Verification by Abstract Interpretation,
Soundness and Abstract Induction

Patrick COUSOT
Courant Institute of Mathematical Sciences, New York University
pcousot@cims.nyu.edu, cims.nyu.edu/~pcousot

Abstract
Automatic program verification tools have to cope with programming language and machine semantics, undecidability, and mathematical induction, and so are all complex and imperfect. The ins and outs of automatic program verification will be discussed in light of the theory and practice of abstract interpretation [18, 19, 22].

Programming language semantics
If Edsger W. Dijkstra could claim in his 1972 Turing lecture [29, p. 863] that “When FORTRAN has been called an infantile disorder, full PL/1, with its growth characteristics of a dangerous tumor, could turn out to be a fatal disease.”, PL/1 and its 1000 pages formal operational semantics defined in VDL [7] now appears as a marvel of simplicity compared to present-day programming languages [57] and their informal definitions [1].

The formal specification of the semantics of programming languages is now a well-established and well-taught subject in computer science, with many proposals at various levels of abstraction such as small/big step operational, denotational, relational and predicate transformer, axiomatic, and algebraic semantics. Surprisingly, very few languages have received a formal definition of their semantics (with few exceptions such as ML [47, 48]). None of the most popular programming languages has a formal and complete definition of its semantics that can reasonably be used as a basis for formal verification. This may come from the high complexity of present-day programming languages.

This problem is often solved by the verification community by considering models or programs that are never checked to be valid or toy languages hardly ever used for programming. Typically there are reals no floats, arrays but infinite, lists or trees but no pointers, etc.

Considering actual languages is a huge and long effort [24, 46], not necessarily more highly rewarded.

Machine semantics
For a long time, computers have been miraculously correct (when ever bugs were found the corresponding pages of the user manual had just to be teared up). Since the exponential improvement of the power of computers by Moore’s Law started to slow down significantly a decade ago, computer industry technology moved to multiprocessors. Because memory access is about 100 times slower than CPU, modern microprocessors optimize memory accesses by re ordering memory operations using caches and memory banks. On most modern multiprocessors memory operations are not executed in the order specified by the program code. In multi-threaded environments (or when interfacing with other hardware via memory busses) this out-of-order execution may lead to serious problems. Barriers or fences have been introduced to synchronize memory accesses. Their functioning is far from being well-documented, sometimes wrong, and their semantics is complex [2]. Of course different machines have different semantics of concurrency so programs and program verification methods are no longer portable.

This problem is often solved by the verification community by considering sequential consistency [44] (which was valid with classical memory models [20] but no longer on modern architectures [6, 31]), or by putting severe restrictions on acceptable implementations (such as multi-programming on a single processor [50]) with potential terrible slowdowns (potentially by factors of tenth of thousands) resulting from naïve compilation.

Soundness
Soundness refers to the possibility of proving that any automatically verified program specification for any program of a programming language is valid with respect to a formal definition of the semantics of the programming language. This does not necessarily mean absolute correctness when putting restrictions (such as no parameter aliasing in Clousot [10] or soundness up to the first error with unpredictable effects in Astrée [23]), preferably machine checkable ones (such as the former restrictions on dynamic memory allocation in Astrée [23], now dropped).

The most common method to circumvent the problem is to ignore it completely [58]. For example unsound static analyzers are easier, indeed much much easier, to develop than sound ones, and they have commercial successes [55], based on heuristic organization of the reported errors. This does not extend to subtle errors and high quality software as required e.g. in aerospace [56] which market is unfortunately much smaller than that of easily analyzable dirty software.

On the contrary, the ultimate objective is that of automatically verified automatic verification methods. Unfortunately, automatic soundness verifications are still is their infancy [9] since the mechanization of abstract interpretation theory is a difficult problem.

Undecidability
Assuming that what is to be verified is well-defined, undecidability is the next challenge. In full generality, mathematical induction, is required to verify programs, and this induction may be non-trivial.
(mathematicians sometimes take centuries to solve apparently simple problems [59]).

First attempts in data-flow analysis [3] considered finite abstractions to get fixpoint equations in finite abstract domains. There was no soundness consideration. Only informal explanations where provided to justify the correctness of the static analysis, without formal connection to a clearly worded program semantics [18].

Unfortunately, finite abstractions, although complete for a given program and a given property [15], are always surpassed by inductive abstraction for a programming language, as proved in [21].

A similar and even more radical and trivial approach has been to consider finite state systems [11, 53] thus avoiding all problems but combinatorial explosion. After enormous but fruitless efforts of the computer aided verification community, this simplistic approach is progressively recognized as a useful toy with no hope of scalability. The simplest solution is then to claim that scalability is unnecessary. This is a bit in contradiction with the software everywhere where perspectives of modern societies.

Another easy approach to undecidability, is to completely forget about termination. This is the case a.o. of finite abstraction refinements [12], even when refined to avoid useless refinements [26]. There is even a best way to refine abstractions for a given program and a property [34, 35] which has never been exploited in practice, probably because it is more complicated and does not terminate either. Most engineers are patient and can wait for hours or days if the results are useful, provided they know an upper bound on how long they have to wait. Of course time-out, which is a too trivial widening, is unsatisfactory. Moreover, refinement is based on the idea than the first imprecise analyses may be able to prove results quickly without resorting to complex abstract domains. This might work on small programs, although the difference is not significative, but has not been shown to scale to very large programs which inevitably have their intricacies requiring complex abstractions [32] beyond naïve refinements towards popular abstract domains e.g. [27, 49].

A refinement per false alarm will hardly scale.

Undecidability can also be handled by looking for specific decidable cases. The most trivial case is programs without loops. Decidability requires restrictions on the properties (such as no type union in ML [14]) and the programming language (such as both branches of a conditional must have the same type [14]). For more general properties, the problem is that the considered cases are almost never applicable in practice and, despite that, the algorithms are of very high complexity so do not scale.

When all these biases will have to be too limited in practice, the next step is to use mathematical induction.

Mathematical induction

Induction is the basic mathematical method for reasoning on computer systems, most often on their structure (like structural induction in operational semantics [52]) or the progress of the computations [33], or inversely [41], or both [38]). Apart from the fact that program properties may not be computer-representable and that implication is undecidable (which is solved by considering abstract domains in SMT solvers [25, 28, 30] or by using proof checkers [51] instead of theorem provers [42]), the main problem is to find an inductive argument for the proof by induction, which is nothing but an infinite fixpoint computation [18]. Convergence must definitely be accelerated (unless the solution can be computed symbolically once and for all, in which case the widening just plugs in the appropriate solution [36]).

The first work-around is to ask the programmer to provide the inductive argument (such as all loop invariants [45]). This is a trivial widening to the proposed solution and to unknown if it is not valid. This is great for tiny programs but more difficult for larger ones, in particular because the inductive argument is often much larger than the program itself. This is also costly at the development time (even for projects that greatly simplify real life software [9, 40]). This is even more costly in the maintenance phase since any modification of the program requires a manual modification of the corresponding inductive argument (and even of the proof itself for interactive verification).

The second work-around is to ask the programmer not for the inductive argument but for a pattern of the inductive argument [37]. This always works when the pattern is per program since the limit of this approach is to ultimately provide the inductive argument itself. This can become rapidly very complex when the pattern is general enough (e.g. for non-linear invariants [16]) and very large programs. Moreover this does the solve the analysis termination problem so that convergence acceleration is necessary [4].

Abstract mathematical induction

Abstract domains generalize patterns of properties essentially by providing infinitely many patterns. Infinite abstract domains require convergence acceleration by extrapolation (widening/dual-widening) and interpolation operators (narrowing, dual-narrowing, which are equivalent up to the exchange of the parameters) to automate mathematical induction in the abstract domain [17, 27].

Trivial extrapolations include bounding computations [43] or abstracting to finite domains [39]. This is subject to the finite abstraction limitations [21]. It cannot be used to prove anything, but the presence of bugs at the beginning of the considered program executions.

The power of extrapolation/interpolation operators lies in the ability to enforce convergence in infinitely many different ways for infinitely many different programs.

In general, an increasing iterative static analysis using extrapolation of successive iterates by widening followed by a decreasing iterative static analysis using interpolation of successive iterates by narrowing (both bounded by the specification) can be further improved by a increasing iterative static analysis using interpolation of iterates with the specification by dual-narrowing until reaching a fixpoint and checking whether it is inductive for the specification. This can be done locally [5]. The soundness and termination properties are independent [13]. An example is dual-narrowing which generalizes Craig interpolation in first order logic pre-ordered by implication to arbitrary abstract domains and may not terminate.

Conclusion

A large fraction of research on program verification has been devoted to avoiding hard and difficult fundamental problems. We know, by undecidability, that full verification will always be automatically unsolvable. So the game is to anticipate how far the limit can be pushed. From what we learned from the past, we anticipate that abstract interpretation will be able to explain how it works [22]. It is another challenge to find what will work.

Moreover predicting the future is well-known to be error-prone. For example, [13] concluded that “Les techniques de mise au point encore largement utilisées dans l’industrie informatique du logiciel peuvent être en partie évitées (du moins pour les fautes de programmation si ce n’est pour les fautes de conception), en utilisant nos propositions d’analyse sémantique automatique des programmes, et ce, sans attendre les dix ans (ou plus) qui seront nécessaires pour que les techniques de vérification des programmes utilisant des démonstrateurs de théorèmes soient opérationnelles. Il est d’ailleurs certain que les méthodes que nous proposons sont complémentaires et offrent pour certains types d’analyses un rapport coût/bénéfice
très rentable.”. Replacing years by decades might still be a valid conclusion.

References


