На правах рукописи

Игнатьев Алексей Сергеевич

МЕТОДЫ ОБРАЩЕНИЯ ДИСКРЕТНЫХ ФУНКЦИЙ С ПРИМЕНЕНИЕМ ДВОИЧНЫХ РЕШАЮЩИХ ДИАГРАММ

05.13.18 – Математическое моделирование, численные методы
и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Иркутск – 2010

Работа выполнена в Учреждении Российской академии наук Институте динамики систем и теории управления Сибирского отделения РАН.

Научный руководитель:	кандидат технических наук, доцент Семенов Александр Анатольевич
Официальные оппоненты:	доктор физико-математичеких наук, профессор Перязев Николай Алексеевич
	кандидат физико-математических наук, доцент Панкратова Ирина Анатольевна
Ведущая организация:	Институт математики им. С. Л. Соболева СО РАН

Защита состоится 18 июня 2010 г. в 14:00 на заседании диссертационного совета Д 212.074.01 при ГОУ ВПО «Иркутский государственный университет» по адресу: 664003, г. Иркутск, бульвар Гагарина, 20, Институт математики, экономики и информатики.

С диссертацией можно ознакомиться в библиотеке Иркутского государственного университета (г. Иркутск, бульвар Гагарина, 24).

Автореферат разослан 17 мая 2010 г.

Ученый секретарь
диссертационного совета,
канд. физ.-мат. наук, доцент  Антоник В. Г.

Общая характеристика работы

Актуальность работы. В последние годы неуклонно растет интерес к дискретным моделям в различных областях информатики и кибернетики. Данный класс моделей чрезвычайно широк и включает модели безопасности компьютерных систем, модели процессов передачи и защиты информации, а также различные автоматные модели. К последним можно отнести автоматные сети, спектр применения которых варьируется от теории принятия решений до компьютерной биологии. Для численного исследования дискретных моделей далеко не всегда успешны «традиционные» методы, оперирующие с действительными числами. Во многих случаях получаемые такими методами результаты являются весьма грубыми приближениями и могут не удовлетворять требуемым критериям точности. Обширный класс дискретных моделей, тем не менее, допускает точные алгоритмы поиска решений. К данному классу относятся, в частности, модели, поведение которых может быть описано алгоритмически вычислимыми дискретными функциями, то есть функциями, преобразующими двоичные слова в двоичные слова. В задачах компьютерной безопасности и при исследовании автоматных моделей одной из наиболее часто возникающих является проблема обращения дискретной функции, вычислимой детерминированным образом за полиномиальное от длины входа время (то есть по известному алгоритму вычисления функции и известному образу требуется найти некоторый прообраз). В контексте данной постановки можно, например, рассматривать подавляющее большинство задач криптоанализа. Используя идеи С. А. Кука (изложенные им еще в 1971г.), можно строго доказать эффективную сводимость проблем обращения полиномиально вычислимых дискретных функций к задачам поиска решений логических (булевых) уравнений. Причем, в конечном счете, возможен эффективный переход к одному уравнению вида КНФ=1 (КНФ — конъюнктивная нормальная форма).

Задачи поиска решений логических уравнений дают пример проблем, чья аргументированная вычислительная сложность (в общей постановке они NP-трудны) не является препятствием для появления новых методов и алгоритмов. Об актуальности данной проблематики свидетельствует хотя бы факт издания в Нидерландах специализированного журнала «JSAT» (см. `http://jsat.ewi.tudelft.nl/`), подавляющее большинство статей в котором посвящено SAT-задачам (SAT-задачами называются задачи поиска решений логических уравнений вида КНФ=1).

На данный момент можно выделить (в качестве наиболее успешных) два общих подхода к поиску решений логических уравнений, высокая эффективность которых делает их широко используемыми. В первую очередь речь идет о SAT-подходе, в основе которого лежат процедуры приведения разнородных

систем логических уравнений к уравнениям вида «КНФ=1». Второй класс методов базируется на использовании двоичных решающих диаграмм (BDD), а точнее сокращенных упорядоченных BDD или «ROBDD» — формата представления булевых функций в виде направленных помеченных графов специального вида. О популярности «ROBDD-подхода» в задачах синтеза и верификации дискретных систем говорит тот факт, что на протяжении ряда лет ключевая статья по алгоритмике двоичных решающих диаграмм[1] лидировала в рейтинге цитируемости международной системы мониторинга научных публикаций «Citeseer» (см. `http://citeseer.ist.psu.edu/source.html`).

Преимуществом SAT-подхода является простота базовых структур данных и возможность их эффективного представления и оперирования с ними в памяти ЭВМ. Один из главных недостатков SAT-подхода состоит в фактической неполноте современных SAT-решателей, проявляющейся на аргументированно трудных тестах (например, на задачах криптоанализа ряда систем шифрования). Следует отметить, что наиболее быстрые (по результатам специализированных конкурсов) алгоритмы решения SAT-задач эксплуатируют идеологию «накопления ограничений» (подобно методам отсечений в целочисленном линейном программировании). Для хранения ограничений и быстрой работы с ними используется только оперативная память компьютера. При переполнении памяти возникает необходимость «чистки» баз накопленных ограничений, то есть удаления некоторых ограничений (дизъюнктов), признаваемых нерелевантными. Все известные практические оценки релевантности имеют характер эвристик. Данный факт не дает гарантии того, что алгоритм в дальнейшем не породит уже отброшенные ограничения. В такого рода ситуациях возможно зацикливание алгоритма (потеря полноты).

Основной недостаток ROBDD-подхода состоит в том, что даже в отношении простых в контексте SAT-подхода логических уравнений можно строго показать общую неэффективность в применении к ним ROBDD-подхода. Главное преимущество ROBDD-подхода состоит в том, что произвольная ROBDD единственным образом (с точностью до изоморфизма графов) представляет соответствующую булеву функцию. Тем самым, ROBDD можно рассматривать как некоторую компактную форму представления булевых функций в специальном классе графов.

Исходя из всего вышесказанного, актуальными представляются проблемы разработки, обоснования эффективности и программной реализации методов решения логических уравнений, в которых сочетались бы преимущества скорости обработки данных, присущей SAT-подходу, и компактности представления данных, присущей ROBDD-подходу. Решатели, базирующиеся на таких методах, могут использоваться при исследовании широкого класса дис-

[1] Bryant R. E. Graph-Based Algorithms for Boolean Function Manipulation // IEEE Transactions on Computers. 1986. Vol. 35, no 8. Pp. 677–691.

кретных моделей, поведение которых допускает описание полиномиально вычислимыми дискретными функциями.

Цель и задачи исследования. Целью диссертационной работы является разработка и практическая реализация гибридного (SAT+ROBDD)-метода решения систем логических уравнений, кодирующих проблемы обращения полиномиально вычислимых дискретных функций.

Для достижения указанной цели ставятся и решаются перечисленные ниже задачи.

1. Разработать стратегию логического вывода, в которой SAT-подход сочетается с ROBDD-подходом в следующем смысле: DPLL (как базовый алгоритм решения SAT) порождает массивы ограничений-дизъюнктов, которые в дальнейшем хранятся и обрабатываются в форме ROBDD-представлений соответствующих булевых функций (данный шаг предполагает сокращение объема оперативной памяти, используемой для хранения накопленных ограничений); разработать и реализовать ROBDD-аналоги основных механизмов вывода, используемых в современных SAT-решателях; строго обосновать корректность и эффективность работы соответствующих процедур.

2. Разработать и программно реализовать ROBDD-решатель систем логических уравнений, использующий новые эвристические алгоритмы.

3. Программно реализовать гибридный (SAT+ROBDD)-решатель логических уравнений, ориентированный на задачи обращения полиномиально вычислимых дискретных функций; разработать параллельные гибридные (SAT+ROBDD)-алгоритмы решения логических уравнений и обращения дискретных функций; программно реализовать разработанные алгоритмы с использованием стандарта MPI (Message Passing Interface); интегрировать все разработанные алгоритмы в программный комплекс.

4. Протестировать построенный программный комплекс на задачах обращения криптографических функций и задачах исследования некоторых автоматных моделей.

Методы и инструменты исследования. Теоретическая часть исследования использует аппарат теории множеств, дискретной математики, теории вычислительной сложности, теории булевых функций, теории параллельных вычислений, а экспериментальная — современные средства разработки программного обеспечения, а также многопроцессорные вычислительные системы.

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

- эвристические алгоритмы решения систем логических уравнений при помощи двоичных решающих диаграмм, использующие декомпозиции исходной системы;
- ROBDD-аналоги базовых алгоритмов и процедур, используемых в современных DPLL-решателях, а также новый алгоритм модификации порядка означивания переменных в ROBDD;
- программный комплекс, включающий ROBDD-решатель систем логических уравнений, а также параллельную и последовательную версии гибридного (SAT+ROBDD)-решателя;
- вычислительные эксперименты, включающие решение задач обращения некоторых криптографических функций и исследование автоматных моделей генных сетей (построенный программный комплекс на тестовых задачах превзошел по эффективности сторонние разработки).

Основные результаты, выносимые на защиту.

1. Метод обращения полиномиально вычислимых дискретных функций, в основе которого лежит гибридный (SAT+ROBDD) логический вывод; строгое обоснование (в форме теорем) математических свойств ROBDD, рассматриваемых в роли баз булевых ограничений; ROBDD-аналоги основных процедур, используемых в нехронологическом DPLL-выводе (правило единичного дизъюнкта, процедура «Clause Learning», процедуры работы с дизъюнктами, использующие идеологию «отсроченных вычислений»).
2. Новые эвристические алгоритмы решения систем логических уравнений, использующие двоичные решающие диаграммы (ROBDD).
3. Программный комплекс, представляющий собой новый гибридный (SAT+ROBDD)-решатель, ориентированный на решение задач обращения дискретных функций и функционирующий в распределенных вычислительных средах (PBC).
4. Результаты численных экспериментов, включающие исследование дискретно-автоматных моделей из компьютерной биологии, а также решение задач обращения некоторых криптографических функций в PBC.

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

Соответствие специальности. В диссертации разработан новый вычислительный метод, использующий алгоритмы обработки дискретных данных и применимый к широкому спектру практических задач (тестирование и верификация дискретных управляющих систем, исследование различных дискретно-автоматных моделей, обращение дискретных функций, криптоанализ). Разработанный метод реализован в виде программного комплекса, функционирующего в распределенных вычислительных средах. Комплекс протестирован на аргументированно трудных задачах обращения некоторых криптографических функций.

Теоретическая и практическая значимость работы. Теоретическая значимость работы заключена в возможности применения предложенных методов к исследованию алгоритмических аспектов задач обращения дискретных функций. Практическая значимость состоит в возможности использовать предложенные методы и применяемые в них вычислительные алгоритмы в исследовании широкого класса моделей дискретных систем, поведение которых описывается полиномиально вычислимыми дискретными функциями.

Апробация работы. Результаты диссертации докладывались и обсуждались на 3-ей Международной научной конференции «Параллельные вычислительные технологии» (Нижний Новгород, 2009 г.); на VII Всероссийской конференции с международным участием «Новые информационные технологии в исследовании сложных структур» (Томск, 2008 г.); на Всероссийской школе-семинаре с международным участием Sibecrypt-09 (Омск, 2009 г.); на VIII и на IX школах-семинарах «Математическое моделирование и информационные технологии» (Иркутск, 2006, 2007 гг.), на ежегодных конференциях из серии «Ляпуновские чтения» (Иркутск, 2006, 2007, 2008, 2009 гг.), а также на научных семинарах Института динамики систем и теории управления СО РАН, научных семинарах кафедры математической информатики Восточно-Сибирской государственной академии образования; научном семинаре лаборатории дискретного анализа Института математики им. С. Л. Соболева СО РАН; научном семинаре кафедры защиты информации и криптографии Томского государственного университета.

Результаты диссертации были получены в процессе исследований по следующим проектам:

- проект СО РАН «Интеллектные методы и инструментальные средства

создания и анализа интегрированных распределенных информационно-аналитических и вычислительных систем для междисциплинарных исследований с применением ГИС, GRID и Веб-технологий» 2007–2009 гг.;

- грант РФФИ №07-01-00400-а «Характеризация сложности обращения дискретных функций в задачах криптографии и интервального анализа»;

- грант Президента РФ НШ-1676.2008.1.

Публикации и личный вклад автора. По теме диссертации опубликовано 14 работ. Наиболее значимые результаты представлены в работах [1–7]. В число указанных работ входят 2 статьи [1, 2] из Перечня ведущих рецензируемых журналов и изданий ВАК РФ (2010 г.), 3 статьи [3–5] в научных журналах, 2 полных текста докладов [6, 7] в материалах международных конференций.

Результаты, относящиеся к разделу 2.3, получены совместно с научным руководителем Семеновым А. А. и являются неделимыми. Из совместных работ с Беспаловым Д. В., Заикиным О. С., Хмельновым А. Е. в диссертацию включены результаты, принадлежащие лично автору.

Структура работы. Диссертация состоит из введения, трех глав, заключения и списка литературы, из 111 наименований. Объем диссертации — 110 страниц, включая 27 рисунков и 7 таблиц.

Содержание работы

Первая глава является обзорной и содержит теоретическую базу для последующего материала. В данной главе приведены необходимые сведения из теории дискретных функций, кратко описаны основные алгоритмы решения SAT-задач. Заключительный раздел первой главы посвящен основам теории двоичных решающих диаграмм и их применению к логическим уравнениям и задачам обращения дискретных функций.

В этой же главе сформулирована основная проблема, исследуемая в диссертации и состоящая в следующем. Рассматривается натуральное семейство, образованное всюду определенными дискретными функциями вида $f_n : \{0,1\}^n \to \{0,1\}^*$, вычислимыми за полиномиальное от n время. Проблемой обращения произвольной функции f_n из данного семейства в точке y называется следующая задача: по известному $y \in range\ f_n$ и известному алгоритму вычисления f_n требуется найти такой $x \in \{0,1\}^n$, что $f_n(x) = y$. Данная проблема может быть эффективно сведена к задаче поиска выполняющего набора КНФ $C(f)$ над множеством булевых переменных $X = \{x_1, \ldots, x_{p(n)}\}$, $p(\cdot)$ — некоторый полином.

Во **второй главе** развивается гибридный (SAT+ROBDD)-подход к решению проблемы обращения полиномиально вычислимых дискретных функций. Теоретические результаты данной главы дают основу для разработки и программной реализации гибридного (SAT+ROBDD)-решателя.

В **разделе 2.1** детально описана ROBDD-алгоритмика работы с системами логических уравнений произвольного вида. Здесь же приведено семейство новых эвристик и механизмов, позволяющих повысить эффективность процедур построения ROBDD-представлений характеристических функций систем логических уравнений. Перечисленные компоненты составляют основу архитектуры ROBDD-решателя логических уравнений, используемого в гибридном (SAT+ROBDD)-подходе к обращению дискретных функций.

Рассматривается произвольная система (кратко обозначаемая через S) логических уравнений вида

$$S : \begin{cases} L_1(x_1, \ldots, x_n) & = 1, \\ \cdots & \\ L_m(x_1, \ldots, x_n) & = 1. \end{cases}$$

Здесь $L_j(x_1, \ldots, x_n)$, $j \in \{1, \ldots, m\}$, — формулы исчисления высказываний (ИВ), $X = \{x_1, \ldots, x_n\}$ — множество булевых переменных. Характеристической функцией системы S называется булева функция $\delta_S : \{0,1\}^n \to \{0,1\}$, заданная формулой

$$L_1(x_1, \ldots, x_n) \cdot \ldots \cdot L_m(x_1, \ldots, x_n).$$

Несложно понять, что по ROBDD-представлению $B(\delta_S)$ функции δ_S проблема совместности системы S, а также проблема поиска некоторого ее решения могут быть решены за линейное от $|B(\delta_S)|$ время ($|B(\delta_S)|$ — число вершин в $B(\delta_S)$). При построении $B(\delta_S)$ можно использовать различные эвристики организации порядка означивания переменных, а также строить декомпозиции исходной системы с целью разбиения процесса построения $B(\delta_S)$ на независимые этапы. В диссертации для этих целей предложена новая эвристическая процедура разбиения системы S на подсистемы, называемые слоями. Пусть выбран некоторый порядок означивания переменных из $X = \{x_1, \ldots, x_n\}$, $\tau : x_{i_1} \prec x_{i_2} \prec \ldots \prec x_{i_{n-1}} \prec x_{i_n}$, в соответствии с которым предполагается строить $B(\delta_S)$. Число $r \in \{1, \ldots, n\}$ называем индексом переменной x_{i_r} относительно выбранного порядка. Используя порядок τ, разбиваем систему S на слои. Первый слой образован всеми уравнениями системы, в которые входит переменная x_{i_1}. Второй слой образован всеми оставшимися уравнениями, содержащими переменную x_{i_r}, которая имеет наименьший относительно τ индекс по всем уравнениям, не входящим в первый слой. И так далее. Пусть R — число определяемых описанным образом слоев системы.

Очевидно, что $1 \leq R \leq n$. ROBDD каждого слоя B_j, $j \in \{1, \ldots, R\}$, строится в соответствии с порядком τ по следующей рекурсивной схеме, использующей алгоритм «Apply» Р. Брайанта:

$$B_j = Apply\left(B_{j_1} \cdot Apply\left(B_{j_2} \cdot \ldots \cdot Apply\left(B_{j_{k-1}} \cdot B_{j_k}\right)\right)\right),$$

где $B_{j_1}, \ldots, B_{j_k}$ — ROBDD булевых функций, выраженных формулами в левых частях уравнений системы, образующих j-тый слой. Итоговая ROBDD характеристической функции системы строится в соответствии с порядком τ по аналогичной рекурсивной схеме:

$$B = Apply\left(B_1 \cdot Apply\left(B_2 \cdot Apply\left(B_3 \cdot \ldots \cdot Apply\left(B_{R-1} \cdot B_R\right)\right)\right)\right).$$

Представленная техника показала высокую эффективность на системах логических уравнений, описывающих поведение некоторых классов дискретных автоматов, используемых в компьютерной биологии[2] (результаты численных экспериментов приведены в третьей главе).

В **разделе 2.2** рассмотрена проблема модификации ROBDD в соответствии с новым порядком означивания переменных. Исследован описанный в литературе подход к этой проблеме, использующий идеологию «пузырьковой сортировки [3]». Предложен новый алгоритм решения данной проблемы, приведена оценка его трудоемкости и обоснование его большей эффективности в сравнении с известными подходами.

Пусть дана произвольная ROBDD $B(f)$, представляющая булеву функцию $f : \{0,1\}^n \to \{0,1\}$ и построенная в соответствии с порядком означивания переменных $\tau : x_1 \prec x_2 \prec \ldots \prec x_n$ (корень ROBDD помечен переменной x_1). Рассмотрим произвольную подстановку на множестве $\{1, \ldots, n\}$

$$\sigma\left(\tau \to \tau'\right) = \begin{pmatrix} 1 & 2 & \ldots & n \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{pmatrix},$$

$\{\alpha_1, \ldots, \alpha_n\} = \{1, \ldots, n\}$. Будем говорить, что данная подстановка задает изменение исходного порядка τ на порядок τ', подразумевая, что столбец подстановки с номером i, $i \in \{1, \ldots, n\}$, имеющий вид $\begin{pmatrix} i \\ j \end{pmatrix}, j \in \{1, \ldots, n\}$, интерпретирует факт нахождения переменной x_j в новом порядке τ' на позиции с номером i.

---

[2] Системная компьютерная биология. Под ред. Н. А. Колчанова, С. С. Гончарова, В. А. Лихошвая, В. А. Иванисенко. Новосибирск: Изд-во СО РАН, 2008. С. 767.

[3] Meinel Ch., Theobald T. Algorithms and Data Structures in VLSI-Design: OBDD-Foundations and Applications. Springer-Verlag, 1998; Колпаков А. В., Латыпов Р. Х. Приближенные алгоритмы минимизации двоичных диаграмм решений на основе линейных преобразований переменных // Автоматика и телемеханика. 2004. Т. 6. С. 112–128.

Определение. *Дана ROBDD $B(f)$, представляющая булеву функцию $f : \{0,1\}^n \to \{0,1\}$, построенная в соответствии с заданным порядком τ : $x_1 \prec x_2 \prec \ldots \prec x_n$. Требуется построить ROBDD $B'(f)$, представляющую ту же самую функцию, в которой порядок означивания переменных из X есть τ', отличный от τ. Назовем данную проблему проблемой модификации ROBDD в соответствии с новым порядком. Ее частным случаем является проблема перестройки $B(f)$ в соответствии с произвольным порядком τ' таким, что i-тый столбец подстановки $\sigma\,(\tau \to \tau')$ имеет вид $\binom{i}{j}$, $j \neq i$. Эту проблему называем проблемой установки переменной x_j на позицию с номером i.*

Доказана следующая теорема (везде далее нумерация теорем соответствует их нумерации в тексте диссертации).

Теорема 2.1. *Пусть ROBDD $B(f)$ представляет булеву функцию f от n переменных в соответствии с порядком τ : $x_1 \prec x_2 \prec \ldots \prec x_n$. Тогда для произвольных i, $j \in \{1, \ldots, n\}$ проблема установки переменной x_j на позицию с номером i решается детерминированным образом за время $O\left(|B(f)|^2\right)$.*

Следствие теоремы 2.1. *Пусть ROBDD $B(f)$ представляет булеву функцию f от n переменных в соответствии с порядком τ : $x_1 \prec x_2 \prec \prec \ldots \prec x_n$. Тогда проблема модификации $B(f)$ в соответствии с произвольным порядком τ' сводится к l-кратному ($l \leq n$) решению проблемы установки переменной на позицию с заданным номером.*

Доказательство данного факта дает алгоритм модификации порядка в ROBDD, применяемый в дальнейшем в гибридном (SAT+ROBDD)-решателе.

В **разделе 2.3** кратко описывается общая схема гибридного (SAT+ROBDD)-подхода к решению задач обращения полиномиально вычислимых дискретных функций. В основе данного подхода лежит понятие ядра DPLL-вывода и возможности декомпозиционного разбиения решаемой задачи обращения на SAT- и ROBDD-части. Последующий процесс решения — это сочетание нехронологического DPLL-вывода на SAT-части с выводом на ROBDD. Основные результаты данного раздела получены в неделимом соавторстве с научным руководителем диссертанта и приведены без доказательств.

Раздел 2.4 содержит описание и оценки трудоемкости новых алгоритмов работы с ROBDD как с модифицируемыми базами булевых ограничений, которые накапливаются в процессе нехронологического DPLL-вывода.

Пусть $B(f)$ — ROBDD-представление произвольной булевой функции f от переменных $x_1, \ldots, x_n$. Каждой переменной x_i, $i \in \{1, \ldots, n\}$, и терминальным вершинам «0», «1» поставим в соответствие множества значений данной переменной, задаваемых всевозможными путями в $B(f)$ из корня в соответствующую терминальную вершину. Данные множества обозначим че-

рез $\Delta^0(x_i)$, $\Delta^1(x_i)$.

Предположим, что в некоторой ROBDD $B(f)$ выполнены перечисленные ниже условия.

1) Для некоторой переменной $x_k \in X$, $X = \{x_1, \ldots, x_n\}$, любой путь π из корня $B(f)$ в терминальную «1» обязательно проходит через некоторую вершину, помеченную переменной x_k.

2) Справедливо $\left|\Delta^1(x_k)\right| = 1$.

Результатом данной ситуации является заключение о том, что в любом наборе значений истинности переменных из множества X, на котором значение функции f, представленной $B(f)$, равно 1, переменная x_k может принимать только одно значение (соответствующее значение в $\Delta^1(x_k)$).

Определение. *Определяемую условиями 1)-2) ситуацию далее называем ROBDD-следствием соответствующего значения для переменной x_k.*

Лемма 2.1. *Выполнимость 1)-2) относительно некоторой переменной x_k означает, что выполнено лишь одно из следующих условий:*

1. *для каждой вершины, помеченной x_k, ее h-ребенком является терминальная вершина «0»;*

2. *для каждой вершины, помеченной x_k, ее l-ребенком[4] является терминальная вершина «0».*

Лемма 2.2. *Процедура проверки возникновения ROBDD-следствий в ROBDD $B(f)$ требует детерминированного времени, ограниченного сверху величиной $O\left(n \cdot |B(f)|\right)$.*

Далее изучается проблема отслеживания ROBDD-следствий в результате подстановок в ROBDD конкретных значений некоторых переменных. Такого рода подстановки осуществляются при помощи известной процедуры «Restrict», описанной Р. Брайантом. Предложен новый алгоритм, проверяющий возникновение ROBDD-следствий как результатов подстановок в ROBDD значений булевых переменных. С использованием лемм 2.1–2.2 доказана следующая теорема.

Теорема 2.5. *Пусть в ROBDD $B(f)$ подставляются значения переменных*

$$x_{i_1} = \alpha_{i_1}, \ldots, x_{i_m} = \alpha_{i_m}, m \leq n, \alpha_{i_j} \in \{0,1\}, j \in \{1, \ldots, m\}.$$

Тогда сложность детерминированной процедуры, осуществляющей данную подстановку и проверяющей наличие всевозможных ROBDD-следствий, ограничена сверху величиной $O\left(n \cdot |B(f)|\right)$.

[4] От «high-child» и «low-child» в английской нотации.

Процедура, о которой идет речь в теореме 2.5, получила название *assign()*. Справедливо следующее весьма важное в практическом отношении свойство.

Теорема 2.6. *Если результатом применения процедуры assign() к* $B(f)$ *является ROBDD-следствие* $x_k = \alpha_k$, $\alpha_k \in \{0, 1\}$, *для некоторой* $x_k \in X$, *то подстановка в* $B(f)$ «$x_k = \alpha_k$» *не может привести к возникновению нового ROBDD-следствия, индуцированного данной подстановкой.*

Данный факт демонстрирует очень привлекательное свойство ROBDD, рассматриваемой в роли базы булевых ограничений. Напомним, что подстановка значения некоторой переменной в КНФ может приводить к выводу по правилу единичного дизъюнкта (unit clause) ряда индуцированных присвоений, подстановка которых также не исключает дальнейших срабатываний unit clause и т. д. В этом смысл стратегии распространения булевых ограничений (BCP). В общем случае полная реализация BCP может приводить к многократному обходу КНФ, что сопряжено с существенными вычислительными затратами. Полученное свойство ROBDD означает, что порождаемые произвольной подстановкой ROBDD-следствия сами по себе новых ROBDD-следствий породить не могут и, таким образом, вся информация, индуцируемая данной подстановкой, извлекается в результате однократного обхода ROBDD.

Еще одним полезным свойством гибридного вывода является возможность естественной организации на ROBDD т. н. «отсроченных вычислений». Рассмотрим следующие условия, которые определяют ситуацию, в некотором роде двойственную ситуации возникновения ROBDD-следствия.

i) Для некоторой переменной $x_q \in X$ в ROBDD $B(f)$ любой путь π из корня в терминальную «0» обязательно проходит через некоторую вершину, помеченную переменной x_q.

ii) Справедливо $\left|\Delta^0(x_q)\right| = 1$.

Устанавливается справедливость следующей теоремы.

Теорема 2.7. *Пусть* $B(f)$ *— произвольная ROBDD и относительно некоторой переменной* x_q *в* $B(f)$ *справедливы условия i) и ii). Тогда в ROBDD* $B(f)$ *невозможны ROBDD-следствия ни для каких переменных из множества* $X \setminus \{x_q\}$. *Трудоемкость процедуры проверки условий i)–ii) ограничена сверху величиной* $O\left(n \cdot |B(f)|\right)$.

Данный результат позволяет сформировать механизмы отсроченных вычислений при подстановке выведенных значений некоторых переменных: если для текущей ROBDD $B(f)$ выполнены i)–ii) относительно x_q, и из КНФ-части выведено значение некоторой переменной x_k, $k \neq q$, нет смысла на данном этапе подставлять соответствующее значение в $B(f)$ — ничего нового выведено не будет. После присвоения или вывода из КНФ-части некоторого значения

для x_q целесообразно осуществить в $B(f)$ подстановку сразу всех накопленных к этому моменту значений переменных, а также вывод всех возможных ROBDD-следствий, используя для этого процедуру *assign()*.

Приведем краткое резюме основных результатов второй главы, касающихся алгоритмики гибридного (SAT+ROBDD)-подхода к задачам обращения полиномиально вычислимых дискретных функций.

Некоторый алгоритм на основе DPLL действует в отношении КНФ $C(f)$, кодирующей задачу обращения функции f в некоторой точке. После первого рестарта вместо чистки базы конфликтных дизъюнктов, имеющей вид КНФ C', строится ROBDD-представление булевой функции, заданной формулой C'. Дальнейший вывод идет как на исходной КНФ, так и на ROBDD, представляющей базу накопленных ограничений. При этом на ROBDD действуют аналоги механизмов вывода, используемых в современных быстрых SAT-решателях на базе DPLL: аналог подстановки и правила единичного дизъюнкта реализован в виде процедуры *assign()*, описанной при доказательстве теоремы 2.5; аналогом CL-процедуры является применение алгоритма Р. Брайанта «Apply» к текущей ROBDD и новому ограничению-дизъюнкту; отсроченные вычисления (аналог структуры «watched literals») организуются с учетом условий i)–ii) и результата теоремы 2.7.

Третья глава посвящена программной реализации гибридного (SAT+ +ROBDD)-подхода к обращению полиномиально вычислимых дискретных функций и построению программного комплекса, функционирующего в распределенных вычислительных средах (РВС).

В **разделе 3.1** описывается архитектура ROBDD-решателя систем логических уравнений, в котором используются предложенные во второй главе эвристические алгоритмы «реорганизации» рассматриваемых систем посредством разбиения их на слои, а также специальный менеджер памяти, значительно повышающий эффективность процедур работы с оперативной памятью ЭВМ при построении ROBDD. Реализованный в соответствии с описанными принципами ROBDD-решатель был протестирован на задачах исследования некоторых дискретных автоматов, используемых в компьютерной биологии. А именно, были решены задачи поиска неподвижных точек автоматных отображений, которые определяются графами, задающими регуляторные контуры дискретных моделей генных сетей[5] (удавалось успешно решить соответствующие задачи для графов на 100 вершинах). Следует отметить, что эвристика разбиения на слои приводила к увеличению в разы эффективности процесса построения ROBDD-представлений характеристических функций рассматриваемых систем логических уравнений.

[5] Григоренко Е. Д., Евдокимов А. А., Лихошвай В. А., Лобарева И. А. Неподвижные точки и циклы автоматных отображений, моделирующих функционирование генных сетей // Вестник Томского гос. ун-та. Приложение. 2005. Т. 14. С. 206–212.

Раздел 3.2 целиком посвящен описанию программного комплекса, в котором реализованы все описанные в работе алгоритмы. Данный программный комплекс представляет собой гибридный (SAT+ROBDD)-решатель, ориентированный на решение в РВС задач обращения полиномиально вычислимых дискретных функций. Главной мотивацией (SAT+ROBDD)-подхода к обращению таких функций стало наблюдение о высокой степени «ROBDD-сжатия» массивов конфликтных дизъюнктов, накапливаемых в процессе нехронологического DPLL-вывода на КНФ, кодирующих соответствующие задачи обращения. Это наблюдение стало результатом большого числа вычислительных экспериментов. В задачах обращения некоторых криптографических функций массивы конфликтных дизъюнктов объемами в сотни мегабайт «сжимались» (при помощи описанного выше ROBDD-решателя) в ROBDD, состоящие из десятков вершин и занимающих в памяти ЭВМ несколько килобайт. Такой малый размер ROBDD-представлений баз конфликтных ограничений дает возможность эффективно обмениваться ими в РВС. Именно этот подход был реализован в построенном параллельном (SAT+ROBDD)-решателе.

Основу параллельного решателя составляет последовательный гибридный (SAT+ROBDD)-решатель, названный «hsat». Данный решатель функционирует в соответствии со схемой, представленной на рисунке 1 (логический вывод ведется как на исходной КНФ, так и на ROBDD, представляющей базу накопленных ограничений-дизъюнктов).

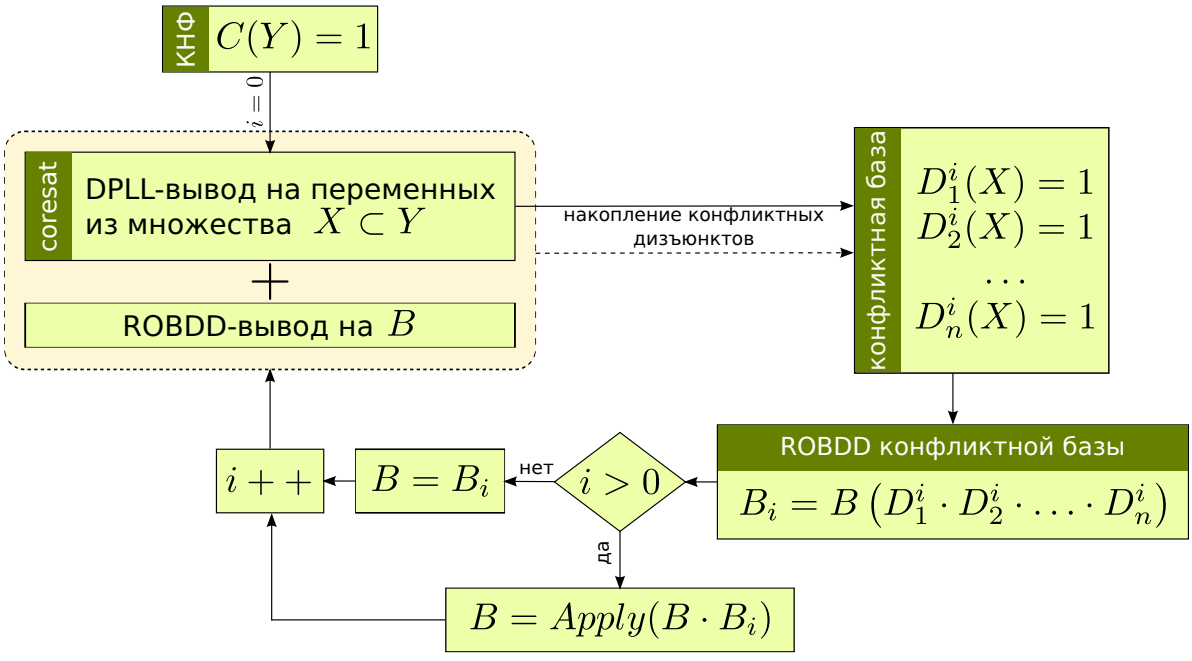

Рис. 1. Схема работы гибридного (SAT+ROBDD)-решателя hsat

Параллельный гибридный (SAT+ROBDD)-решатель получил название «mhsat». Далее описаны основные принципы его работы. На k вычислитель-

ных ядрах запускается k версий решателя hsat, получающих на входе, вообще говоря, произвольную КНФ C. Все версии hsat стартуют с различными начальными порядками угадывания переменных. Процесс вывода является итерационным. Каждая итерация разбивается на два этапа. На первом этапе все ядра работают независимо и на каждом происходит накопление конфликтных дизъюнктов, причем база конфликтных ограничений имеет вид ROBDD (при этом используются все описанные выше механизмы гибридного (SAT+ROBDD)-вывода). На втором этапе все ядра обмениваются накопленными ограничениями (при этом возникает необходимость изменения в некоторых ROBDD порядка означивания переменных). Обмен происходит в соответствии со схемой, представленной на рисунке 2. После обмена ограничениями решатель снова переходит в режим независимой работы ядер (следующая итерация). Работа продолжается до момента решения соответствующей SAT-задачи на некотором ядре.

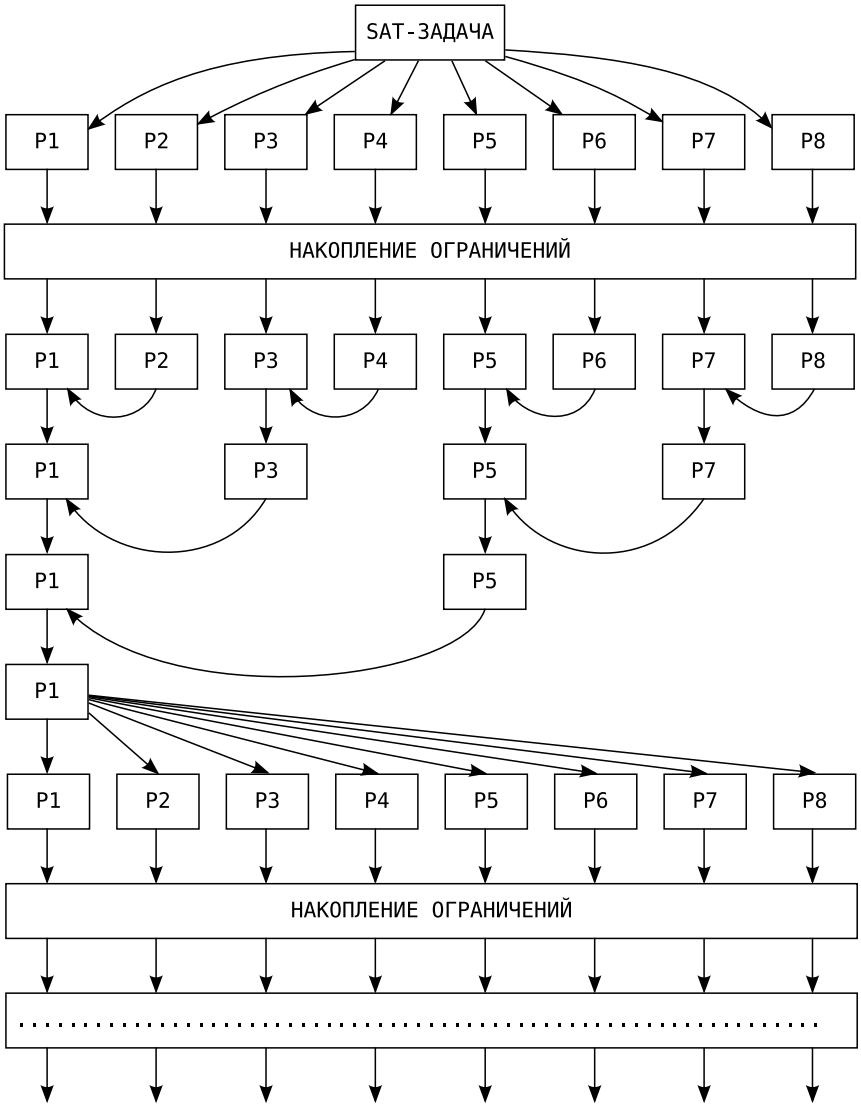

Рис. 2. Схема обмена ограничениями в решателе mhsat (рассмотрен случай 8 ядер: P1–P8)

Решатель mhsat был реализован с использованием стандарта MPI и протестирован на задаче обращения дискретной функции, задающей порождение ключевого потока в системе шифрования A5/1. Следует отметить, что задача поиска выполняющего набора КНФ $C(\text{A5/1})$, непосредственно кодирующей криптоанализ A5/1, является очень сложной. Поэтому были рассмотрены ослабления этой задачи, а именно, рассматривались КНФ, из декомпозиционного семейства $\Delta_d = \{C_1^*(\text{A5/1}), \ldots, C_{2^d}^*(\text{A5/1})\}$, полученного в результате подстановок в $C(\text{A5/1})$ всевозможных значений некоторых d булевых переменных, выбираемых специальным образом. Именно такой подход используется при крупноблочном распараллеливании[6] SAT-задач, кодирующих задачи криптоанализа различных шифров (в том числе и A5/1).

В численных экспериментах рассматривались КНФ из декомпозиционного семейства $\Delta_{20} = \{C_1^*(\text{A5/1}), \ldots, C_{2^{20}}^*(\text{A5/1})\}$ ($d = 20$). Решатель mhsat запускался на 4 ядрах процессора Intel® Xeon® E5345 с тактовой частотой 2,33 ГГц. Проводилось сравнение по эффективности mhsat, известного параллельного решателя MiraXT[7], а также решателя dminisat[6]. Было сгенерировано 50 тестов (КНФ, выбираемые случайным образом из семейства Δ_{20}). В среднем mhsat по эффективности на данном наборе тестов превзошел MiraXT в 2,88 раза, а dminisat — в 5,12 раза.

В **заключении** сформулированы основные результаты диссертационной работы.

1. Разработан новый вычислительный метод решения задач обращения полиномиально вычислимых функций, базирующийся на гибридном (SAT+ROBDD) логическом выводе; основу метода составили новые алгоритмы логического вывода на ROBDD и ROBDD-аналоги основных механизмов, применяемых в нехронологическом DPLL-выводе; базовые свойства всех алгоритмов обоснованы в форме теорем, построены оценки их вычислительной трудоемкости.

2. Разработан ROBDD-решатель систем логических уравнений, использующий оригинальные эвристические алгоритмы разбиения рассматриваемых систем на слои и эффективные процедуры управления оперативной памятью ЭВМ. Решатель показал высокую эффективность на задачах исследования некоторых автоматных моделей генных сетей.

---

[6] Семенов А. А., Заикин О. С., Беспалов Д. В., Буров П. С., Хмельнов А. Е. Решение задач обращения дискретных функций на многопроцессорных вычислительных системах // Труды Четвертой Международной конференции «Параллельные вычисления и задачи управления» (PACO'2008). Москва: 26–29 октября 2008. С. 152–176.

[7] Schubert T., Lewis M., Becker B. PaMiraXT: Parallel SAT Solving with Threads and Message Passing // Journal on Satisfiability, Boolean Modeling and Computation. Special Issue on Parallel SAT Solving. 2009. Vol. 6. Pp. 203–222.

3. Разработан программный комплекс, реализующий метод гибридного (SAT+ROBDD)-вывода и ориентированный на решение задач обращения полиномиально вычислимых дискретных функций. Данный программный комплекс реализован с использованием стандарта MPI и предназначен для работы в распределенных вычислительных средах. Принципиальная новизна комплекса состоит возможности эффективного обмена в распределенных средах булевыми ограничениями, представляемыми в виде ROBDD.

4. Проведено тестирование построенного программного комплекса на аргументированно трудных задачах обращения некоторых криптографических функций. Эффективность комплекса на рассмотренных классах тестов оказалась существенно выше эффективности известных программных систем.

Таким образом, основным теоретическим результатом диссертации является представленный в ней новый вычислительный метод обращения полиномиально вычислимых функций, включающий в себя новые алгоритмы обработки дискретных данных, которые могут быть использованы в решении практических задач широкого спектра (исследование дискретно-автоматных моделей, обращение дискретных функций, криптоанализ и т. п.). Основным практическим результатом диссертации является программный комплекс, в котором были реализованы все разработанные алгоритмы.

Основные публикации по теме диссертации

1. **Игнатьев А. С. Двоичные диаграммы решений в логических уравнениях и задачах обращения дискретных функций / А. Е. Хмельнов, А. С. Игнатьев, А. А. Семенов // Вестник НГУ. Серия: Информационные технологии. — 2009. — Т. 7, № 4. — С. 36–52.**

2. **Игнатьев А. С. Использование двоичных диаграмм решений в задачах обращения дискретных функций / А. С. Игнатьев, А. А. Семенов, А. Е. Хмельнов // Вестник Томского гос. ун-та. Серия: Управление, вычислительная техника. — 2009. — № 1(6). — С. 115–129.**

3. Игнатьев А. С. Алгоритмы работы с ROBDD как с базами булевых ограничений / А. С. Игнатьев, А. А. Семенов // Прикладная дискретная математика. — 2010. — № 1. — С. 86–104.

4. Игнатьев А. С. Решение систем логических уравнений с использованием BDD / А. С. Игнатьев, А. А. Семенов, А. Е. Хмельнов // Вестник Томского гос. ун-та. Приложение. — 2006. — № 17. — С. 25–29.

5. Игнатьев А. С. Логические уравнения и двоичные диаграммы решений / А. А. Семенов, А. С. Игнатьев // Прикладные алгоритмы в дискретном анализе. Серия: Дискретный анализ и информатика. — 2008. — Т. 2. — С. 99–126. — ISBN: 978-5-9624-0287-1.

6. Игнатьев А. С. Двоичные диаграммы решений в параллельных алгоритмах обращения дискретных функций / А. С. Игнатьев, А. А. Семенов, Д. В. Беспалов // Труды III Международной научной конференции ПАВТ'09. Нижний Новгород, ННГУ. — 2009. — С. 688–696. — (ISBN: 978-5-696-03854-4).

7. Игнатьев А. С. Гибридный подход (SAT+ROBDD) в задачах криптоанализа поточных систем шифрования / А. С. Игнатьев, А. А. Семенов, Д. В. Беспалов, О. С. Заикин // Труды VIII школы-семинара с международным участием Sibecrypt'09. Омск, ОмГТУ. — 2009. — С. 19–20.

Редакционно-издательский отдел
Института динамики систем и теории управления СО РАН
664033, Иркутск, ул. Лермонтова, д. 134
Подписано к печати 11.05.2010
Формат бумаги 60 × 84 1/16, объем 1,2 п. л.
Заказ 4. Тираж 100 экз.

Отпечатано в ИДСТУ СО РАН