОСТЬ. Альтернативным подходом является предварительная обработка исходной КНФ с помощью правил резолюции и уже затем сведение к 3-КНФ.

Стратегия применения правил резолюции. Решение многих практических задач сводится к решению специальным образом сформированных КНФ. Проблема состоит в том, что получаемые КНФ имеют большое количество дизъюнктов и переменных даже при небольшой размерности исходной задачи. С помощью применения правил резолюции можно добиться преобразования исходной КНФ в эквивалентную КНФ, с меньшим количеством дизъюнктов и переменных. Эквивалентность означает, что множества решающих наборов обеих формул одинаковы.

Стратегия резолюции основана на классическом методе резолюций математической логики для автоматического доказательства теорем.

Используя правила вывода:

$\text{TRUE} \vee x \sim \text{TRUE}$

$\text{FALSE} \vee x \sim x$

$x \vee \neg x \sim \text{TRUE}$

$x \wedge \neg x \sim \square$ (пустой дизъюнкт)

наряду с вычислением резольвент (метод бинарной резолюции):

$(x \vee P) \wedge (\neg x \vee Q) \rightarrow P \vee Q$

удается значительно сократить размер КНФ (число дизъюнктов), а в отдельных случаях даже получить решение.

Бинарная резолюция сама по себе является полным алгоритмом решения задачи ВЫПОЛНИМОСТЬ, но из-за больших временных затрат не нашла применения в данной области. Поэтому при практическом использовании бинарной резолюции необходимо ограничивать глубину рекурсии. Резольвенты могут вычисляться из дизъюнктов, которые сами являются резольвентами, добавленными на предыдущих шагах метода. Глубина рекурсии устанавливает, до

каких пор нужно брать резольвенты от резольвент. Например, глубина рекурсии 1 устанавливает, что можно брать резольвенты только от дизъюнктов исходной КНФ.

Выбор начального приближения. Сходимость методов последовательных приближений к решению зависит от выбора начального приближения. Во всех случаях лучше выбирать начальное приближение достаточно близкое к решению. Методы преодоления локальных минимумов, позволяют перейти к хорошему приближению, стартуя с любого начального приближения. В общем случае, выбор начального приближения основан на информации, которая известна из исходной задачи априори. Когда условия задачи не дают информации для выбора начального приближения, оправдано стартовать со случайного целочисленного (0 или 1) вектора.

Добавление весовых множителей (регуляризация). Обобщением метода последовательных приближений с «инерцией» с использованием весовых параметров (4) является добавление весовых множителей в правые части системы нелинейных уравнений (3) и к самим компонентам приближения:

$$\left[\sum_{k=0}^{K-1}\alpha_k\left(\sum_{\xi\in\Xi}\prod_{j\neq i}^{N}p_{i,j}(x(t-k))\right)\right]\cdot\beta_i\cdot x_i(t+1)=\sum_{k=0}^{K-1}\alpha^*{}_k\left(\sum_{\xi\in\Lambda}\prod_{j\neq i}^{N}p_{i,j}(x(t-k))\right)\sim$$

$$A_i(t)\cdot x_i(t+1)=B_i(t)$$

$$\sum_{k=0}^{K-1}\alpha_k=1,\ \alpha_k\in R,\ \ \alpha_k\geq 0,$$

$$\sum_{k=0}^{K-1}\alpha^*{}_k=1,\ \alpha^*{}_k\in R,\ \ \alpha^*{}_k\geq 0\qquad\qquad \beta_i\in R,\ \ \beta_i\geq 0,\ K\geq 1$$

Смысл весовых множителей α_k заключается в том, чтобы установить влияние предыдущих итераций при формировании следующего приближения. Смысл весовых множителей $\alpha^*{}_k$ заключается в том, чтобы установить влияние предыдущих итераций на правые части системы нелинейных уравнений при формировании следующего приближения. Смысл весовых множителей β_i заключается в том, чтобы установить степень влияния каждой компоненты вектора переменных при формировании следующего приближения.

Таким образом, на каждой итерации происходит неравномерное продвижение по каждой оси многомерного пространства.

Сдвиг по антиградиенту является поправкой текущего приближения, для улучшения сходимости модифицированного метода последовательных приближений с «инерцией» и производится по формуле:

$$\bar{x}(t+1)=\bar{x}(t)+\frac{\delta}{A(x(t))}\cdot\bar{r}(t),\quad \bar{r}(t)=A(x(t))\cdot x(t)-B(x(t)),\quad \delta\in R$$

Метод смены траектории и туннелирования. Поиск точки глобального минимума градиентными методами может затрудняться из-за наличия точек локальных минимумов функционала (2), так как траектория, образованная

последовательными приближениями имеет тенденцию сходиться к ближайшей точке локального минимума. Возможное наличие точек локальных минимумов обосновано теоремой 3.

Теорема 3. Существует класс КНФ, для которых ассоциированный функционал (2) имеет стационарные точки, отличные от решения соответствующей задачи ВЫПОЛНИМОСТЬ.

Метод смены траектории позволяет выйти из локального минимума, с помощью формирования нового вектора приближений, который бы обладал свойствами не худшими, чем текущий вектор приближений, но позволял бы продолжить поиск решения.

Рассмотрим множества переменных:

$$E_K = \{x_i \mid \text{сумма } G_i(x(t)) = K, \text{причем} x_i \text{ или } \bar{x}_i \text{ входит в } G_i \} \qquad (5)$$

Введем вероятности $P_k > 0 : \sum_{i=1}^{K} P_i = 1$. С вероятностью P_k поменяем каждое значение $x_i(t) \in E_K$ на противоположное. При этом, вероятность того, что другие слагаемые функционала $F(x)$ (2) примут значение >0, зависит от P_k и размерности слагаемых (скобок) из E_K и тем ниже, чем меньше K.

Экспериментально установлено, что при K=0 или K=1 полученный вектор обладает свойствами не худшими, чем исходный вектор приближения. Количество слагаемых, принимающих значение >0 до и после операции примерно одинаково и может быть оценено с помощью теоремы 4.

Теорема 4. Любой дизъюнкт 3-КНФ, сгенерированной псевдослучайным заполнением дизъюнктов (распределение равномерное) из n переменных (n>2) после применения метода смены траектории (K=0) примет значение 0 с вероятностью не выше $[3(2n-1)(2n-2)+18(2n-2)+57]/[8n(2n-1)(2n-2)]$.

Используя полученный вектор в качестве нового начального приближения, метод приближений быстро (в большинстве случаев за 5-10 итераций) находит следующее приближение, на котором функционал $F(x)$ достигает значения не больше, чем на векторе $x(t)$. При этом, очень часто, удается проскочить область локального минимума. При дальнейшем движении по новой траектории метод может зациклиться в другой области локального минимума. Тогда метод смены траектории повторяется.

Ясно что, чем выше K и больше вероятности P_k для «больших» K, тем меньшую роль играет рестарт программы.

Рестарт. Когда процедура поиска решения зацикливается и применение методов туннелирования и смены траектории эффекта не дает, применяют рестарт. То есть процедура стартует заново с новым начальным приближением. Рестарт может включать в себя не только смену начального приближения, но и коррекцию параметров процедуры поиска решения. В случае модифицированного

метода последовательных приближений с «инерцией» — это параметры методов сдвига по антиградиенту, туннелирования, смены траектории и др.

Выбор метода проектирования. Проектирование текущего вещественного вектора приближений на булево пространство и обратный переход от булева вектора к вещественному вектору (с координатами 0 или 1) используется во многих подпрограммах модифицированного метода последовательных приближений. Например, при проверке текущего приближения на решение исходной задачи ВЫПОЛНИМОСТЬ, в методе туннелирования для определения следующего приближения при преодолении локальных минимумов, методе смены траектории для определения следующего приближения при преодолении плато и др. Экспериментально установлено, что следующий метод проектирования является достаточно эффективным:

$$\begin{cases} x_i > 0.5 + \varepsilon_4 \Rightarrow y_i = \text{ИСТИНА} \\ x_i < 0.5 - \varepsilon_4 \Rightarrow y_i = \text{ЛОЖЬ} \\ x_i \in [0.5 - \varepsilon_4, 0.5 + \varepsilon_4] \Rightarrow y_i = RAND(\text{ИСТИНА}, \text{ЛОЖЬ}) \end{cases}$$

И в обратную сторону: $y_i = \text{ИСТИНА} \Rightarrow x_i = 1,\ y_i = \text{ЛОЖЬ} \Rightarrow x_i = 0$

RAND(x, y) – функция, которая равновероятным образом возвращает одно из двух значений x или y.

Увеличение разрядности. Для уменьшения влияния ошибок округления на результат при реализации итерационной процедуры было предусмотрено использование вычислений с произвольной точностью. Исследования сходимости метода при увеличении разрядности вычислений показали преимущество использования типа DOUBLE (двойная точность, 64 бит) по сравнению с типом FLOAT (одинарная точность, 32 бит). При этом дальнейшее увеличение разрядности к значимому эффекту не приводит.

Гибридный метод последовательных приближений с «инерцией» заключается в объединении описанных методик в единый метод, названный гибридным (модифицированным) методом последовательных приближений с «инерцией». Метод аккумулирует наиболее эффективные приемы, используемые в ведущих алгоритмах решения задачи ВЫПОЛНИМОСТЬ, а также новые разработанные приемы повышения эффективности методов глобального градиентного спуска.

Основная процедура состоит из последовательных итераций, которые совмещают метод последовательных приближений с «инерцией» (по схеме Зейделя) и сдвиг по антиградиенту.

Способы распараллеливания метода. Гибридный алгоритм допускает целый спектр способов распараллеливания.

Способ 1. ДНФ, эквивалентная исходной КНФ делится на n независимых частей (подформул). Для каждой из подформул, с помощью основного алгоритма, ищется вектор решения. Полученные вектора используются в качестве начального приближения для поиска решения следующей подформулы. После нескольких

итераций вычисляется «усредненный» вектор, который используется в качестве начального приближения для поиска решения исходной КНФ.

Способ 2. Может быть реализован при наличии нескольких эквивалентных КНФ с различной структурой. Дело в том, что метод последовательных приближений с «инерцией» формирует различные траектории поиска для КНФ с различной структурой. Это объясняется, например, тем, что в различных КНФ веса одноименных переменных могут различаться. Различные эквивалентные КНФ можно получить, например, с помощью преобразования с применением правил резолюции либо добавлением избыточной информации к исходной КНФ. Вычисления проводятся барабанным методом. Очередное приближение для итерационной процедуры, ассоциированной с одной из КНФ, подставляется в качестве начального приближения итерационной процедуры, ассоциированной со следующей КНФ. Процедура повторяется до тех пор, пока не будет найден выполняющий набор для одной из КНФ.

В **третьей главе** проводится анализ разработанного метода последовательных приближений с «инерцией». Приводятся результаты испытания метода и его модификаций на различных типах задач. Основное внимание уделяется задаче ВЫПОЛНИМОСТЬ, ассоциированной с задачей факторизации чисел больших размерностей. Показано, что метод может быть с успехом применен к другим задачам информатики.

Для проведения численных экспериментов было отобрано по 500 экземпляров тестовых примеров для каждого типа выбранных генераторов КНФ. Всего около 5000 тестовых КНФ. Были отобраны следующие генераторы: RTI (КНФ, сформированные случайным образом), BMS (КНФ с минимальным хребтом), CBS (КНФ с хребтом, фиксированного размера), UF (унифицированные случайные 3-КНФ), BW (КНФ, ассоциированные с задачей «перекладывание колоды»), GCP (КНФ, ассоциированные с задачей раскраски графов), LPP (КНФ, ассоциированные с задачей логистики), FACTOR (КНФ, ассоциированные с задачей факторизации), DLOG (КНФ, ассоциированные с задачей дискретного логарифмирования), EDLOG (КНФ, ассоциированные с задачей дискретного логарифмирования на эллиптической кривой). Входные параметры алгоритмов сведения задач криптоанализа выбирались исходя из рекомендуемых соответствующими стандартами условий обеспечения критпостойкости.

После каждой модификации проводилось тестирование метода для определения эффективности проделанных изменений.

Стратегия применения правил резолюции. Для определения эффективности стратегии резолюции были проведены численные эксперименты.

Результаты показывают, что структура выбранных типов тестовых КНФ из библиотеки SATLib, позволяет успешно применять стратегию резолюции. Почти всегда удается разрешить более 40% переменных и уменьшить число дизъюнктов КНФ до 50%, что приводит к очень быстрому (за 5-10 итераций) поиску точного решения методом последовательных приближений с «инерцией». В то же время,

на синтезированных (не основанных на прикладных задачах) тестовых примерах данная стратегия практического эффекта не дает.

В применении к КНФ, ассоциированных с задачей факторизации (размерность факторизуемого числа до 1024 бит включительно), показано, что с помощью стратегии резолюции можно разрешать до 1,1% переменных и сокращать до 67% дизъюнктов исходной КНФ. Как показал эксперимент, трудоемкость по времени данной стратегии с глубиной рекурсии 1 кубическая от количества бит в факторизуемом числе.

Для других задач криптоанализа также выявлено, что структура тестовых примеров позволяет успешно применять стратегию резолюции. Следует отметить, что всегда удается разрешить несколько значащих битов двоичного представления искомых чисел. В том числе, и старшие биты чисел, которые могут быть полезны для определения вспомогательных битов переноса.

Основной параметр стратегии применения правил резолюции – глубина рекурсии бинарной резолюции. Трудоемкость методики бинарной резолюции экспоненциально возрастает с ростом глубины рекурсии. В то же время число сокращенных дизъюнктов и количество разрешенных переменных сильно падает.

Переход к задаче 3-ВЫПОЛНИМОСТЬ. Данные вычислительных экспериментов по увеличению размерности КНФ в процентах от исходного числа дизъюнктов при сведении к 3-КНФ до и после предварительного применения метода резолюции показывают, что увеличение количества дизъюнктов в основном приемлемо. Хотя, размерность для некоторых типов задач увеличивается значительно, что приводит к значительному расширению области поиска решения.

Добавление весовых множителей (регуляризация). Добавление весовых множителей позволяет увеличить эффективность метода последовательных приближений с «инерцией». Хотя методики для определения оптимального набора весовых множителей разработано не было, тем не менее, ясен тот факт, что число решенных тестов стабильно уменьшается при последовательном уменьшении числа «левых» весовых множителей в наборе (параметр K из (4)).

Сдвиг по антиградиенту. Сдвиг по антиградиенту хорошо сокращает погрешности и ускоряет сходимость алгоритма. Число решенных примеров увеличилось примерно на 20%. Применение исключительно только данного приема позволило достаточно эффективно решать тестовые задачи из библиотеки SATLib, например, для КНФ серии UF20-91 удалось решить 703 теста из 1000. На примерах UF250-1065 алгоритм показал результат 6% от количества тестов (версия алгоритма без сдвига — 1%). На остальных тестах метод зацикливается в одной из областей локального минимума, что говорит о необходимости применения инструментов обхода локальных минимумов.

Метод смены траектории и туннелирование. Метод смены траектории и туннелирования (в качестве инструмента обхода локальных минимумов) значительно увеличивает число решаемых примеров. Удалось решить 100% тестов библиотеки SATLib за время сравнимое с лучшими решателями 2005 года.

Сравнительные результаты тестирования метода последовательных приближений с «инерцией» со всеми модификациями приведены в таблице 1.

Таблица 1. Гибридный метод последовательных приближений с «инерцией» — сравнительные результаты численных экспериментов для задач библиотеки SATLib.

Количество литералов (N)	Количество дизъюнктов (M)	Число КНФ группе	Время решен МППИ	Время решения RANOV	Время решения SATz
BW (КНФ, ассоциированные с задачей «перекладывание колоды»)					
48<N<6325	261<M<131973	7	23 сек.	2,20 мин.	2,03 мин.
BMS (КНФ с минимальным хребтом)					
100	<430	1000	6 сек.	12 сек.	6 сек.
CBS (КНФ с хребтом, фиксированного размера)					
100	449	2000	5 сек.	4 сек.	12 сек.
UF (унифицированные случайные 3-КНФ)					
250	1065	100	9,82 мин.	3 сек.	19 сек.
GCP (КНФ, ассоциированные с задачей раскраски графов)					
600	2237	100	11,23 мин.	2 сек.	0.5 сек.
GCP2 (КНФ, ассоциированные с задачей раскраски графов)					
500	3100	101	2 сек.	4 сек.	2 сек.

Обозначения полей:
Время решения МППИ — среднее время, затраченное на решение всей группы примеров метод последовательных приближений с «инерцией».
Время решения RANOV — среднее время, затраченное на решение всей группы пример неполным алгоритмом локального поиска RANOV (победитель соревнований решателей 20 года)
Время решения SATz — среднее время, затраченное на решение всей группы примеров полн (переборным) алгоритмом локального поиска SATz (4-е место на соревнованиях решателей 20 года)

Определение наиболее вероятных значений верных бит множителей в задаче факторизации. В исследованиях особое внимание уделяется задаче факторизация. Предложены дополнительные методики, увеличивающие эффективность метода в применении к КНФ, эквивалентных задаче факторизация. Среди них эквивалентные преобразования исходной КНФ путем добавления избыточной информации, известной о задаче априори.

1. Так, алгоритм представления задачи факторизация в виде КНФ, предложенный В.И. Дулькейтом кодирует операцию умножения простого числа p на простое число q классическим «столбиком». Без потери общности к условиям задачи факторизации числа $p \cdot q = q \cdot p$ можно добавить логическое ограничение $(p > q) \wedge (q < p)$. Представляя данное ограничение в виде КНФ и добавляя (знаком конъюнкции) к исходной КНФ получим эквивалентную формулу с избыточной информацией.

2. Рекомендующие документы по выбору параметров при генерации криптографических ключей для алгоритма RSA устанавливают необходимость выбора простых сомножителей p и q для формирования открытого ключа $n = p \cdot q$. Информация $n \mod r \neq 0$, где $r = 2,3,5,7,11,...$ легко проверяется.

Условие неделимости факторизуемого числа на выбранные числа записывается в виде КНФ и добавляется (знаком конъюнкции) к исходной КНФ.

Первая модификация позволяет строить функционал с единственной точкой глобального минимума. Исследования показали уменьшение времени работы модифицированного метода последовательных приближений в среднем на 7%.

Вторая модификация позволяет увеличить частоту переменных, отвечающих значимым битам искомых множителей, и таким образом «улучшить» характеристики функционала. Исследования показали, что такие формулы, обработанные методом резолюции, решаются до 50% быстрее по сравнению с исходными формулами. Если метод резолюций не применять, то размерность КНФ становиться очень большой.

Сравнительные результаты работы отобранных решателей на тестовых КНФ, ассоциированных с задачей факторизации, приводятся в таблице 2.

Таблица 2. Гибридный метод последовательных приближений с «инерцией» — сравнительные результаты численных экспериментов для задач ВЫПОЛНИМОСТЬ, эквивалентных задаче факторизация.

Число бит в фактории-зуемом числе	Кол-во переменных	Кол-во дизъюнктов	Время решения МППИ (мин)	Время решения RANOV (мин)	Время решения SATz (мин)
16	3001	156	0,0001	0,0000	0,0083
20	4983	245	0,0005	0,1922	0,0180
24	7457	354	0,0031	0,3174	0,1255
28	10435	483	0,0340	> 60,0000	30,2568
32	13901	632	0,1183	> 60,0000	> 60,0000
36	17871	801	0,0881	> 60,0000	> 60,0000
40	22333	990	1,8817	> 60,0000	> 60,0000
44	27291	1199	10,6301	> 60,0000	> 60,0000
48	32745	1428	61,9942	> 600,0000	> 600,0000
52	38691	1677	487,4967	> 600,0000	> 600,0000
56	45137	1946	11,3225	> 600,0000	> 600,0000
60	52079	2235	4959,6667	> 6000,0000	> 6000,0000
72	75885	3222	11520,0028	> 60000,0000	> 60000,0000

Обозначения полей:
Время решения МППИ — среднее время, затраченное на решение всей группы примеров метод последовательных приближений с «инерцией».
Время решения RANOV — среднее время, затраченное на решение всей группы пример неполным алгоритмом локального поиска RANOV (победитель соревнований решателей 20 года)
Время решения SATz — среднее время, затраченное на решение всей группы примеров полн (переборным) алгоритмом локального поиска SATz (4-е место на соревнованиях решателей 20 года)
Знак «<>» означает, что за указанное время решение найдено не было.

В рамках работы проводились исследования близости формируемых методом векторов приближений к вектору решения для задач факторизации больших размерностей. В качестве исходного материала для тестирования были выбраны

по 100 независимых примеров размерностей 1024, 2048, 3072 бит. Результаты показывают стабильное формирование 68% верных бит при росте размерности задачи. При этом величина доверительного интервала по всем размерностям не превышает 0,18. Максимальное (минимальное) число совпадающих бит так же стабильно - 68,3% (67,7%). При этом число верно определенных бит, отвечающих именно битам сомножителей равно 67,9%. При этом, результат достигается всего за 100-500 итераций, стартуя со случайно сформированного приближения. Отметим, что найденные переменные являются ключевыми для решения задачи, т.е. после подстановки их верных значений в исходную КНФ, формула оказывается легко разрешимой относительно оставшихся переменных. Дополнительно, в работе предложена система тестов, основанная на проверке обстоятельства кластеризации ненулевых строк в матрице умножения классическим «столбиком» в двоичной системе счисления. Тесты позволяют с вероятностью 0.8 определить до 31% конкретных неизвестных в задаче факторизации размерности 512 бит.

Полученные результаты позволяют сделать следующее **заключение**.

Разработаны и обоснованы методики, равномерно улучшающие сходимость метода последовательных приближений с «инерцией» на всех типах задач ВЫПОЛНИМОСТЬ. Эффективность гибридного метода не уступает эффективности «лучших» решателей 2005 года.

Разработана стратегия применения правил резолюции. Исследована и подтверждена эффективность использования стратегии резолюции как препроцессора КНФ в методе последовательных приближений с «инерцией».

Исследована применимость метода последовательных приближений с «инерцией» к КНФ, ассоциированным с задачами криптоанализа асимметричных шифров.

Для КНФ, эквивалентных задаче факторизации (с соблюдением всех условий криптостойкости RSA) размерностью до 72 бит были получены точные решения. При этом эффективность предложенного метода превосходит многие решатели задачи ВЫПОЛНИМОСТЬ. Приближения, формируемые методом, более чем на 68% совпадают с решением независимо от размерности задачи (до 3072 бит включительно), причем с ростом размерности наблюдается рост доли совпадающих компонент вектора приближения. Кроме того, разработана система дополнительных тестов, позволяющая с высокой долей вероятности определять конкретные биты сомножителей в задаче факторизации.

Список опубликованных работ

в ведущих рецензируемых научных изданиях, определенных Высшей аттестационной комиссией:

1. Дулькейт В.И., Файзуллин Р.Т., Хныкин И.Г. Алгоритм минимизации функционала, ассоциированного с задачей 3-SAT и его практические применения // Компьютерная оптика, т. 32, №1. - 2008. - С.68-73.

2. Дулькейт В.И., Файзуллин Р.Т. Хныкин И.Г. Метод решения задачи ВЫПОЛНИМОСТЬ и его применение для криптографического анализа асимметричных шифров // Доклады Томского государственного универистета систем управления и радиоэлектроники. - 2008. - ч. 1. 2(18). - С.54-56.

3. Дулькейт В.И., Файзуллин Р.Т., Хныкин И.Г. Непрерывные аппроксимации решения задачи ВЫПОЛНИМОСТЬ применительно к криптографическому анализу асимметричных шифров // Компьютерная оптика, т. 33, №1. - 2009. - С.86-91.

4. Хныкин И.Г. Эквивалентное преобразование КНФ, ассоциированных с практическими задачами с помощью правил резолюции // Системы управления и информационные технологии, 2(36), 2009. - С. 54-58.

в других изданиях:

5. Дулькейт В.И., Файзуллин Р.Е., Хныкин И.Г. Минимизация функционалов, ассоциированных с задачами криптографического анализа асимметричных шифров // Прикладная дискретная математика. - 2008. - № 2. - С.113-119.

6. Дулькейт В.И., Файзуллин Р.Т., Хныкин И.Г. Минимизация функционалов, ассоциированных с задачами криптографического анализа // Дифференциальные уравнения. Функциональные пространства. Теория приближений. Тез. докл. Международная конференция, посвященная 100-летию со дня рождения С.Л. Соболева. - Новосибирск: Ин-т математики СО РАН, 2008. - С.484-485.

7. Дулькейт В.И., Файзуллин Р.Т., Хныкин И.Г. Сведение задач криптоанализа асимметричных шифров к решению ассоциированных задач ВЫПОЛНИМОСТЬ // Сборник докладов XIII Всероссийской конференции «Математические методы распознавания образов». - М.: МАКС Пресс, 2007. - С.249-251.

8. Хныкин И.Г. Алгоритм минимизации функционала ассоциированного с задачей 3-SAT // Проблемы теоретической и прикладной математики: Труды 38-й Региональной молодежной конференции. - Екатеринбург: УрО РАН, 2007. - С.427-432.

9. Хныкин И.Г. Алгоритм минимизации функционала, ассоциированного с задачей SAT // Труды 13-й Всероссийской конференции «Математическое программирование и приложения», Информационный бюллетень АМП №11. (Тез. докл. Всерос. конф.) - Екатеринбург: УрО РАН, 2007.

10. Хныкин И.Г. Эквивалентное преобразование КНФ, ассоциированных с задачами криптографического анализа, с помощью правил резолюции // Прикладная дискретная математика, приложение. (Тез. докл. конф. SIBECRYPT'09) - Томск: ТГУ, 2009.