Translating Scala Programs to Isabelle/HOL
System Description

Lars Hupel[1] and Viktor Kuncak[2]

[1] Technische Universität München
[2] École Polytechnique Fédérale de Lausanne (EPFL)

Abstract. We present a trustworthy connection between the Leon verification system and the Isabelle proof assistant. Leon is a system for verifying functional Scala programs. It uses a variety of automated theorem provers (ATPs) to check verification conditions (VCs) stemming from the input program. Isabelle, on the other hand, is an interactive theorem prover used to verify mathematical specifications using its own input language Isabelle/Isar. Users specify (inductive) definitions and write proofs about them manually, albeit with the help of semi-automated tactics. The integration of these two systems allows us to exploit Isabelle's rich standard library and give greater confidence guarantees in the correctness of analysed programs.

Keywords: Isabelle, HOL, Scala, Leon, compiler

1 Introduction

This system description presents a new tool that aims to connect two important worlds: the world of interactive proof assistant users who create a body of verified theorems, and the world of professional programmers who increasingly adopt functional programming to develop important applications. The Scala language (`www.scala-lang.org`) enjoys a prominent role today for its adoption in industry, a trend most recently driven by the Apache Spark data analysis framework (to which, e.g., IBM committed 3500 researchers recently [16]). We hope to introduce some of the many Scala users to formal methods by providing tools they can use directly on Scala code. Leon system (`http://leon.epfl.ch`) is a verification and synthesis system for a subset of Scala [2, 10]. Leon reuses the Scala compiler's parsing and type-checking frontend and subsequently derives verification conditions to be solved by the automated theorem provers, such as *Z3* [13] and *CVC4* [1]. Some of these conditions arise naturally upon use of particular Scala language constructs (e.g. completeness for pattern matching), whereas others stem from Scala assertions (`require` and `ensuring`) and can naturally express universally quantified conjectures about computable functions.

Interactive proof assistants have long contained functional languages as fragments of the language they support. Isabelle/HOL [14,20] offers definitional facilities for functional programming, e.g. the **`datatype`** command for inductive data types and **`fun`** for recursive functions. A notable feature of Isabelle is its code generator: certain executable specifications can be translated into source code in target languages such as ML, Haskell, Scala, OCaml [5, 7]. Yet many Scala users do not know Isabelle today.

Aiming to bring the value of trustworthy formalized knowledge to many programmers familiar with Scala, we introduce a mapping in the opposite direction: instead of generating code from logic, we show how to map programs in the purely functional fragment of Scala supported by Leon into Isabelle/HOL. We use Isabelle's built-in tactics to discharge the verification conditions. Compared to use of automated solvers in Leon alone, the connection with Isabelle has two main advantages:

1. Proofs in Isabelle, even those generated from automated tactics, are justified by a minimal inference kernel. In contrast to ATPs, which are complex pieces of software, it is far less likely that a kernel-certified proof is unsound.
2. Isabelle's premier logic, HOL, has seen decades of development of rich mathematical libraries and formalizations such as Archive of Formal Proofs. Proofs carried out in Isabelle have access to this knowledge, which means that there is a greater potential for reuse of existing developments.

Establishing the formal correspondence means embedding Scala in HOL, requiring non-trivial transformations (§2). We use a *shallow embedding*, that is, we do not model Scala's syntax, but rather perform a syntactic mapping from Scala constructs to their equivalents in HOL. For our implementation we developed an idiomatic Scala API for Isabelle based on previous work by Wenzel [18, 21] (§3). We implemented as much functionality as possible inside Isabelle to leverage checking by Isabelle's proof kernel. The power of Isabelle's tactics allows us to prove more conditions than what is possible with the Z3 and CVC4 backends (§4). We are able to import Leon's standard library and a large amount of its example code base into Isabelle (§5), and verify many of the underlying properties.

Contribution We contribute a mechanism to import functional Scala code into Isabelle, featuring facilities for embedding Isabelle/Isar syntax into Scala via Leon and reusing existing constants in the HOL library without compromising soundness. This makes Isabelle available to Leon as a drop-in replacement for Z3 or CVC4 to discharge verification conditions. We show that Isabelle automation is already useful for processing such conditions.

Among related works we highlight a Haskell importer for Isabelle [6], which also uses a shallow embedding and has a custom parser for Haskell, but does not perform any verification. Breitner et. al. have formalised "large parts of Haskell's standard prelude" in Isabelle [4]. They use the HOLCF logic, which is extension on HOL for domain theory, and have translated library functions manually. Mehnert [12] implemented a verification system for Java in Coq using separation logic.

In the following text, we are using the term "Pure Scala" to refer to the fragment of Scala supported by Leon [2, §3], whereas "Leon" denotes the system itself. More information about Leon and Pure Scala is available from the web deployment of Leon at `http://leon.epfl.ch` in the Documentation section.

2 Bridging the gap

Isabelle is a general specification and proof toolkit with the ability of functional programming in its logic Isabelle/HOL. Properties of programs need to be stated and

```
sealed abstract class List[A]
case class Cons[A](head: A, tail: List[A]) extends List[A]
case class Nil[A]() extends List[A]

def size[A](l: List[A]): BigInt = (l match {
  case Nil => BigInt(0)
  case Cons(_, xs) => 1 + size(xs)
}) ensuring(_ >= 0)
```

(a) Pure Scala version

```
datatype 'a list = Nil | Cons 'a "'a list"

fun size :: "'a list => int" where
"size Nil = 0" |
"size (Cons _ xs) = 1 + size xs"

lemma "size xs >= 0" by (induct xs) auto
```

(b) Isabelle version

Fig. 1: Example programs: Linked lists and a size function

proved explicitly in an interactive IDE. While the system offers *proof tactics*, the order in which they are called and their parameters need to be specified by the user. Users can also write custom tactics which deal with specific classes of problems.

Leon is more specialised to verification of functional programs and runs in batch mode. The user annotates a program and then calls Leon which attempts to discharge resulting verification conditions using ATPs. If that fails, the user has to restructure the program. Leon has been originally designed to be fully automatic; consequently, there is little support for explicitly guiding the prover. However, because of its specialisation, it can leverage more automation in proofs and counterexample finding on first-order recursive functions.

Due to their differences, both systems have unique strengths. Their connection allows users to benefit from this complementarity.

2.1 Language differences

Both languages use different styles in how functional programs are expressed. Figure 1 shows a direct comparison of a simple program accompanied by a (trivial) proof illustrating the major differences:

- Pure Scala uses an object-oriented encoding of algebraic data types (*sealed classes* [15]), similar to Java or C#. Isabelle/HOL follows the ML tradition by having direct syntax support [3].
- (Pre-) and postconditions in Leon are annotated using the `ensuring` function, whereas Isabelle has a separate `lemma` command. In a sense, verification conditions in Leon are "inherent", but need to be stated manually in Isabelle.

– Pure Scala does not support top-level pattern matching (e.g. $rev\ (x : xs) = \ldots$).

The translation of data types and terms is not particularly interesting because it is mostly a cavalcade of technicalities and corner cases. However, translating functions and handling recursion poses some interesting theoretical challenges.

2.2 Translating functions

A *theory* is an Isabelle/Isar source file comprising a sequence of definitions and proofs, roughly corresponding to the notion of a "module" in other languages. Theory developments are strictly monotonic. Cyclic dependencies between definitions are not allowed [11], however, a definition may consist of multiple constants. In Pure Scala, there are no restrictions on definition order and cyclicity.

Consequentially, the Isabelle integration has to first compute the dependency graph of the functions and along with it the set of strongly connected components. A single component contains a set of mutually-recursive functions. Collapsing the components in the graph then results in a directed acyclic graph which can be processed in any topological ordering.

The resulting function definitions are not in idiomatic Isabelle/HOL style; in particular, they are not useful for automated tactics. Consider Figure 1: the naive translation would produce a definition $size\ xs = \texttt{case}\ xs\ \texttt{of}\ y \# ys \rightarrow \ldots size\ ys \ldots$ Isabelle offers a generic term rewriting tactic (the *simplifier*), which is able to substitute equational rules. Such a rule, however, constitutes a non-terminating simplification chain, because the right-hand side contains a subterm which matches the left-hand side.

This can be avoided by splitting the resulting definition into cases that use Haskell-style top-level pattern matching. A verified routine to perform this translation is integrated into Isabelle, producing terminating equations which can be used by automated tactics. From this, we also obtain a better induction principle which can be used in subsequent proofs.

When looking at the results of this procedure, the example in Figure 1 is close to reality. The given Pure Scala input program produces almost exactly the Isabelle theory below, modulo renaming. Because of our implementation, the user normally does not see the resulting theory file (see §3). However, for this example, the internal constructions we perform are roughly equivalent to what Isabelle/Isar would perform (see §5).

2.3 Recursion

Leon has a separate termination checking pass, which can run along with verification and can be turned off. Leon's verification results are only meant to be valid under the assumption that its termination checker succeeded (i.e. ensuring partial correctness).

Isabelle's proof kernel does not accept recursive definitions at all. We use the *function package* by Krauss [9] to translate a set of recursive equations into a low-level, non-recursive definition. To automate this construction, the package provides a **fun**

command which can be used in regular theories (see Figure 1), but also programmatically. To justify its internal construction against the kernel, it needs to prove termination. By default, it searches for a lexicographic ordering involving some subset of the function arguments.

This also means that when Leon is run using Isabelle, termination checking is no longer independent of verification, but rather "built in". Krauss' package also supports user-specified termination proofs. In the future, we would like to give users the ability to write those in Scala.

A further issue is recursion in data types. Negative recursion can lead to unsoundness, e.g. introducing non-termination in non-recursive expressions. While Leon has not implemented a wellformedness check yet, Isabelle correctly rejects such data type definitions. Because we map Scala data types syntactically, we obtain this check for free when using Isabelle in Leon.

2.4 Cross-language references

One of the main reasons why we chose a shallow embedding of Pure Scala into Isabelle is the prospect of reusability of Isabelle theories in proofs of imported Pure Scala programs. For example, the dominant collection data structure in functional programming – and by extension both in Pure Scala and Isabelle/HOL – are lists. Both languages offer dozens of library functions such as `map`, `take` or `drop`. Isabelle's `List` theory also contains a wealth of theorems over these functions. All of the existing theorems can be used by Isabelle's automated tactics to aid in subsequent proofs, and are typically unfolded automatically by the simplifier.

However, when importing Pure Scala programs, all its data types and functions are defined again in a runtime Isabelle theory. While the imported `List.map` function may end up having the same shape as HOL's `List.map` function, they are nonetheless distinct constants, rendering pre-existing theorems unusable.

The naive approach of annotating Pure Scala's `map` function to not be imported and instead be replaced by HOL's `map` function is unsatisfactory: The user would need to be trusted to correctly annotate Pure Scala's library, negatively impacting correctness. Hence, we implemented a hybrid approach: We first import the whole program unchanged, creating fresh constants. Later, for each annotated function, we try to prove an equivalence of the form $f' = f$ where f' is the imported definition and f is the existing Isabelle library function, and register the resulting theorem with Isabelle's automated tools. This establishes a trustworthy relationship between the imported Pure Scala program and the existing Isabelle libraries.

Depending on the size of the analysed program (including dependencies), this approach turns out to be rather inefficient.[3] According to Leon conventions, we introduced a flag which skips the equivalence proofs for Pure Scala library functions and just asserts the theorems as axioms. This also alleviates another practical problem: not all desired equivalences can be proven automatically by Isabelle. Support for specifying manual equivalence proofs would be useful, but is not yet implemented.

[3] Because our implementation uses Isabelle in interactive instead of in batch mode, we cannot produce pre-computed heap images to be loaded for later runs.

3 Technical considerations

Isabelle has been smoothly integrated into Leon by providing an appropriate instance of a *solver*. In that sense, Isabelle acts as "yet another backend" which is able to check validity of a set of assertions.

3.1 Leon integration

A solver in Leon terminology is a function checking the consistency of a set of assumptions. A pseudo-code type signature could be given as $\mathcal{P}(\mathcal{F}) \rightarrow \{\texttt{sat}, \texttt{unsat}, \texttt{unknown}\}$, where $\mathcal{F}$ is the set of supported formulas. According to program verification convention, a result of `unsat` means that no contradiction could be derived from the assumptions, i.e. that the underlying program is correct. If a solver however returns `sat`, it is expected to produce a counterexample which violates verification conditions, e.g. a value which is not matched by any clause in a pattern match.

The Isabelle integration is exactly such a function, but with the restriction that it never returns `sat`, because a failed proof attempt does not produce a suitable counterexample. Since Leon offers a sound and complete counterexample procedure for higher-order functions [17], implementing this feature for Isabelle would not be useful.

3.2 Process communication

Communication between the JVM process running Leon and the Isabelle process works via our *libisabelle* library which extends Wenzel's PIDE framework [19, 21] to cater to non-IDE applications. It introduces a remote procedure call layer on top of PIDE, reusing much of its functionality. Leon is then able to update and query state stored in the prover process. Procedure calls are typed and asynchronous, using an implementation of type classes in ML and Scala's *future* values by Haller et al. [8], respectively.

While being a technologically more complicated approach, it offers benefits over textual Isabelle/Isar source generation. Most importantly, because communication is typed, the implementation is much more robust. Common sources of errors, e.g. pretty printing of Isabelle terms or escaping, are completely eliminated.

4 Example

Figure 2 shows a fully-fledged example of an annotated Pure Scala program. As background, assume the `List` definition from the previous example enriched with some standard library functions, a `Nat` type, and a `listSum` function.[4] The functions in the example are turned into lemma statements in Isabelle. The string parameter of the `proof` annotation is an actual Isar method invocation, that is, it is interpreted by the Isabelle system. For hygienic purposes, names of Pure Scala identifiers are not preserved during translation, but suffixed with unique numbers. To allow users to refer back to syntactic entities using their original names, the `<var _>` syntax has been introduced.

[4] The full example is available at `https://git.io/vznVH`.

```
def sumReverse[A](xs: List[Nat]) =
  (listSum(xs) == listSum(xs.reverse)).holds

@proof(method = """(induct "<var xs>", auto)""")
def sumConstant[A](xs: List[A], k: Nat) =
  (listSum(xs.map(_ => k)) == length(xs) * k).holds

@proof(method = "(clarsimp, induct rule: list_induct2, auto)")
def mapFstZip[A, B](xs: List[A], ys: List[B]) = {
  require(length(xs) == length(ys))
  xs.zip(ys).map(_._1)
} ensuring { _ == xs }
```

Fig. 2: Various induction proofs about lists

Running Leon with the Isabelle solver on this example will show that all conditions hold. The first proof merely reuses a lemma which is already in the library. The other two need specific guidance, i.e. an annotation, for them to be accepted by the system. The proofs involve Isabelle library theorems, for example distributivity of $(+, *)$ on natural numbers. For comparison, Leon+Z3 cannot prove any proposition. When also instructed to perform induction, it can prove `sumConstant`. (Same holds for Leon+CVC4.) There is currently no way in Leon to concisely specify the use of a custom induction rule for Z3 (or CVC4) as required by the last proposition (simultaneous induction over two lists of equal length).

This example also demonstrates another instance of the general Isabelle philosophy of *nested languages:* Pure Scala identifiers may appear inside Isar text which appears inside Pure Scala code. Further nesting is possible because Isabelle text can itself contain nested elements (e.g. ML code, ...).

5 Evaluation

In this section, we discuss implementation coverage of Pure Scala's syntactic constructs, trustworthiness of the translation and overall performance.

Coverage. The coverage of the translation is almost complete. A small number of Leon primitives, among them array operations have not been implemented yet.[5] All other primitives are mapped as closely as possible and adaptations to Isabelle are proven correct when needed. Leon's standard library contains – as of writing – 177 functions with a total of 289 verification conditions, out of which Isabelle can prove 206 ($\approx 71\%$).

Trustworthiness. Our mapping uses only definitional constructs of Isabelle and thus the theorems it proves have high degree of trustworthiness. Using a shallow embedding always carries the risk of semantics mismatches. A concern is that since the translation

[5] In fact, while attempting to implement array support we discovered that Leon's purely functional view of immutably used arrays does not respect Scala's reference equality implementation of arrays, leading to a decision to disallow array equality in Leon's Pure Scala.

of Pure Scala to Isabelle works through an internal API, the user has no possibility to convince themselves of the correctness of the implemented routines short of inspecting the source code. For that reason, all operations are logged in Isabelle. A user can request a textual output of an Isar theory file corresponding to the imported Pure Scala program, containing all definitions and lemma statements, but no proofs. This file can be inspected manually and re-used for other purposes, and represents faithfully the facts that Isabelle actually proved in a readable form.

Performance. On a contemporary dual-core laptop, just defining all data types from the Pure Scala library (as of writing: 13), but no functions or proofs, Leon+Isabelle takes approximately 30 seconds. Defining all functions adds another 70 seconds to the process. Using Leon+Z3, this is much faster: it takes less than 10 seconds. The considerable difference (factor ≈ 10) can be explained by looking at the internals of the different backends. Z3 has data types and function definitions built into its logic. Isabelle itself does not: both concepts are implemented in HOL, meaning that every definition needs to be constructed and justified against the proof kernel. The processing time of an imported Pure Scala programs is comparable to that of a hand-written, idiomatic Isabelle theory file. In fact, during processing the Pure Scala libraries, thousands of messages are passed between the JVM and the Isabelle process, but the incurred overhead is negligible compared to the internal definitional constructions.

6 Conclusion

We have implemented an extension to Leon which allows using Isabelle to discharge verification conditions of Pure Scala programs. Because it supports the vast majority of syntax supported by Leon, we consider it to be generally usable. It is incorporated in the Leon source repository,[6] supporting the latest Isabelle version (Isabelle2016).

With this work, it becomes possible to co-develop a specification in both Pure Scala and Isabelle, use Leon to establish a formal correspondence, and prove interesting results in Leon and/or Isabelle/Isar. Because of the embedded Isar syntax, complicated correctness proofs can also be expressed concisely in Leon. To the best of our knowledge, this constitutes the first bi-directional integration between a widespread general purpose programming language and an interactive proof assistant.

An unintended consequence is that since Isabelle can export code in Haskell and now import code from Pure Scala, there is a fully-working Scala-to-Haskell cross-compilation pipeline. The transformations applied to the Pure Scala code to make it palatable to Isabelle's automation also results in moderately readable Haskell code.

Acknowledgements We would like to thank the people who helped "making the code work": Ravi Kandhadai, Etienne Kneuss, Manos Koukoutos, Mikäel Mayer, Nicolas Voirol, Makarius Wenzel. Cornelius Diekmann, Manuel Eberl, and Tobias Nipkow suggested many textual improvements to this paper.

[6] `https://github.com/epfl-lara/leon`

References

1. Barrett, C., Conway, C.L., Deters, M., Hadarean, L., Jovanović, D., King, T., Reynolds, A., Tinelli, C.: CVC4. In: Gopalakrishnan, G., Qadeer, S. (eds.) Computer Aided Verification, Lecture Notes in Computer Science, vol. 6806, pp. 171–177. Springer (2011)
2. Blanc, R.W., Kneuss, E., Kuncak, V., Suter, P.: An overview of the Leon verification system: Verification by translation to recursive functions. In: Scala Workshop (2013)
3. Blanchette, J.C., Hölzl, J., Lochbihler, A., Panny, L., Popescu, A., Traytel, D.: Truly modular (co)datatypes for Isabelle/HOL. In: Interactive Theorem Proving, Lecture Notes in Computer Science, vol. 8558, pp. 93–110. Springer International Publishing (2014)
4. Breitner, J., Huffman, B., Mitchell, N., Sternagel, C.: Certified HLints with Isabelle/HOLCF-Prelude (Jun 2013), Haskell And Rewriting Techniques (HART)
5. Haftmann, F.: Code Generation from Specifications in Higher-Order Logic. Ph.D. thesis, Technische Universität München (2009)
6. Haftmann, F.: From higher-order logic to Haskell: there and back again. In: Proceedings of the 2010 ACM SIGPLAN workshop on Partial evaluation and program manipulation. pp. 155–158. ACM (2010)
7. Haftmann, F., Nipkow, T.: Code generation via higher-order rewrite systems. In: Blume, M., Kobayashi, N., Vidal, G. (eds.) Functional and Logic Programming: 10th International Symposium: FLOPS 2010. Lecture Notes in Computer Science, vol. 6009. Springer (2010)
8. Haller, P., Prokopec, A., Miller, H., Klang, V., Kuhn, R., Jovanovic, V.: Futures and promises (2012), `http://docs.scala-lang.org/overviews/core/futures.html`
9. Krauss, A.: Partial and nested recursive function definitions in higher-order logic. Journal of Automated Reasoning 44(4), 303–336 (2009)
10. Kuncak, V.: Developing verified software using Leon. In: NASA Formal Methods - 7th International Symposium, NFM 2015, Pasadena, CA, USA, April 27-29, 2015, Proceedings. pp. 12–15 (2015)
11. Kunčar, O.: Correctness of Isabelle's cyclicity checker: Implementability of overloading in proof assistants. In: Proceedings of the 2015 Conference on Certified Programs and Proofs. pp. 85–94. CPP '15, ACM, New York, NY, USA (2015)
12. Mehnert, H.: Kopitiam: Modular incremental interactive full functional static verification of java code. In: NASA Formal Methods: Third International Symposium, NFM 2011. pp. 518–524. Springer, Berlin, Heidelberg (2011)
13. de Moura, L., Bjørner, N.: Z3: An efficient SMT solver. In: Tools and Algorithms for the Construction and Analysis of Systems, pp. 337–340. Springer (2008)
14. Nipkow, T., Klein, G.: Concrete Semantics. Springer (2014)
15. Odersky, M., Spoon, L., Venners, B.: Programming in Scala, Second Edition. Artima Inc (2010)
16. Terdoslavich, W.: IBM Bets On Apache Spark As 'The Future Of Enterprise Data' (Jun 2015), `http://www.informationweek.com/big-data/ibm-bets-on-apache-spark-as-the-future-of-enterprise-data/d/d-id/1320855`
17. Voirol, N., Kneuss, E., Kuncak, V.: Counter-example complete verification for higher-order functions. In: Scala Symposium (2015)
18. Wenzel, M.: Isabelle as document-oriented proof assistant. In: Conference on Intelligent Computer Mathematics/Mathematical Knowledge Management (2011)
19. Wenzel, M.: Isabelle/jEdit – a Prover IDE within the PIDE framework. In: Intelligent Computer Mathematics, pp. 468–471. Springer (2012)
20. Wenzel, M.: The Isabelle/Isar Reference Manual (2013)
21. Wenzel, M.: Asynchronous user interaction and tool integration in Isabelle/PIDE. In: Klein, G., Gamboa, R. (eds.) Interactive Theorem Proving, Lecture Notes in Computer Science, vol. 8558, pp. 515–530. Springer International Publishing (2014)