

Федеральное государственное автономное образовательное учреждение
высшего образования Университет ИТМО

Чухарев Константин Игоревич

**Методы декомпозиции задачи булевой выполнимости для синтеза
и верификации моделей автоматных программ**

Специальность 2.3.5

Математическое и программное обеспечение вычислительных систем,
комплексов и компьютерных сетей

Научный доклад
об основных результатах диссертации
на соискание учёной степени
кандидата технических наук

Научный руководитель:
Семёнов Александр Анатольевич
к.т.н., доцент

Санкт-Петербург, Россия
2024

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В современном мире существенную роль играет проблема эффективной верификации разнообразных автоматизированных систем на предмет удовлетворения конкретным спецификациям, а также задача синтеза систем под конкретные спецификации. Под термином «Автоматизированные системы» понимается широкий класс объектов, объединенных общей вычислительной природой — любой такой объект, решая задачу, вычисляет значения некоторой вполне конкретной функции. Для исследования различных свойств таких объектов, в том числе относящихся к практическим приложениям, имеет смысл использовать абстрактные модели, в рамках которых вычисляемые функции задаются программами. Широкий класс автоматизированных систем допускает описание на основе концепций, опирающихся на понятие состояния вычисляющей модели, впервые введенное, по-видимому, А. Тьюрингом в фундаментальной статье¹. Данное понятие является весьма плодотворным, поскольку имеет массу практических приложений, таких как разработка языков программирования, трансляторов и компиляторов, разработка микроконтроллеров под решение конкретных производственных задач, разработка микропроцессоров общего назначения и многое другое. Многие такие практические системы могут быть представлены моделями, в которых состояния понимаются и рассматриваются в рамках конечно-автоматной парадигмы. Такой подход к построению программ подразумевает разбиение программы на более простые модули, которые сами по себе могут рассматриваться как вычислительные единицы и находиться в отдельных состояниях. Данный подход получил известность как концепция автоматного программирования². Для автоматизированной системы, сценарий работы которой представляется в форме автоматной программы, проблемы синтеза и верификации сводятся к аналогичным проблемам для формальных моделей данных систем. В настоящей диссертации рассматриваются два класса таких моделей: конечные автоматы и булевы схемы. Между этими двумя моделями имеется тесная взаимосвязь: конечный автомат задает функцию, которая на вход может принимать данные, вообще говоря, произвольной конечной длины. Булева схема же работает с данными конкретной длины. Соответственно, функции, задаваемой конечным автоматом, соответствует счетное число функций, задаваемых булевыми схемами. Для решения конкретных вычислительных задач, которые,

¹Turing A. M. On Computable Numbers, with an Application to the Entscheidungsproblem // Proceedings of the London Mathematical Society. 1937. Vol. s2–42, no. 1. P. 230–265.

²Поликарпова Н. И., Шалыто А. А. Автоматное программирование. Питер, 2009. 176 с.

как правило, являются комбинаторными, конечно же, приходится рассматривать конечные входные данные и, таким образом, переходить к булевым схемам. Задачи верификации и синтеза для упомянутых моделей являются вычислительно сложными. Даже для булевых схем, работающих с входными конечными данными, задачи, связанные с синтезом и верификацией, относятся к NP-трудным и, таким образом, не могут быть решены известными алгоритмами за полиномиальное время. В данной ситуации (как и во многих других примерах, касающихся NP-трудных задач) для решения конкретных примеров рассматриваемых проблем используются некоторые комбинаторные задачи с хорошо развитой алгоритмической базой. Одной из таких является задача булевой выполнимости (Boolean Satisfiability Problem — SAT), для решения которой за последние 20 лет разработаны весьма эффективные на практике эвристические алгоритмы, применяемые для решения задач символьной верификации³, компьютерной безопасности и криптографии⁴, построению расписаний и планированию⁵ и многим другим прикладным областям. В применении к перечисленным задачам программные реализации алгоритмов решения SAT (так называемые SAT-решатели) дают мощные вычислительные инструменты, позволяющие решать частные случаи рассматриваемых задач таких размерностей, перед которыми другие подходы оказываются бессильны.

Таким образом, актуальной является проблема разработки алгоритмов и программных комплексов на основе алгоритмов решения SAT, используемых для решения задач верификации и синтеза формальных моделей конечно-автоматных программ — конечных автоматов и булевых схем. При решении поставленной задачи возникает целый ряд новых проблем, основной из которых является отсутствие априорных оценок времени работы SAT-решателя на трудной формуле, кодирующей рассматриваемую задачу. Грубо говоря, решатель, получив на вход формулу, может работать час, два, неделю, месяц или даже больше, и нет общего способа определить, сколько времени ему потребуется для завершения работы, притом что формально данный алгоритм является полным и на любой формуле завершает свою работу за конечное время. Описанный феномен известен как «*heavy-tailed behavior phenomenon*»⁶ (НТВ). В рамках настоящей диссертации для борьбы с

³Kroening D. Software Verification // Handbook of Satisfiability. Vol. 336 / ed. by A. Biere [et al.]. 2nd ed. IOS Press, 2021. P. 791–818. (Frontiers in Artificial Intelligence and Applications).

⁴Bard G. V. Algebraic Cryptanalysis. Boston, MA : Springer US, 2009.

⁵Prestwich S. CNF Encodings // Handbook of Satisfiability. Washington, 2021. P. 75–100.

⁶Gomes C. P., Sabharwal A. Exploiting Runtime Variation in Complete Solvers // Handbook of Satisfiability. Vol. 185 / ed. by A. Biere [et al.]. IOS Press, 2009. P. 271–288. (Frontiers in Artificial Intelligence and Applications).

НТВ используются специальные декомпозиционные представления булевых формул, кодирующих описания рассматриваемых моделей. С использованием разработанных алгоритмов удалось решить ряд экстремально сложных задач, относящихся к верификации и синтезу конкретных примеров моделей автоматных программ.

Цель работы. Целью настоящей диссертации является повышение эффективности (снижение времени работы) полных алгоритмов решения задачи булевой выполнимости (SAT) применительно к синтезу и верификации моделей автоматных программ за счет разработки оригинальных методов и техник декомпозиции булевых формул.

Задачи работы. Для достижения поставленной цели были решены следующие научно-технические задачи:

1. Разработаны методы кодирования в SAT задач синтеза конечно-автоматных моделей с заданным поведением и свойствами, отличающиеся от существующих добавлением кодирования структуры охранных условий в виде деревьев разбора соответствующих формул.
2. Разработаны методы кодирования в SAT задачи синтеза модульных конечно-автоматных моделей с заданным поведением и свойствами, отличающиеся от существующих автоматизированным модульным разбиением.
3. Разработаны методы кодирования в SAT задачи синтеза булевых схем и булевых формул по заданной таблице истинности, отличающиеся от существующих возможностью использования произвольных элементарных гейтов.
4. Разработаны методы декомпозиции булевых формул, кодирующих задачи синтеза конечно-автоматных моделей и верификации булевых схем, отличающиеся от существующих возможностью построения оценок декомпозиционной трудности.
5. С применением разработанных методов решены трудные примеры синтеза и верификации моделей автоматных программ — конечных автоматов и логических схем.
6. Разработана программная библиотека `kotlin-satlib`, обеспечивающая взаимодействие с SAT-решателями через программный интерфейс, контроль за различными этапами построения SAT кодировок, а также предоставляющая возможности манипуляции переменными с произвольными конечными доменами.

7. Разработан программный комплекс fVSAT для синтеза и верификации конечно-автоматных моделей с помощью SAT-решателей.
8. Проведены масштабные вычислительные эксперименты для подтверждения практической эффективности всех разработанных методов.

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

1. Алгоритмы синтеза булевых формул и схем, основанные на сведениях к проблеме выполнимости (SAT) и использующие инкрементальные SAT-решатели.
2. Алгоритмы синтеза монолитных и модульных конечно-автоматных моделей, основанные на сведениях к проблеме выполнимости (SAT) и содержащие явное кодирование структуры охранных условий в виде деревьев разбора соответствующих формул.
3. Алгоритмы декомпозиции и методы оценивания декомпозиционной трудности булевых формул, кодирующих задачи синтеза и верификации моделей автоматных программ.
4. Решение экстремально трудных задач синтеза моделей автоматных программ при помощи разработанных алгоритмов.
5. Программная библиотека kotlin-satlib для взаимодействия с SAT-решателями и контроля за процессом построения SAT кодировок.
6. Программный комплекс fVSAT для синтеза и верификации конечно-автоматных моделей с помощью SAT-решателей.

Основные положения, выносимые на защиту.

1. Методы декомпозиции булевых формул, применяемые к задачам тестирования и верификации моделей автоматных программ и использующие SAT-решатели, отличающиеся от известных методов тем, что с целью получения более точных верхних оценок трудности формул в предлагаемых методах используются специальные конструкции SAT-разбиений.
2. Метод синтеза минимальных представлений булевых функций в виде формул и схем, использующий сведения к задаче выполнимости (SAT), отличающийся от существующих подходов тем, что с целью достижения более высокой эффективности (относительно времени и точности решения) предлагаемый метод использует инкрементальные SAT-решатели.
3. Методы синтеза и верификации монолитных и модульных конечно-автоматных моделей по примерам поведения и формальной спецификации,

использующие сведения к задаче выполнимости (SAT) и контрпримеры (Counterexample-Guided Inductive Synthesis — CEGIS), отличающиеся от существующих подходов тем, что с целью повышения эффективности (относительно времени решения) применяется техник явного кодирования структуры охранных условий.

4. Программная библиотека `kotlin-satlib`⁷ для взаимодействия с SAT-решателями и обеспечения контроля над всеми этапами построения SAT-кодировок, отличающаяся от известных библиотек тем, что с целью расширения области применимости разработанная библиотека предоставляет широкий выбор различных SAT-решателей и возможность манипулировать переменными с произвольными конечными доменами.
5. Программный комплекс `fbSAT`⁸ для синтеза и верификации конечно-автоматных моделей с помощью SAT-решателей, включающий в себя реализацию всех предложенных методов.

Теоретическая и практическая значимость. Теоретическая значимость диссертации заключается в разработанных в ней концепциях и алгоритмах решения задач синтеза моделей автоматных программ и методах построения оценок декомпозиционной трудности таких задач. Практическая значимость диссертации состоит в том, что основные разработанные в ней алгоритмы применимы к индустриальным задачам проектирования, синтеза и верификации моделей автоматных программ, а также в том, что на целом ряде конкретных примеров практическая реализация и апробация разработанных алгоритмов демонстрируют лучшую эффективность (меньшее время решения) в сравнении с известными подходами.

Методы и инструменты исследования. Теоретическая часть работы использует методологию дискретной математики, математической логики и теории вычислительной сложности. Для синтеза конечно-автоматных моделей был использован программный комплекс `fbSAT`, разработанный в рамках данной диссертации. Для верификации конечно-автоматных моделей использовался символьный верификатор `NuSMV`⁹. При построении вычислительных задач из области проверки логической эквивалентности схем использовалась программная система `Transalg`¹⁰. Для решения конкретных инстансов задачи SAT использовались различные современные SAT-

⁷<https://github.com/Lipen/kotlin-satlib>

⁸<https://github.com/ctlab/fbSAT>

⁹<https://nusmv.fbk.eu>

¹⁰<https://gitlab.com/transalg/transalg>

решатели, находящиеся в открытом доступе, такие как MiniSAT¹¹, Glucose¹², Kissat¹³, CaDiCaL¹⁴. Для взаимодействия с SAT-решателями через программный интерфейс использовалась программная библиотека `kotlin-satlib`, разработанная в рамках данной диссертации. В вычислительных экспериментах задействовался вычислительный кластер.

Соответствие специальности. Содержание диссертации соответствует паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» в перечисленных ниже пунктах 1 («Модели, методы и алгоритмы проектирования, анализа, трансформации, верификации и тестирования программ и программных систем») и 3 («Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем»):

- в диссертации представлено семейство методов и алгоритмов, применяемых для трансформации автоматных программ в булевы схемы и формулы с целью вычислительного решения задач синтеза и верификации автоматных программ (пункт 1);
- представлены алгоритмы решения задач синтеза, верификации и тестирования моделей автоматных программ при помощи декомпозиционных представлений булевых формул, кодирующих исходные задачи (пункт 1);
- разработанная программная библиотека `kotlin-satlib` обеспечивает взаимодействие между алгоритмами кодирования в SAT задач синтеза и верификации автоматных программ и современными эффективными SAT-решателями (пункт 3).

Достоверность результатов проведённых исследований. Достоверность полученных в работе результатов обеспечивается теоретической корректностью предложенных алгоритмов, эффективность которых обоснована масштабными вычислительными экспериментами.

Апробация работы. Основные результаты диссертации докладывались на следующих конференциях:

- VIII Конгресс молодых ученых, Университет ИТМО, Санкт-Петербург, 2019.
- Конференция СПИСОК-2019, СПбГУ, Санкт-Петербург, 2019.

¹¹<https://github.com/niklasso/minisat>

¹²<https://github.com/audemard/glucose>

¹³<https://github.com/arminbiere/kissat>

¹⁴<https://github.com/arminbiere/cadical>

- IX Конгресс молодых ученых, Университет ИТМО, Санкт-Петербург, 2020.
 - Конференция ППС 2021, Университет ИТМО, Санкт-Петербург, 2021.
 - X Конгресс молодых ученых, Университет ИТМО, Санкт-Петербург, 2021.
 - Воркшоп SAT/SMT Solvers: Theory and Practice, Санкт-Петербург, 2021.
 - XI Конгресс молодых ученых, Университет ИТМО, Санкт-Петербург, 2022.
- Диссертационная работа была выполнена при поддержке грантов и проектов:
- Грант РФФИ №19-07-01195 А «Разработка методов машинного обучения на основе SAT-решателей для синтеза модульных логических контроллеров киберфизических систем».
 - Грант №19-37-51066 Научное наставничество «Разработка методов синтеза конечно-автоматных алгоритмов управления для программируемых логических контроллеров в распределенных киберфизических системах».
 - НИР-ФУНД 77051 «Исследование алгоритма тестирования на основе обучения и улучшения его эффективности», 2020-2021.
 - НИР-ПРИКЛ 222004 «Алгоритмы решения SAT для логических схем и анализа программ», 2021-2023.
 - НИР-ПРИКЛ 223099 «Алгоритмы решения SAT для логических схем и анализа программ», 2023-2024.

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в 10 публикациях. Из них одна издана в журналах, рекомендованных ВАК, а четыре — в изданиях, индексируемых в базе цитирования Scopus.

Личный вклад автора. Методы синтеза монолитных и модульных конечно-автоматных моделей по примерам поведения и формальной спецификации, основанные на сведениях к задаче SAT, разработаны соискателем в соавторстве с Чивилихиным Д. С. Методы синтеза и верификации модульных конечно-автоматных моделей по примерам поведения и формальной спецификации, основанные на сведениях к задаче SAT и использовании контрпримеров, разработаны соискателем в соавторстве с Чивилихиным Д. С. и Суворовым Д. М. Программный комплекс fVSAT разработан лично соискателем. Реализация всех разработанных методов синтеза и верификации конечно-автоматных моделей в программном комплексе fVSAT выполнена лично соискателем. Метод синтеза булевых формул и схем по заданной таблице истинности, основанный на сведениях к задаче SAT, предложен и разработан лично соискателем. Прототип программной библиотеки `kotlin-satlib` для взаимодействия с SAT-решателями через унифицированный программный интерфейс разработан в соавторстве с Гречишкиной Д. С. Дальнейшая разработка

и расширение программной библиотеки `kotlin-satlib`, что включает в себя поддержку дополнительных SAT-решателей (например, Kissat) и разработку модуля для манипуляции переменными с конечными доменами, производилась лично соискателем. Общая стратегия оценивания трудности формул, кодирующих проверку эквивалентности (задача верификации) булевых схем, разработана соискателем в соавторстве с Семёновым А. А., Кондратьевым В. С., Кочемазовым С. Е. и Тарасовым Е. А. Описанные конструкции декомпозиций формул, кодирующих эквивалентность булевых формул, предложены лично соискателем. Теоретическое обоснование корректности предложенных конструкций выполнены соискателем в соавторстве с Семёновым А. А.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В Главе 1...

В Главе 2...

В Главе 3...

В заключении сформулированы основные результаты научной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ И ВЫВОДЫ

В данной диссертации были получены следующие основные результаты и выводы:

- Разработаны алгоритмы декомпозиции булевых формул, которые могут применяться к задачам верификации моделей автоматных программ. Эти алгоритмы позволяют строить оценки декомпозиционной трудности рассматриваемых формул.
- Предложены новые алгоритмы кодирования в SAT задачи синтеза булевых схем и булевых формул по заданной таблице истинности. Эти алгоритмы отличаются от существующих возможностью использования произвольных элементарных гейтов, что расширяет их применение.
- Разработаны алгоритмы кодирования в SAT монолитных и модульных конечно-автоматных моделей для решения задач синтеза и верификации. Эти алгоритмы включают явное кодирование структуры охранных условий в виде деревьев разбора соответствующих формул, что позволяет их минимизировать, что в свою очередь повышает человеко-читаемость полученных выражений.

- Созданы программные библиотеки для взаимодействия с различными SAT-решателями через специально разработанный программный интерфейс. Эти библиотеки отличаются возможностью контроля всех этапов построения SAT-кодировок и манипулирования переменными с произвольными конечными доменами.
- Проведены масштабные вычислительные эксперименты, подтвердившие практическую эффективность всех разработанных методов. Эти эксперименты продемонстрировали значительное улучшение в решении трудных примеров синтеза и верификации моделей автоматных программ — конечных автоматов и булевых схем.

Таким образом, разработанные методы и алгоритмы значительно повышают эффективность (снижают время работы) полных алгоритмов решения задачи булевой выполнимости (SAT) и вносят важный вклад в области синтеза и верификации моделей автоматных программ.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. D. Chivilikhin, S. Patil, **K. Chukharev**, A. Cordonnier, V. Vyatkin. Automatic State Machine Reconstruction from Legacy PLC Using Data Collection and SAT Solver // IEEE Transactions on Industrial Informatics. 2020. Vol. 16, issue 12. P. 7821–7831. (**Q1, Scopus**)
2. **K. Chukharev**, D. Suvorov, D. Chivilikhin. SAT-based Counterexample-Guided Inductive Synthesis of Distributed Controllers // IEEE Access. 2020. Vol. 8. P. 207485–207498. (**Q1, Scopus**)
3. **K. Chukharev**, D. Chivilikhin. fbSAT: Automatic Inference of Minimal Finite-State Models of Function Blocks Using SAT Solver // IEEE Access. 2022. Vol. 10. P. 131592–131610. (**Q1, Scopus**)
4. Andreev A., **Chukharev K.**, Kochemazov S., Semenov A. Solving Influence Maximization Problem under Deterministic Linear Threshold Model using Metaheuristic Optimization // MIPRO, 2024. (**Scopus**)
5. **Чухарев К.** Применение инкрементальных SAT-решателей для решения NP-трудных задач на примере задачи синтеза минимальных булевых формул // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20, 6(130). С. 841—847. (**ВАК**)

6. **Чухарев К.**, Чивилихин Д. Построение минимальных конечно-автоматных моделей функциональных блоков по обучающим примерам // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. СПб: Университет ИТМО, 2018. URL: <http://openbooks.ifmo.ru/ru/file/8061/8061.pdf> (дата обр. 17.06.2020).
7. **Чухарев К.** Построение конечно-автоматных моделей функциональных блоков по примерам поведения и темпоральным свойствам // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. СПб: Университет ИТМО, 2019. URL: <https://kmu.itmo.ru/digests/article/1283> (дата обр. 17.06.2020).
8. **Чухарев К.** Автоматический синтез минимальных конечно-автоматных моделей функциональных блоков по примерам поведения и темпоральным свойствам // Материалы 8-й всероссийской научной конференции по проблемам информатики СПИСОК-2019. СПб.: ВВМ, 2019.
9. **Чухарев К.** Синтез конечно-автоматных моделей модульных логических контроллеров по примерам поведения с помощью SAT-решателей // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. СПб: Университет ИТМО, 2020. URL: <https://kmu.itmo.ru/digests/article/3555> (дата обр. 17.06.2020).
10. Гречишкина Д., **Чухарев К.** Программный интерфейс для SAT-решателей на основе технологии JNI // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. СПб: Университет ИТМО, 2020. URL: <https://kmu.itmo.ru/digests/article/4456> (дата обр. 17.06.2020).