

АНОТАЦІЯ

Тарасіч Ю.Г. Статичний аналіз лінійно визначених програм і його застосування. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 121 Інженерія програмного забезпечення (12 Інформаційні технології). – Херсонський державний університет, Херсон, 2021.

Загальна проблема дослідження полягає у дослідженні та розробці методів створення моделей ефективних програмних систем статичного аналізу, програм та математичних моделей об'єктних програм для статичного аналізу, дослідженні математичних методів та розробці ефективних алгоритмів комп'ютерної алгебри статичного аналізу програм.

Підхід проведення досліджень полягає у:

- аналізі методів та функціональності існуючих систем статичного аналізу програм,
- теоретичному аналізі результатів досліджень з теорії програмних інваріантів, що включає аналіз математичних моделей програм та алгоритмів комп'ютерної алгебри, що використовуються у теорії програмних інваріантів,
- створенні нових алгоритмів комп'ютерної алгебри.

Мета роботи полягає в аналізі існуючих алгоритмів комп'ютерної алгебри та інсерційного моделювання в системах статичного аналізу та верифікації програмного забезпечення і побудові ефективних алгоритмів статичного аналізу програм.

Основні завдання дослідження:

- Дослідження існуючих спеціалізованих систем статичного аналізу програм.
- Аналіз алгоритмів комп'ютерної алгебри та інсерційного моделювання в системах статичного аналізу та верифікації програмного забезпечення.

- Аналіз існуючих алгоритмів побудови канонічних форм логічних формул та інструментів спрощення формул (солверів).

- Створення нових засобів та алгоритмів статичного аналізу програм, зокрема, алгоритмів доказу інваріантних нерівностей.

- Розробка нових алгоритмів побудови канонічних форм лінійних напівалгебраїчних формул.

- Реалізація алгоритму побудови канонічних форм лінійних напівалгебраїчних формул.

У першому розділі дисертації (*Статичний аналіз лінійно визначених програм*) надано аналіз існуючих методів та систем статичного аналізу, приведено базисні дефініції дисертаційного дослідження – визначено загальні поняття графових та алгебраїчних моделей програм, приведено визначення поняття лінійно визначених програм.

У другому розділі дисертаційного дослідження (*Генерація інваріантів програм*) приведено поняття власного полінома лінійного оператора та алгоритм побудови власних поліномів, визначено зв'язок між власними поліномами лінійних операторів та поліноміальними інваріантами лінійних циклів програм (підрозділ 2.1), представлено нові методи доказу інваріантності системи лінійних нерівностей і завершуваності лінійно визначених ітеративних циклів імперативних програм (підрозділ 2.2).

Основні алгоритми проблеми доказу та генерації поліноміальних інваріантних рівностей лінійно визначених циклів - побудова множини L – інваріантів циклів для операторів, жорданова форма яких містить нетривіальні блоки; теорема 2.12, в якій встановлено структуру базису Гребнера ідеалу поліноміальних інваріантів лінійного циклу з довільним невідродженим лінійним оператором в тілі циклу і теорема 2.14 про алгоритмічну розв'язуваність проблеми побудови базису ідеалу інваріантів «діагоналізованої частини» лінійного оператора в разі неприводимості його мінімального характеристичного полінома. У зв'язку з цим, за теоремою 2.12, інваріанти лінійного оператора можна класифікувати як

внутрішньоклітинкові - ті, які притаманні кожній жордановій клітинці лінійного оператора, і міжклітинкові - ті, які притаманні його діагоналізуємій частині. Внутрішньоклітинкові інваріанти обчислюються безпосередньо за

формулами
$$x_j = \frac{z}{c} \left(a_j + \frac{C_1 \left(\lambda \frac{cy-bz}{cz} \right)}{\lambda} a_{j+1} + \dots + \frac{C_{n-j} \left(\lambda \frac{cy-bz}{cz} \right)}{\lambda^{n-j}} a_n \right).$$
 Існування

міжклітинкових інваріантів залежить від існування нетривіальних мультиплікативних співвідношень між власними числами лінійного оператора (теорема 2.2).

Для лінійних операторів з непривідним мінімальним характеристичним многочленом проблема побудови базису множини мультиплікативних співвідношень між його власними числами алгоритмічно розв'язна, але алгоритм теореми 2.14 є неефективним, зважаючи на дуже великі степені многочлена $S(x)$, який потрібно розкласти на множники.

У методі доказу інваріантності системи лінійних нерівностей, а також завершуваності лінійно визначених ітеративних циклів імперативних програм враховується передумова циклу та умова його повторення у вигляді сукупності системи лінійних нерівностей. В основі методу лежить побудова та аналіз спектру лінійного оператора тіла циклу та обчислення числа його ітерацій, після виконання яких інваріантність або підтверджується, або спростовується.

Запропоновані методи доказу та генерації інваріантних рівностей та доказу інваріантних нерівностей може бути покладено в основу загального алгоритму доказу інваріантності системи лінійних нерівностей та поліноміальних рівностей для лінійно-визначених програм.

У розділі 3 (*Канонічні форми лінійних напівалгебраїчних формул над типами даних та їх використання в задачах верифікації програм*) представлено результати аналізу функціональних особливостей існуючих систем побудови доказів (на прикладі CVC4, MathSAT5, QEPcad, Singular, COCOA, MiniZinc, STP, RedLog, Satallax, Isabelle, E-SETHEO, Minisat,

SMTInterpol, TPS / ETPS, Paradox, Gandalf, Vampire), зокрема, на наявність інструментів спрощення формул (підрозділ 3.1); представлено новий алгоритм побудови канонічної форми лінійних півалгебраїчних (ЛПФ) формул (підрозділи 3.2-3.3), включаючи алгоритми побудови канонічних форм логічних формул над перелічуваним та множиним типами даних (підрозділ 3.4); представлено модифікацію алгоритму побудови канонічної форми лінійних напівалгебраїчних формул за рахунок застосування алгоритму поповнення (підрозділ 3.6), приведено реалізацію алгоритму побудови канонічної форми ЛПФ засобами IMS та C++ (підрозділ 3.5).

Основний результатом є визначення канонічної форми ЛПФ, що володіє властивістю єдиності й іншими корисними властивостями, а також представлення алгоритму її побудови. Пропонована канонічна форма ЛПФ - пряме узагальнення канонічної форми представлення системи лінійних нерівностей і алгоритму її побудови, наведених у [3.37, 3.38].

Наукова новизна одержаних результатів.

— Уперше представлено новий метод визначення канонічної форми ЛПФ, що володіє властивістю єдиності та іншими корисними властивостями, а також алгоритм її побудови.

— Уперше реалізовано алгоритм побудови канонічних форм логічних формул засобами інсерційного моделювання (система інсерційного моделювання IMS).

— Набуло подальшого розвитку дослідження проблеми пошуку інваріантів циклів, а саме, - наведено алгоритм обчислення основних інваріантів лінійного оператора жорданової клітинки та алгоритм обчислення основних інваріантів діагоаналізуемого лінійного оператора з непривідним мінімальним характеристичним поліномом та представлено новий метод доказу інваріантності системи лінійних нерівностей і завершення деяких лінійних ітераційних циклів імперативних програм, дані яких є елементами конструктивного лінійно впорядкованого поля.

Практичне значення наукових результатів полягає в можливості

застосування наукових положень і висновків дослідження в задачах верифікації програмного забезпечення, а саме в розробці спеціалізованих програмних систем верифікації формальних моделей програм.

Розвиток даного підходу у верифікації дозволить досить ефективно та надійно розробляти програмне забезпечення для медичної апаратури, військової техніки, бортових систем управління для авіації та космонавтики, технології блокчейн, загалом, критичних програмних систем.

На теперішньому етапі виконується дослідження застосування методів формальної верифікації, зокрема, методів статичного аналізу та використання систем верифікації для верифікації формальних моделей програм, економічних, біологічних процесів, нормативно-правових документів, тощо, продовжується робота над дослідженням математичних методів та розробкою ефективних алгоритмів комп'ютерної алгебри статичного аналізу програм.

Ключові слова: статичний аналіз, верифікація, лінійно визначені програми, інваріант циклу, програмні інваріанти.

ABSTRACT

Yuliia H. Tarasich. The static analysis of linearly defined programs and its application. - Qualification scientific paper, manuscript.

Thesis of a Doctor of Philosophy in specialty 121 Software engineering (12 Information technologies). - Kherson State University, Kherson, 2021.

The general problem of study is research and development of methods of creation of models of effective software systems for static programs analysis and mathematical models of object programs for static analysis, research of mathematical methods and development of effective algorithms of computer algebra of programs static analysis.

The research approach consist of:

- the analysis of methods and of functionality of existing systems of static programs analysis,
- the theoretical analysis of research results in the theory of program invariants, including analysis of mathematical models of programs and computer algebra algorithms that are used in the theory of program invariants,
- the development of new computer algebra algorithms.

The **main aim of the work** is to analyze the existing algorithms of computer algebra and insertion modeling in the systems of software static analysis and verification and to build an effective algorithms for static analysis of programs.

The main tasks of the study:

- Research of existing specialized systems for static analysis of programs.
- Analysis of algorithms of computer algebra and insertion modeling in static analysis and software verification systems.
- Analysis of existing algorithms for constructing canonical forms of logical formulas and simplification tools (solvers).
- Creation of new tools and algorithms for static analysis of programs, in particular, creation of algorithms for proving of invariant inequalities.

- Development of new algorithms for constructing canonical forms of linear semi-algebraic formulas.

- Implementation of the algorithm for constructing canonical forms of linear semi-algebraic formulas.

In the first section of the dissertation (*The static analysis of linearly defined programs*) an analysis of existing methods and systems of static analysis and basic definitions of dissertation research (such as general concepts of graph and algebraic models of programs, the concept of linearly defined programs) are provided.

In the second section of the dissertation research (*Generation of invariants of programs*) the concept of eigenpolynom of a linear operator is introduced, the algorithm of construction of eigenpolynomials is formulated and the connection between eigenpolynoms and polynomial invariants of linear cycles of programs is established (subsection 2.1); a new method for proving the invariance of the system of linear inequalities, and also of the completion of linearly defined iterative cycles of imperative programs is presented (subsection 2.2).

The main algorithms of the problem of proof and generation of polynomial invariant equations are the construction of the set L - invariants of cycles for operators whose Jordan form contains nontrivial blocks; Theorem 2.13, in which the structure of the Grebner basis of the ideal of polynomial invariants of a linear cycle with an arbitrary nondegenerate linear operator in the body of the cycle is establishing and Theorem 2.15 about the algorithmic solvability of the problem of constructing the basis of the ideal of invariants of the "diagonalizable part" of a linear operator in the case of irreducibility of its minimal characteristic polynomial. Thus, by theorem 2.13, the invariants of a linear operator can be classified as intracellular - that are inherent to each Jordan cell of linear operator, and intercellular - those that are inherent in its diagonalisable part.

Intracellular invariants are computed directly from the formulas

$$x_j = \frac{z}{c} \left(a_j + \frac{C_1 \left(\lambda \frac{cy-bz}{cz} \right)}{\lambda} a_{j+1} + \dots + \frac{C_{n-j} \left(\lambda \frac{cy-bz}{cz} \right)}{\lambda^{n-j}} a_n \right).$$

The existence of intercellular invariants depend on the existence of nontrivial multiplicative relations between the eigenvalues of the linear operator (theorem 2). For linear operators with an irreducible minimum characteristic polynomial problem of constructing a basis of set of multiplicative relations between its eigenvalues is algorithmically solvable, but the algorithm of theorem 8 is ineffective due to a very large degree of the polynomial $S(x)$, which is necessary to decompose into factors.

The method of proving the invariance of the system of linear inequalities, and also of the completion of linearly defined iterative cycles of imperative programs takes into account the premise of the cycle, and also the condition of cycle repetition in the form of a set of systems of linear inequalities. The method is based on the construction and analysis of the spectrum of the cycle body and on the calculation of the number of iterations of the cycle, after which the invariance is confirmed or refuted.

The proposed methods of proof and generation of invariant equations and proof of invariant inequalities can be used as a basis for the general algorithm for proving the invariance of the system of linear inequalities and polynomial equalities for linearly defined programs.

In Section 3 (*Canonical forms of linear semi-algebraic formulas over the data types and their use in program verification*) the results of the analysis of functional features of existing proving systems, in particular on the availability of formula simplification tools, (CVC4, MathSAT5, QEPCAD, Singular, COCOA, MiniZinc, STP, RedLog, Satallax Isabelle, E-SETHEO, Minisat, SMTInterpol, TPS / ETPS, Paradox, Gandalf, Vampire) are presented (section 3.1); a new algorithm for constructing the canonical form of linear semi-algebraic (LSF) formulas (subsections 3.2-3.3) is presented, including algorithms for constructing canonical forms of logical formulas over enumerated and multiple types

(subsection 3.4); the modification of the algorithm for constructing the canonical form of linear semi-algebraic formulas due to the application of the replenishment algorithm is presented (subsection 3.6), the implementation of the algorithm for constructing the canonical form of LSF by means of IMS and C ++ (subsection 3.5) is given.

The main result is the definition of the canonical form of LSF, which has the property of unity and other useful properties, and also the description of the algorithm for its construction. The proposed canonical form of LSF is a direct generalization of the canonical form of representation of the system of linear inequalities and the algorithm of its construction that were given in [3.37, 3.38].

Scientific novelty of the obtained results:

- the problem of generating of polynomial invariants of iterative cycles with a loop initialization operator and a nondegenerate linear operator in the loop body is considered.

- the algorithm of calculation of the basic invariants of the linear operator of a Jordan cell and the algorithm of calculation of the main invariants of the diagonalized linear operator with the irreducible minimum characteristic polynomial are shown.

- the new method of proving the invariance of a system of linear inequalities and the completion of some linear iterative cycles of imperative programs, the data of which are elements of a constructive linearly ordered field, is presented.

- a new method of determining the canonical form of LSF, which has the property of uniqueness and other useful properties, and also the description of the algorithm for its construction were given.

- the algorithm of construction of canonical forms of logical formulas is realized.

The practical significance of the expected scientific results lies in the possibility of applying scientific provisions and research conclusions in software verification tasks, namely in the development of specialized software systems for verification of formal program models.

The development of this approach in verification will allow to effectively and reliably develop a medical equipment, a military equipment, an onboard control systems for aviation and aerospace, a blockchain technology, in general, the software, where the cost of error and unreliability of software can be expensive.

At the present stage we work on the study of the application of formal verification methods, including methods of static analysis, and of the using of verification systems to verify formal models of programs, economic, biological processes, legal documents, etc. We also continue to work on the study of mathematical methods and of the development of efficient algorithms of computer algebra of static program analysis.

Keywords: static analysis, verification, linearly defined programs, cycle invariant, program invariants.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Основні наукові результати дисертації

1. Lvov M. The Static Analysis of Linear Loops [Електронний ресурс] / M. Lvov, Y. Tarasich // CEUR Workshop Proceedings. – 2015. – Режим доступу до ресурсу: http://ceur-ws.org/Vol-1356/paper_79.pdf.

Особистий внесок здобувачки полягає в аналізі та описі основних результатів проблеми доказу та генерації поліноміальних інваріантних рівностей, визначенні нових засобів та алгоритмів статичного аналізу програм, а саме - нового алгоритму доказу інваріантних нерівностей.

2. The canonical forms of logical formulas off the data types / M. Lvov, V. Peschanenko, O. Letychevskyi, Yu. Tarasich // Теоретичні та прикладні аспекти побудови програмних систем: матеріали міжнародної наукової конференції, м. Київ, 5-9 грудня 2016 р. / редкол.: М.С. Нікітченко, Д.Б. Буй та ін. — Кіровоград: ПП «Центр оперативної поліграфії «Авангард»», — 2016. — XXX с. (тези)

Особистий внесок здобувачки полягає у представлені нових алгоритмів обчислення канонічної форми логічних формул над перелічуванням та множинним типами даних.

3. The canonical forms of logical formulae over the data types and their using in programs verification [Електронний ресурс] / M.Lvov, V. Peschanenko, O. Letychevskyi, Y. Tarasich // CEUR Workshop Proceedings. – 2017. – Режим доступу до ресурсу: <http://ceur-ws.org/Vol-1844/10000536.pdf>.

Особистий внесок здобувачки полягає у стислому огляді засобів спрощення формул, представлені нового алгоритму обчислення канонічної форми логічних формул над упорядкованими типами даних.

4. Algorithm and Tools for Constructing Canonical Forms of Linear Semi-Algebraic Formulas / [M. Lvov, V. Peschanenko, Y. Tarasich та ін.]. // Cybernetics and Systems Analysis. – 2018. – №54. – С. 993–1002.

Особистий внесок здобувачки полягає в тестуванні засобів спрощення формул та аналізі їх основних функціональних особливостей; представленні нового алгоритму побудови канонічних форм лінійних напівалгебраїчних формул.

Апробація матеріалів дисертації

5. Formalization and algebraic modeling of tokenomics projects [Електронний ресурс] / [O. Letychevskyi, V. Peschanenko, V. Radchenko та ін.] // CEUR Workshop Proceedings. – 2019. – Режим доступу до ресурсу: http://ceur-ws.org/Vol-2393/paper_409.pdf.

Особистий внесок здобувачки полягає у застосуванні розроблених алгоритмів для моделювання визначених задач.

6. Our Approach to Formal Verification of Token Economy Models / O.Letychevskyi, V. Peschanenko, M. Poltoratskyi, Y. Tarasich, 2020. – (Communications in Computer and Information Science). – (1175 CCIS). – С. 348 – 363.

Особистий внесок здобувачки полягає у застосуванні розроблених алгоритмів для моделювання визначених задач.

7. Platform for Modeling of Algebraic Behavior: Experience and Conclusions / Letychevskyi, O., Peschanenko, V., Poltoratskyi, M., Tarasich, Y...// CEUR Workshop Proceedings. - 2020. – Режим доступу до ресурсу: <http://ceur-ws.org/Vol-2732/20200042.pdf>

Особистий внесок здобувачки полягає у застосуванні розроблених алгоритмів для моделювання визначених задач.