Федеральное государственное автономное образовательное учреждение
высшего образования Университет ИТМО

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

Чухарев Константин

Специальность 2.3.5 (05.13.11)

Математическое и программное обеспечение вычислительных машин,
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 статье[1]. 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 программирования[2]. 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, как правило, являются

[1] *Turing 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.

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

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 верификации[3], компьютерной безопасности и криптографии[4], построению расписаний и планированию[5] 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 — то есть 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] (HTB). В рамках настоящей диссертации для борьбы с

[3] *Kroening 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).

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

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

[6] *Gomes 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.

Учитывая все сказанное выше, сформулируем основные цели и задачи работы.

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 оригинальные алгоритмы 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 оригинальные алгоритмы кодирования в SAT задачи синтеза булевых схем и булевых формул по заданной таблице истинности, отличающиеся от существующих возможностью использования произвольных элементарных гейтов.
3. Разработаны оригинальные алгоритмы декомпозиции булевых формул, кодирующих задачи синтеза конечно-автоматных моделей и верификации булевых схем, отличающиеся от существующих возможностью построения оценок декомпозиционной трудности.
4. С применением разработанных алгоритмов решены трудные примеры синтеза и верификации моделей автоматных программ — как конечных автоматов, так и логических схем.
5. Разработана оригинальная программная библиотека, обеспечивающая взаимодействие с SAT-решателями через программный интерфейс, контроль за рахличными этапами построения SAT кодировок, а также предоставляющая возможности манипуляции переменными с произвольными конечными доменами.
6. 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. Новые алгоритмы синтеза моделей автоматных программ (конечных автоматов и булевых схем), основанные на сведении к проблеме булевой выполнимости (SAT).
2. Новые алгоритмы декомпозиции и новые методы оценивания декомпозиционной трудности булевых формул, кодирующих задачи синтеза и верификации моделей автоматных программ.
3. Решение экстремально трудных задач синтеза моделей автоматных программ при помощи разработанных алгоритмов.
4. Новая программная библиотека для взаимодействия с SAT-решателями.

Основные положения, выносимые на защиту.

1. Оригинальные алгоритмы декомпозиции булевых формул, применяемые к задачам верификации моделей автоматных программ и отличающиеся от существующих возможностью построения оценок декомпозиционной трудности рассматриваемых формул.
2. Оригинальные алгоритмы кодирования в SAT задачи синтеза булевых схем и булевых формул по заданной таблице истинности, отличающиеся от существующих возможностью использования произвольных элементарных гейтов.
3. Оригинальные алгоритмы кодирования в SAT моделей автоматных программ, применяемые к решению задач синтеза и верификации данных моделей и отличающиеся от существующих наличием явного кодирования структуры охранных условий, представленных в виде деревьев разбора соответствующих формул.
4. Программная библиотека[7,8] для взаимодействия с различными SAT-решателями через программный интерфейс, отличающаяся от существующих возможностью контроля всех этапов построения SAT кодировки, а также возможностью манипулировать переменными с произвольными конечными доменами.

Теоретическая и практическая значимость. Теоретическая значимость диссертации заключается в разработанных в ней концепциях и алгоритмах решения

[7] Для языка Kotlin: `https://github.com/Lipen/kotlin-satlib`

[8] Для языка Rust: `https://github.com/Lipen/sat-nexus`

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 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 и математической логики, теории вычислительной сложности,а также теорию эволюционных вычислений. Для синтеза конечно-автоматных моделей был использован программный комплекс fbSAT, разработанный в рамках данной диссертации[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-решатели, находящиеся в открытом доступе, такие как MiniSAT[11], Glucose[12], Kissat[13], CaDiCaL[14]. В вычислительных экспериментах задействовался вычислительный кластер.

Соответствие специальность. Содержание научно-квалификационной работы соответствует паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» в перечисленных ниже пунктах:

- 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 содержит программные инструменты, обеспечивающие взаимодействие между алгоритмами кодирования в

[9] https://github.com/ctlab/fbSAT

[10] https://gitlab.com/transalg/transalg

[11] https://github.com/Lipen/minisat

[12] https://github.com/Lipen/glucose

[13] https://github.com/Lipen/kissat

[14] https://github.com/Lipen/cadical

SAT задач синтеза и верификации автоматных программ и современными эффективными SAT решателями (пункт 3).

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.
- 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 9 публикациях. Из них 3 опубликовано в изданиях, индексируемых в базе цитирования Scopus.

Краткое содержание работы

В Главе 1...

В Главе 2...

В Главе 3...

В заключении сформулированы основные результаты научной работы.

Основные выводы и результаты работы

В данной диссертации были получены следующие основные результаты и выводы:

- Разработаны оригинальные 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 оригинальные алгоритмы кодирования в SAT моделей автоматных программ для решения задач синтеза и верификации данных моделей. 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, как конечных автоматов, так и логических 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.

Основные публикации по теме диссертации

1. Automatic State Machine Reconstruction from Legacy PLC Using Data Collection and SAT Solver / D. Chivilikhin, S. Patil, **K. Chukharev**, A. Cordonnier, V. Vyatkin // IEEE Transactions on Industrial Informatics. 2020. Vol. 16, issue 12. P. 7821–7831. (**Q1, Scimago**)
2. SAT-based Counterexample-Guided Inductive Synthesis of Distributed Controllers / **K. Chukharev**, D. Suvorov, D. Chivilikhin // IEEE Access. 2020. Vol. 8. P. 207485–207498. (**Q1, Scimago**)
3. fbSAT: Automatic Inference of Minimal Finite-State Models of Function Blocks Using SAT Solver / **K. Chukharev**, D. Chivilikhin // IEEE Access. 2022. Vol. 10. P. 131592–131610. (**Q1, Scimago**)
4. **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**)
5. **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).
6. **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).
7. **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.
8. **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).
9. 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).