На правах рукописи

Заикин Олег Сергеевич

ПАРАЛЛЕЛЬНАЯ ТЕХНОЛОГИЯ РЕШЕНИЯ SAT-ЗАДАЧ И ЕЕ РЕАЛИЗАЦИЯ В ВИДЕ ПАКЕТА ПРИКЛАДНЫХ ПРОГРАММ

05.13.01 – Системный анализ, управление и обработка информации
(в отраслях информатики, вычислительной техники и автоматизации)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Томск – 2009

Работа выполнена в Институте динамики систем и теории управления СО РАН.

Научный руководитель:

кандидат технических наук Семенов Александр Анатольевич

Официальные оппоненты:

доктор физико-математических наук, профессор Корольков Юрий Дмитриевич

кандидат технических наук Тимошевская Наталия Евгеньевна

Ведущая организация:

Институт вычислительного моделирования СО РАН (г. Красноярск)

Защита состоится:

22 января 2008 г. в 10.30 на заседании Диссертационного совета Д 212.267.12 при ГОУ ВПО «Томский государственный университет» по адресу: 634050, г. Томск, пр. Ленина, 36.

С диссертацией можно ознакомиться:

В научной библиотеке Томского государственного университета.

Автореферат разослан 21 декабря 2008 г.

Ученый секретарь диссертационного совета, д.т.н., профессор Смагин В. И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Многие значимые в практическом отношении проблемы, связанные с управлением и обработкой информации в дискретных системах, допускают эффективные сводимости к задачам поиска решений логических уравнений. Сказанное касается проблем синтеза и верификации в микроэлектронике, целого ряда вопросов теоретического программирования, задач обращения дискретных функций, задач управления коммуникационными протоколами и многих других. Один из наиболее важных классов логических уравнений образован уравнениями вида «КНФ = 1» (КНФ – конъюнктивная нормальная форма). Задачи поиска решений данного класса уравнений относятся к так называемым «SAT-задачам». Для решения SAT-задач используются специальные программные комплексы, называемые SAT-решателями. В построении SAT-решателей в последние годы отмечен реальный прогресс.

Высокая вычислительная сложность SAT-задач аргументирует использование для их решения параллельных вычислений. С этой позиции SAT-задачи могут исследоваться по двум направлениям.

Первое направление основывается на концепции мелкозернистого параллелизма. В этом случае распараллеливание осуществляется на уровне базового алгоритма SAT-решателя (SAT-решатели с параллельной вычислительной архитектурой). Однако современные SAT-решатели являются довольно сложными в структурном смысле программами, в связи с чем реализация данного подхода трудна в техническом отношении. Кроме этого, мелкозернистый параллелизм зачастую характеризуется невысокой масштабируемостью.

В основе второго направления лежит концепция крупноблочного параллелизма. В этом случае осуществляется декомпозиция исходной SAT-задачи на семейство подзадач, каждая из которых обрабатывается SAT-решателем на отдельном вычислительном узле параллельной вычислительной системы. При этом параллельная вычислительная система свободна от ряда ограничений, присущих мелкозернистому параллелизму. Кроме этого, крупноблочный подход, как правило, характеризуется высокой масштабируемостью. Крупноблочный параллелизм предпочтителен в решении SAT-задач, в которых размерность пространства полного перебора существенно меньше числа булевых переменных, фигурирующих в КНФ. Речь идет, прежде всего, о SAT-задачах, к которым сводятся задачи обращения дискретных функций из класса, имеющего пересечения с многочисленными областями кибернетики.

Сказанное означает, что проблема создания эффективных технологий решения SAT-задач посредством их крупноблочного распараллеливания является актуальной.

Цель работы и задачи исследования

Целью диссертационной работы является разработка и практическая реализация технологии крупноблочного распараллеливания SAT-задач.

Для достижения указанной цели ставятся и решаются следующие основные задачи.

1. Разработка и исследование различных стратегий построения декомпозиционных представлений SAT-задач.

2. Разработка и алгоритмическая реализация процедуры прогнозирования трудоемкости параллельного решения SAT-задач.

3. Программная реализация технологии крупноблочного распараллеливания SAT-задач в форме пакета прикладных программ (ППП).

4. Решение с помощью разработанного ППП SAT-задач, кодирующих задачи обращения ряда дискретных функций, используемых в криптографии.

Методы исследования

Теоретические исследования используют аппарат теории множеств, дискретной математики, теории вычислительной сложности, параллельных вычислений, а экспериментальные – инструментальные и программные средства решения задач в параллельных вычислительных системах.

Научная новизна

Новыми являются все основные результаты, полученные в диссертации, в том числе:

– общая схема крупноблочного распараллеливания SAT-задач;

– процедура прогнозирования трудоемкости параллельного решения SAT-задач;

– оценка сложности процедуры прогнозирования трудоемкости параллельного решения SAT-задач;

– разработка ППП для реализации параллельной технологии решения SAT-задач;

– параллельное решение с помощью разработанного ППП вычислительно сложных SAT-задач, кодирующих задачи обращения некоторых дискретных функций, используемых в криптографии.

При решении поставленных задач были получены перечисленные ниже вспомогательные результаты, представляющие самостоятельный интерес.

Введено понятие декомпозиционного множества; введено понятие схемы формирования семейств декомпозиционных множеств (далее «схема формирования»); введена классификация схем формирований; разработаны схемы формирования, ориентированные на структуру SAT-задач, кодирующих задачи обращения некоторых дискретных функций; реализована процедура оптимизированной рассылки SAT-задач, снижающая транспортные расходы в параллельных вычислительных системах.

Теоретическая и практическая значимость работы

Теоретическая значимость работы аргументируется возможностью прямого переноса ее базовых концепций на задачи дискретной оптимизации с двоичными данными. Разработанные в диссертации методы и алгоритмы могут применяться для решения широкого класса практически важных задач, допускающих эффективную сводимость к SAT-задачам.

Достоверность результатов

Достоверность полученных в работе теоретических результатов обеспечивается строгостью производимых математических построений. Корректность и эффективность практической реализации подтверждается результатами вычислительных экспериментов.

Основные результаты, выносимые на защиту

1. Параллельная технология решения SAT-задач, включающая общую схему крупноблочного распараллеливания SAT-задач.
2. Разработка и анализ различных схем формирования декомпозиционных представлений SAT-задач.
3. Процедура прогнозирования трудоемкости параллельного решения SAT-задач; оценка сложности данной процедуры.
4. Разработка ППП для практической реализации предложенной параллельной технологии решения SAT-задач.
5. Параллельное решение с помощью разработанного ППП вычислительно сложных SAT-задач, кодирующих задачи обращения некоторых дискретных функций, используемых в криптографии.

Внедрение результатов исследований

Результаты диссертации внедрены:

1) в разработку технологии крупноблочного распараллеливания SAT-задач в рамках следующих грантов РФФИ: № 04-07-90358 (2006 г.), № 07-01-00400-а (2007-2008 гг.), а также грантов поддержки ведущих научных школ: НШ-

9508.2006.1 (2006 г.), НШ-1676.2008.1 (2008 г.);

2) в создание ППП Distributed-SAT (D-SAT), предназначенного для параллельного решения SAT-задач: свидетельство № 2008610423 об официальной регистрации программы для ЭВМ, выданное Федеральной службой по интеллектуальной собственности, патентам и товарным знакам (2008 г.);

3) в учебный процесс (разработка курсовых и дипломных работ) Института математики, экономики и информатики Иркутского государственного университета (ИМЭИ ИГУ).

Апробация работы

Результаты диссертации докладывались и обсуждались на 1-ой и 2-ой Международных научных конференциях «Параллельные вычислительные технологии (ПАВТ)» (Челябинск, 2007; Санкт-Петербург, 2008), на 4-ой Международной конференции «Параллельные вычисления и задачи управления (PACO'2008)» (Москва, 2008), на XV Международной конференции «Вычислительная механика и современные прикладные программные системы (ВМСППС'2007)» (Алушта, 2007), на XIV Байкальской международной школе-семинаре «Методы оптимизации и их приложения» (Северобайкальск, 2008 г.) на VI и VII Сибирских научных школах-семинарах с международным участием «Компьютерная безопасность и криптография» (Горно-Алтайск, 2007; Красноярск, 2008), на II Всероссийской конференции с международным участием «Инфокоммуникационные и вычислительные технологии и системы (ИКВТС'2006)» (Улан-Удэ, 2006), на 6-ой школе-семинаре «Распределенные и кластерные вычисления» (Красноярск, 2006), на научных конференциях «Ляпуновские чтения» (Иркутск, 2006; 2007), на VI и IX школах-семинарах молодых ученых «ММИТ» (Иркутск, 2005; 2007); на научных семинарах Института динамики систем и теории управления СО РАН, а также на научном семинаре кафедры защиты информации и криптографии Томского государственного университета (2008 г.).

Структура работы и личный вклад автора

Диссертация состоит из введения, трех глав, заключения и списка литературы, насчитывающего 104 наименования. Объем диссертации – 116 страниц. Первая глава является обзорной, в ней отражены в основном известные результаты. Часть результатов второй главы (введение прогнозных функций, оценка сложности процедур прогнозирования времени параллельного решения SAT-задач) получены в неделимом соавторстве с научным руководителем. Все остальные результаты, представленные в диссертации, получены автором лично.

Публикации

Основные результаты диссертации опубликованы в 18 печатных работах ([1–18] в списке литературы). Работы [1] и [2] опубликованы в изданиях, входящих в список ВАК.

СОДЕРЖАНИЕ РАБОТЫ

Во введении дается общая характеристика работы, определяется тематика и цель работы, излагаются основные результаты, выносимые на защиту.

В первой главе описывается SAT-подход в решении задач обращения дискретных функций. Приведен обзор основных компонентов архитектуры современных SAT-решателей и концепция логического криптоанализа.

Рассматривается класс логических уравнений, образованный уравнениями вида $C(x_1,\ldots,x_n)=1$, где $C(x_1,\ldots,x_n)$ – некоторая конъюнктивная нормальная форма (КНФ) над множеством булевых переменных $X=\{x_1,\ldots,x_n\}$. Задачи решения таких уравнений называются *SAT-задачами*. К SAT-задачам относятся задача проверки выполнимости произвольной КНФ (исторически первая NP-полная задача) и задача поиска выполняющего набора произвольной КНФ (NP-трудная задача).

Обозначим через $\{0,1\}^n$ множество всех двоичных слов длины n, $\{0,1\}^* = \bigcup_{n\in\mathrm{N}}\{0,1\}^n$. *Дискретной функцией* называется произвольная функция вида $f_n:\{0,1\}^n \to \{0,1\}^*$. Через $dom\ f_n \subseteq \{0,1\}^n$ обозначается область определения функции f_n, а через $range\ f_n \subset \{0,1\}^*$ – ее область значений. Дискретную функцию f_n назовем *всюду определенной*, если $dom\ f_n = \{0,1\}^n$. Семейство всюду определенных дискретных функций, вычислимых некоторой программой машины Тьюринга, обозначается через $f=\{f_n\}_{n\in\mathrm{N}}$. Класс $\Im$ образован всеми алгоритмически вычислимыми за полиномиальное от n время всюду определенными семействами дискретных функций. *Задача обращения дискретной функции* f_n семейства $f\in\Im$ ставится следующим образом: дано двоичное слово $y\in range\ f_n$, требуется найти такое слово $x\in\{0,1\}^n$, что $f_n(x)=y$.

В работе[1] было предложено (со ссылками на результаты С.А. Кука) рассматривать проблемы обращения функций $\{f_n\}_{n\in\mathrm{N}}$ произвольного семейства $f \in \Im$ как задачи поиска решений логических уравнений вида $C(x_1,\ldots,x_{p(n)})=1$, где $p(\cdot)$ – некоторый полином, а $C(x_1,\ldots,x_{p(n)})$ – КНФ над множеством булевых переменных $X=\{x_1,\ldots,x_{p(n)}\}$. Функция объема получаемого при этом семейства КНФ есть некоторый полином от n.

Функция объема получаемого при этом семейства КНФ есть некоторый полином от n.

Консервативность сведения исходного уравнения к уравнению вида $C(x_1,\ldots,x_{p(n)})=1$ означает равномощность множеств решений этих уравнений.

В настоящее время известны консервативные процедуры сведения задач обращения дискретных функций из класса $\Im$ к SAT-задачам. В основе данных процедур лежат преобразования Цейтина.

В соответствии с перечисленными результатами общая концепция SAT-подхода в решении задач обращения дискретных функций состоит в следующем.

– Алгоритм вычисления произвольной функции f_n семейства $f=\{f_n\}_{n\in\mathrm{N}}, f\in\Im$, консервативно преобразуется в уравнение вида $C(x_1,\ldots,x_{p(n)})=1$.

– В уравнение $C(x_1,\ldots,x_{p(n)})=1$ подставляется известный вектор $y \in range\ f_n$.

– Полученная в результате КНФ $C|_y$ является выполнимой, а из выполняющего ее набора можно выделить компоненты вектора $x\in\{0,1\}^n$ такого, что $f_n(x)=y$.

Тем самым задача обращения функции f_n в точке $y \in range\ f_n$ оказывается сведенной к SAT-задаче.

Для решения SAT-задач используются специальные программные комплексы, называемые *SAT-решателями*. Большая часть эффективных (по результатам специализированных конкурсов) SAT-решателей использует в

---

[1] Заикин О.С., Семенов А.А. Технология крупноблочного параллелизма в SAT-задачах // Проблемы управления. – 2008. – № 1. С. 43-50.

качестве основы алгоритм DPLL. Пусть $C(x_1,\ldots,x_n)$ – произвольная КНФ над множеством булевых переменных $X=\{x_1,\ldots,x_n\}$. В основе DPLL лежит процедура итеративного означивания переменных из X в некотором порядке с последующей подстановкой значений в $C(x_1,\ldots,x_n)$. Данная процедура представляется в форме двоичного дерева. При этом вершины дерева интерпретируют выбор соответствующих булевых переменных, ветви дерева интерпретируют присвоения этим переменным значений истинности, а листья дерева интерпретируют значения булевой функции $C(x_1,\ldots,x_n)$ при соответствующем означивании переменных из X. Приведенная процедура в DPLL дополнена отслеживанием возможности срабатывания *правила единичного дизъюнкта*. Схема подстановок в КНФ значений переменных с последующим срабатыванием правила единичного дизъюнкта называется стратегией распространения булевых ограничений (*BCP-стратегия*). Помимо перечисленного современные SAT-решатели, основанные на DPLL, используют различные технологии, позволяющие ускорять поиск (нехронологический бэктрекинг, процедура запоминания информации о конфликтах, рестарты, быстрые структуры данных и др.).

В диссертации в основном рассматриваются SAT-задачи, полученные в результате сведения к ним криптографических задач из области поточного шифрования. Этот выбор обусловлен потребностью в аргументировано сложных тестах. Большинство используемых на практике поточных шифров относится к *двоичным аддитивным*. В таких шифрах на основе секретного ключа с помощью *генераторов ключевого потока* (делее – *генератора*) порождается *ключевой поток*, а шифртекст образуется путем сложения исходного текста по модулю 2 с ключевым потоком. Входная последовательность генератора называется *инициализирующей*.

Генератор можно рассматривать как семейство дискретных функций $g=\{g_n\}_{n\in\mathrm{N}}$ вида $g_n:\{0,1\}^n\to\{0,1\}^*$, $n\in\mathrm{N}$, $g\in\Im$. Алгоритм вычисления функций g_n называется *порождающим алгоритмом генератора*. Под криптоанализом генератора понимается поиск инициализирующей последовательности $x=(x_1,\ldots,x_n)$ по известному фрагменту y ключевого потока. Криптоанализ, рассматриваемый как SAT-задача, называется *логическим криптоанализом*. Обозначим через C^g КНФ, полученную в результате пропозиционального кодирования алгоритма, реализующего g. КНФ, получаемую в результате подстановки в C^g значений переменных, кодирующих

ключевой поток, обозначим через C_0^g. Тем самым задача поиска инициализирующей последовательности генератора g оказывается сведенной к SAT-задаче.

Во второй главе описана технология крупноблочного распараллеливания SAT-задач. Рассмотрены различные виды декомпозиционных представлений, и описана процедура прогнозирования трудоемкости параллельного решения SAT-задач.

Распределенная вычислительная среда (далее *РВС*) – совокупность вычислительных узлов, объединенных коммуникационной сетью. *Вычислительный узел* РВС – программно-аппаратный ресурс, включающий аппаратное и программное обеспечение, требуемое для решения одно или нескольких вычислительных задач. *Вычислительный кластер* (далее *кластер*) – совокупность вычислительных узлов РВС, объединенных единой коммуникационной средой. *Главный (управляющий) узел* РВС – один из вычислительных узлов РВС, обладающий функциями управления заданиями пользователей и ресурсами РВС.

Ниже приведено описание основных элементов разработанной в диссертации технологии крупноблочного распараллеливания SAT-задач.

Рассматривается произвольная КНФ C над множеством булевых переменных $X = \{x_1,\ldots,x_n\}$. Выбираем в множестве X некоторое подмножество

$$X' = \{x_{i_1},\ldots,x_{i_d}\},\ \{i_1,\ldots,i_d\} \subseteq \{1,\ldots,n\},\ d \in \{1,\ldots,n\}.$$

Назовем множество $X' = \{x_{i_1},\ldots,x_{i_d}\}$ *декомпозиционным множеством*, а d – *размерностью декомпозиционного множества*. Дополнительно полагаем, что при $d = 0$ декомпозиционное множество пусто. Декомпозиционному множеству $X' : |X'| = d,\ d > 0$ поставим в соответствие множество $Y(X') = \{Y_1,\ldots,Y_k\}$, состоящее из $k = 2^d$ различных двоичных векторов длины d, каждый из которых является вектором значений переменных $x_{i_1},\ldots,x_{i_d}$. *Декомпозиционным семейством*, порожденным из КНФ C множеством X', называется множество $\Delta_{X'}(C)$ КНФ, полученных подстановками в C векторов $Y_j,\ j \in \{1,\ldots,k\}$:

$$\Delta_{X'}(C) = \{C_1 = C|_{Y_1},\ldots,C_k = C|_{Y_k}\},\ \Delta_{\varnothing}(C) = \{C\}.$$

Несложно видеть, что любой набор $\alpha \in \{0,1\}^n$, выполняющий C ($C|_{\alpha} = 1$), в компонентах, входящих в X', совпадает с некоторым вектором $Y^{\alpha} \in Y(X')$, а в остальных компонентах совпадает с набором, выполняющим КНФ

$C|_{Y^\alpha} \in \Delta_{X'}(C)$. КНФ C при этом невыполнима тогда и только тогда, когда все КНФ в $\Delta_{X'}(C)$ невыполнимы. Таким образом, SAT-задача относительно исходной КНФ C сводится к решению k SAT-задач для КНФ из множества $\Delta_{X'}(C)$. Для обработки множества $\Delta_{X'}(C)$ может быть использована РВС.

Предположим, что зафиксировано некоторое декомпозиционное множество $X', |X'| = d$. В силу сказанного выше, исходная SAT-задача сводится к решению 2^d SAT задач для КНФ из множества $\Delta_{X'}(C)$. Следующий вопрос состоит в том, можно ли в X' выбрать подмножество $X^{\sim}$, декомпозиция на основе которого была бы существенно эффективнее, чем на основе X'? В работе[2] была предложена специальная процедура статистического прогнозирования, предназначенная для улучшения декомпозиционного множества. Краткое описание данной процедуры приведено ниже.

Введем в рассмотрение натуральное число R_0, предназначением которого является разделение ситуаций – когда имеется необходимость формирования случайной выборки, и когда такой необходимости нет. Например, за R_0 можно принять число, близкое к числу процессоров в кластере. Если при некотором $X^{\sim} \subseteq X', |X^{\sim}| = d$ мощность семейства $\Delta_{X^{\sim}}(C)$ слишком велика, то представление о времени соответствующего параллельного вычисления можно составить на основе знания среднего времени решения SAT-задач для серии КНФ, выбранных случайным образом из $\Delta_{X^{\sim}}(C)$.

Через q_d обозначаем объем такой выборки. Через Y^d обозначается множество, образованное всеми различными векторами значений переменных из $X^{\sim} : |X^{\sim}| = d$. Каждому значению параметра $d \in \{0,1,\ldots,|X'|\}$ такому, что $2^d > R_0$, ставится в соответствие множество векторов $\{Y_{j_1},\ldots,Y_{j_{q_d}}\}$, выбираемых из Y^d в соответствии с равномерным распределением, а также множество (выборка) КНФ $\Theta_d = \{C_{j_1} = C|_{Y_{j_1}},\ldots,C_{j_{q_d}} = C|_{Y_{j_{q_d}}}\}$. Каждому значению параметра $d \in \{0,1,\ldots,|X'|\}$ такому, что $2^d \le R_0$, ставится в соответствие множество Y^d и

---

[2] Заикин О.С., Семенов А.А. Технология крупноблочного параллелизма в SAT-задачах // Проблемы управления. – 2008. – №1. С. 43-50.

множество КНФ $\Theta_d = \{C_1 = C|_{Y_1}, \ldots, C_{2^d} = C|_{Y_{2^d}}\}$ (в данном случае $\Theta_d = \Delta_d(C)$). Множество выборок $\{\Theta_d\}_{d \in \{0,1,\ldots,|X'|\}}$ далее обозначаем через Θ.

Фиксируем некоторый SAT-решатель S. Обозначим через $t(C')$ время работы (число битовых операций) SAT-решателя S на произвольном входе C'. Введем в рассмотрение функцию $\tau_S : \Theta \to \mathrm{N}$,

$$\tau_S(\Theta_d) = \sum_{C' \in \Theta_d} t(C').$$

Значением данной функции при каждом фиксированном $d \in \{0,1,\ldots,|X'|\}$ является суммарное время (число битовых операций) работы SAT-решателя S по всем КНФ из Θ_d. Следует учитывать, что при некоторых значениях параметра d (например, при $d = 0$) КНФ из Θ_d могут оказаться очень сложными для SAT-решателя, и в этом случае время подсчета соответствующего значения прогнозной функции может превысить разумные границы. Для учета данного факта вводится в рассмотрение специальная функция $g(C) = p(m \cdot n)$, здесь m – число дизъюнктов в КНФ C, а $p(\cdot)$ – некоторый полином, степень которого больше 1. Допустим, что в соответствии с перечисленными правилами построено семейство выборок $\Theta = \{\Theta_d\}_{d \in \{0,1,\ldots,|X'|\}}$ (при фиксированном R_0). *Прогнозную функцию* определим следующим образом.

$$T(\Theta_d) = \begin{cases} \dfrac{2^d}{q_d} \cdot \tau_S(\Theta_d),\ 2^d > R_0,\ \tau_S(\Theta_d) < g(C); \\ \tau_S(\Theta_d),\ 2^d \le R_0,\ \tau_S(\Theta_d) < g(C); \\ \\ \infty,\ \tau_S(\Theta_d) \ge g(C). \end{cases}$$

Запись «$T(\Theta_d) = \infty$» означает, что функция не определена на выборке Θ_d. Рациональное число $T(\Theta_d)$ является прогнозом общего объема битовых операций, требуемого для решения исходной SAT-задачи при декомпозиции КНФ C на семейство КНФ, порожденное множеством X^d. Тем самым задача прогнозного планирования оптимального по трудоемкости параллельного вычисления сводится к задаче минимизации функции T на множестве $domT \subseteq \Theta$.

Во второй главе диссертации приводится алгоритм типа «динамического программирования» полиномиальной трудоемкости, позволяющий решать описанную задачу минимизации для одного класса прогнозных функций.

Далее приведено описание общей процедуры крупноблочного распараллеливания SAT-задачи с последующим решением полученного семейства SAT-задач в РВС. Параметры декомпозиции исходной SAT-задачи определяются на основе знания глобального минимума прогнозной функции T на множестве $domT$.

Пусть $X_*^{\sim}$ ($X_*^{\sim} \subseteq X'$, $|X_*^{\sim}| = d^*$) – это такое множество, что $\left(d^*, T\left(\Theta_{d^*}\right)\right)$ – точка глобального минимума функции T на области ее определения. Пусть $\Delta^*(C) = \{C_1, \ldots, C_k\}$, $k = 2^{d^*}$, – декомпозиционное семейство КНФ, порожденное из C множеством $X_*^{\sim}$. Пусть имеется РВС, состоящая из r вычислительных узлов, где $r \in \mathrm{N}$. Возможны следующие два случая.

1) $k \le r$, то есть число КНФ в семействе $\Delta^*(C)$ не превосходит числа вычислительных узлов РВС. В этом случае для каждой КНФ из семейства $\Delta^*(C)$ SAT-задача решается на отдельном вычислительном узле РВС. На практике такая ситуация возникает весьма редко.

2) $k > r$ – число КНФ в семействе $\Delta^*(C)$ больше числа вычислительных узлов РВС. Именно такая ситуация наиболее типична для задач обращения криптографических функций из $\Im$. В данном случае каждому вектору $Y_j \in Y^{d^*}$, $j \in \{1, \ldots, k\}$, $k = 2^{d^*}$, ставится в соответствие натуральное число (назовем его *натуральным индексом* КНФ C_j), двоичным представлением которого является вектор Y_j. Пусть КНФ семейства $\Delta^*(C)$ упорядочены некоторым образом (например, по возрастанию их натуральных индексов). Произвольную КНФ из $\Delta^*(C)$ назовем *связанной*, если в рассматриваемый момент времени SAT-задача для нее либо уже решена, либо решается на некотором вычислительном узле РВС. Остальные КНФ называем *свободными*. Выбираются первые r КНФ $C_1, \ldots, C_r$ из семейства $\Delta^*(C)$. Для каждой из выбранных КНФ $C_1, \ldots, C_r$ решается SAT-задача на отдельном вычислительном узле РВС. Как только освобождается некоторый из r вычислительных узлов РВС, на нем запускается процедура решения SAT-задачи для первой (в смысле введенного выше порядка) свободной КНФ семейства $\Delta^*(C)$. Данный процесс продолжается

до тех пор, пока не будет найден выполняющий набор некоторой КНФ из $\Delta^*(C)$, либо пока не будет доказана невыполнимость всех КНФ из $\Delta^*(C)$. В силу сказанного выше данная процедура решает SAT-задачу для произвольной КНФ C корректно.

Отметим, что рассмотренная выше процедура использовала фиксированное некоторым (неопределенным) образом множество X'. Во второй главе диссертации предлагается общая методика построения декомпозиционных множеств. Основой данной методики является понятие схемы формирования декомпозиционного множества.

Схемой формирования семейства декомпозиционных множеств (далее «*схема формирования*») назовем семейство $H = \{h_0,\ldots,h_n\}$ отображений множества булевых переменных $X = \{x_1,\ldots,x_n\}$ в семейство множеств $\{X_H^0,\ldots,X_H^n\}$, где $X_H^d, d \in \{0,\ldots,n\}$, – декомпозиционное множество размерности d, $h_d : X \to X_H^d$, $d \in \{0,\ldots,n\}$. При этом для любого H полагается, что $X_H^0 = \varnothing$.

Одним из результатов второй главы диссертации является совмещение технологии прогнозных функций с использованием различных схем формирования. При этом каждой схеме формирования H ставится в соответствие оптимальное значение прогнозной функции T_H. После чего выбирается схема формирования и соответствующая декомпозиция, дающие наилучшие прогнозные значения трудоемкости параллельного вычисления.

В третьей главе приведено описание архитектуры и основных функций ППП D-SAT, созданного для реализации параллельной технологии, предложенной во второй главе. ППП D-SAT был использован в решении задач логического криптоанализа ряда генераторов, последовательный логический криптоанализ которых не дал приемлемых результатов.

ППП D-SAT функционирует в РВС под управлением инструментального комплекса DIStributed COmputing system of Modular Programming (DISCOMP), разработанного в ИДСТУ СО РАН. Библиотека программ ППП D-SAT включает *модуль расщепления*, *модуль SAT-решателя*, *модуль прогнозирования*, *модуль сравнения*, *аналитический модуль* и *транспортный модуль*.

Модуль расщепления используется для декомпозиции исходной КНФ. В данном модуле реализован ряд схем формирования. Основные входные данные модуля расщепления – это файл с исходной КНФ в формате DIMACS, номер $j \in \{1,\ldots,p\}$ схемы формирования, диапазон значений d_j (размерности декомпозиционного множества) и фиксированное значение q, определяющее

число КНФ в произвольной случайной выборке. Обозначим через $d_j^{\min}$ и $d_j^{\max}$ натуральные числа, определяющие соответственно левую и правую границы интервала, в котором изменяются значения d_j. Для каждого значения диапазона значений $d_j \in \left\{d_j^{\min},\ldots,d_j^{\max}\right\}$ строится отдельное семейство КНФ. В случае $q < 2^{d_j}$ из декомпозиционного семейства случайным образом выбираются q КНФ. Если $q = 0$ или $q \geq 2^{d_j}$, то выборкой является все рассматриваемое семейство. Результатом декомпозиции является набор файлов, содержащий КНФ, представленные в формате DIMACS. Данный набор описывается в виде параллельного списка. Вычислительное ядро модуля SAT-решателя составляет программа minisat[3] версии 2.0. По входному файлу, в котором КНФ представлена в формате DIMACS, модуль SAT-решателя осуществляет поиск выполняющего данную КНФ набора, либо выдает значение NULL в случае отсутствия такового. В выходной файл каждой копии модуля SAT-решателя записывается информация о результатах его работы. Модуль прогнозирования получает на вход файлы с результатами работы модуля SAT-решателя, значение q и число r вычислительных узлов РВС, относительно которого строится прогноз. Модуль строит прогноз оптимального (с точки зрения вычислительных затрат) значения d_j. Данное значение обозначается через d_j^*. Транспортный модуль осуществляет рассылку КНФ из параллельного списка. Модуль сравнения получает на вход прогнозные значения $d_1^*,\ldots,d_p^*$ и выдает значение $d^* \in \{d_1^*,\ldots,d_p^*\}$ с наилучшим прогнозом, а также соответствующий значению d^* индекс схемы формирования $j^* \in \{1,\ldots,p\}$. Аналитический модуль проверяет найденный выполняющий набор на корректность и сравнивает прогнозируемое время решения с реальным временем.

ППП D-SAT может функционировать в следующих режимах: *режим прогнозирования*; *режим решения SAT-задачи* и *режим прогнозирования с последующим решением SAT-задачи*.

Режим прогнозирования может быть использован для прогнозирования на персональных компьютерах (ПК) трудоемкости параллельного решения SAT-

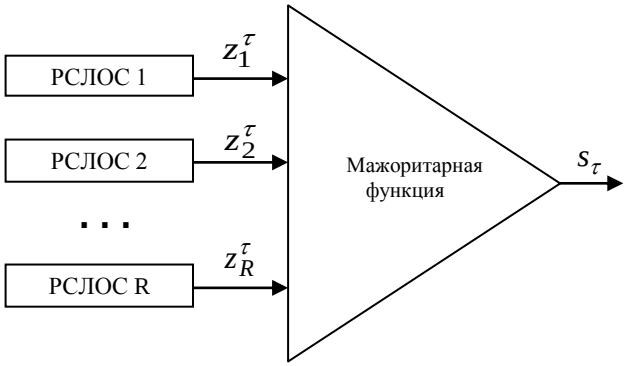
---

[3] http://minisat.se/MiniSat.html

задач на мощных вычислительных кластерах, доступ к которым ограничен. Если мощность ПК сопоставима с мощностью вычислительного узла кластера, то результаты экспериментов на ПК можно с учетом транспортных расходов экстраполировать на кластер.

С помощью ППП D-SAT был осуществлен успешный параллельный логический криптоанализ генератора Гиффорда, а также суммирующего и порогового генераторов (последовательный логический криптоанализ в отношении этих генераторов оказался малоэффективным). В качестве подтверждения эффективности предлагаемой в работе технологии приведем результаты параллельного логического криптоанализа порогового генератора.

В пороговом генераторе выходные биты $z_i^\tau, i \in \{1,\ldots,R\}$, сдвигаемых одновременно (в моменты времени $\tau \in N$) R регистров сдвига с линейной обратной связью[4] (РСЛОС) смешиваются посредством нелинейной булевой функции. В пороговом генераторе для этих целей используется мажоритарная функция.



$$s_\tau = \begin{cases} 1, \sigma_\tau > \dfrac{R}{2} \\ 0, \sigma_\tau \le \dfrac{R}{2} \end{cases}$$

$$\sigma_\tau = \sum_{i=1}^{R} z_i^\tau$$

Рассматривался вариант порогового генератора с 72-битной инициализирующей последовательностью на основе пяти РСЛОС, имеющих следующие порождающие многочлены: $(X^{11} + X^{10} + X^8 + X^3 + 1)$; $(X^{13} + X^9 + X^8 + X^2 + 1)$; $(X^{15} + X^{14} + X^{12} + X^2 + 1)$; $(X^{16} + X^{14} + X^8 + X^3 + 1)$;

---

[4] Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В. // Основы криптографии. – М.: Гелиос АРВ, 2002. – 480 с.

($X^{17}+X^{15}+X^{13}+X^{12}+X^{11}+X^{10}+1$). Анализировался фрагмент ключевого потока длиной 150 бит. В следующей таблице приведены результаты параллельного логического криптоанализа порогового генератора рассмотренного выше вида на вычислительном кластере Blackford ИДСТУ СО РАН[5] (40 четырехядерных процессоров).

Время решения на одном ядре кластера	Прогноз времени решения на кластере	Реальное время решения на кластере
> 2 суток	15 – 20 мин.	6 – 10 мин.

Таблица. Результаты параллельного логического криптоанализа 72-битного порогового генератора (серия тестов).

В заключении сформулированы основные результаты диссертационной работы:

1. Разработана крупноблочная параллельная технология решения SAT-задач. Исследованы декомпозиционные представления SAT-задач. Разработаны и проанализированы различные схемы формирования, ориентированные на структуру SAT-задач, кодирующих задачи криптоанализа некоторых генераторов ключевого потока.
2. Разработана процедура прогнозирования трудоемкости параллельного решения SAT-задач. Получена оценка сложности процедуры прогнозирования времени параллельного решения SAT-задач.
3. Разработан ППП D-SAT для практической реализации предложенной параллельной технологии решения SAT-задач. С помощью ППП D-SAT решены SAT-задачи, кодирующие логический криптоанализ генератора Гиффорда, а также суммирующего и порогового генераторов.

---

[5] http://www.mvs.icc.ru

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Заикин О.С., Семенов А.А. Технология крупноблочного параллелизма в SAT-задачах // Проблемы управления. – 2008. – № 1. – С. 43-50.**
2. **Семенов А.А., Заикин О.С., Беспалов Д.В., Ушаков А.А. SAT-подход в криптоанализе некоторых систем поточного шифрования // Вычислительные технологии. – 2008. Т. 13, № 6. – С. 133-149.**
3. Заикин О.С. Об одной эвристике в задаче поиска выполняющего набора выполнимой КНФ // Тезисы докладов VI школы-семинара молодых ученых «ММИТ». – Иркутск: ИДСТУ СО РАН, 2008. – С. 19.
4. Семенов А.А., Заикин О.С. Об одном подходе к поиску выполняющих наборов одновыполнимых КНФ // Материалы II Всероссийской конференции с международным участием «Инфокоммуникационные и вычислительные технологии и системы» (ИКВТС-06). – Улан-Удэ: БГУ, 2006. – Т. 2. – С. 122-126.
5. Заикин О.С. Параллельный подход к логическому криптоанализу некоторых генераторов двоичных последовательностей // Материалы конференции «Ляпуновские чтения». – Иркутск: ИДСТУ СО РАН, 2006. – С. 16.
6. Заикин О.С. Применение вычислительных кластеров в криптоанализе генераторов двоичных последовательностей. // Избранные материалы Шестой школы-семинара «Распределенные и Кластерные вычисления». Красноярск: ИВМ СО РАН, 2007. – С. 31-36.
7. Заикин О.С., Сидоров И.А. Технология крупноблочного распараллеливания в криптоанализе некоторых генераторов двоичных последовательностей // Труды международной научной конференции ПАВТ’07. Челябинск: ЮУрГУ, 2007. – Т. 1. – С. 158-169.
8. Семенов А.А., Заикин О.С. Технология крупноблочного параллелизма в SAT-задачах // Материалы XV международной конференции по вычислительной механике и современным прикладным программным системам (ВМСППС ’2007). – С. 230-231.
9. Заикин О.С., Семенов А.А., Сидоров И.А., Феоктистов А.Г. Параллельная технология решения SAT-задач с применением пакета прикладных программ D-SAT. // Вестник ТГУ. Приложение. – 2007. – № 23. – С. 83-95.
10. Заикин О.С. Реализация технологии решения SAT-задач с применением пакета прикладных программ D-SAT // Тезисы докладов IX школы-семинара молодых ученых «ММИТ». Иркутск: ИДСТУ СОРАН, 2007. – С. 73-76.
11. Заикин О.С. Использование пакета прикладных программ D-SAT для решения SAT-задач // Материалы конференции «Ляпуновские чтения». Иркутск: ИДСТУ СО РАН, 2007. – С. 9.

12. Семенов А.А., Заикин О.С. Неполные алгоритмы в крупноблочном параллелизме комбинаторных задач // Труды международной научной конференции ПАВТ’08. Санкт-Петербург, 2008. – С. 232-244.

13. Заикин О.С. Пакет прикладных программ Distributed-SAT: Свидетельство об официальной регистрации программы для ЭВМ № 2008610423. – М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2008.

14. Семенов А.А., Заикин О.С. Неполные алгоритмы в крупноблочном параллелизме комбинаторных задач // Вычислительные методы и программирование. – 2008. – Том 9, № 1. – С. 112-122.

15. Заикин О.С. Декомпозиционные представления данных в крупноблочном параллелизме SAT-задач // Прикладные алгоритмы в дискретном анализе. Иркутск: ИГУ, 2008. – Серия: Дискретный анализ и информатика, вып. 2. – С. 49-69.

16. Семенов А.А., Заикин О.С., Отпущенников И.В., Буров П.С. О некоторых особенностях задач обращения дискретных функций // Труды XIV Байкальской Международной школы-семинара «Методы оптимизации и их приложения». – Иркутск: ИСЭМ СО РАН, 2008. – Т. 1. – С. 498-505.

17. Семенов А.А., Заикин О.С., Беспалов Д.В., Хмельнов А.Г., Буров П.С. Анализ некоторых криптографических примитивов на вычислительных кластерах // Прикладная дискретная математика. – 2008. – № 2. – С. 120-130.

18. Семенов А.А., Заикин О.С., Беспалов Д.В., Хмельнов А.Г., Буров П.С. Решение задач обращения дискретных функций на многопроцессорных вычислительных системах // Труды Четвертой Международной конференции «Параллельные вычисления и задачи управления» PACO’2008. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2008. – С. 152-176.

Редакционно-издательский отдел
Института динамики систем и теории управления СО РАН
664033, Иркутск, ул. Лермонтова, д. 143
Подписано к печати 10.12.2008
Формат бумаги 60×84 1/16, объем 1,2 п.л.
Заказ 12. Тираж 100 экз.

Отпечатано в ИДСТУ СО РАН