Pallene: A statically typed companion language for Lua

Hugo Musso Gualandi
PUC-Rio
hgualandi@inf.puc-rio.br

Roberto Ierusalimschy
PUC-Rio
roberto@inf.puc-rio.br

ABSTRACT

The simplicity and flexibility of dynamic languages make them popular for prototyping and scripting, but the lack of compile-time type information makes it very challenging to generate efficient executable code.

Inspired by ideas from scripting, just-in-time compilers, and optional type systems, we are developing Pallene, a statically typed companion language to the Lua scripting language. Pallene is designed to be amenable to standard ahead-of-time compilation techniques, to interoperate seamlessly with Lua (even sharing its runtime), and to be familiar to Lua programmers.

In this paper, we compare the performance of the Pallene compiler against LuaJIT, a just in time compiler for Lua, and with C extension modules. The results suggest that Pallene can achieve similar levels of performance.

CCS CONCEPTS

• **Software and its engineering → Scripting languages**; *Just-in-time compilers*; *Dynamic compilers*; *Imperative languages*;

ACM Reference Format:
Hugo Musso Gualandi and Roberto Ierusalimschy. 2018. Pallene: A statically typed companion language for Lua. In *Proceedings of XXII Brazilian Symposium on Programming Languages (SBLP).* ACM, New York, NY, USA, 8 pages.

1 INTRODUCTION

The simplicity and flexibility of dynamic languages make them popular for prototyping and scripting, but the lack of compile-time type information makes it very challenging to generate efficient executable code. There are at least three approaches to improve the performance of dynamically typed languages: scripting, just-in-time compilation, and optional type systems.

The scripting approach [19] advocates the use of two separate languages to write a program: a low-level *system language* for the parts of the program that need good performance and that interact with the underlying operating system, and a high-level *scripting language* for the parts that need flexibility and ease of use. Its main advantage is that the programmer can choose the programming language most suited for each particular task. The main disadvantages are due to the large differences between the languages. That makes it hard to rewrite a piece of code from one language to the other and also adds run-time overhead in the API between the two language runtimes.

Just-in-time compilers [6] dynamically translate high level code into low-level machine code during the program execution, on demand. To maximize performance, JIT compilers may collect run-time information, such as function parameter types, and use that information to generate efficient specialized code. The most appealing aspect of JIT compilers for dynamic languages is that they can provide a large speedup without the need to rewrite the code in a different language. In practice, however, this speedup is not always guaranteed as programmers often need to rewrite their code anyway, using idioms and "incantations" that are more amenable to optimization [10].

As the name suggests, optional type systems allow programmers to partially add types to programs.[1] These systems combine the static typing and dynamic typing disciplines in a single language. From static types they seek better compile-time error checking, machine-checked lightweight documentation, and run-time performance; from dynamic typing they seek flexibility and ease of use. One of the selling points of optional type systems is that they promise a smooth transition from small dynamically typed scripts to larger statically typed applications [27]. The main challenge these systems face is that it is hard to design a type system that is at the same time simple, correct, and amenable to optimizations.

Both scripting and optional types assume that programmers need to restrict the dynamism of their code when they seek better performance. Although in theory JIT compilers do not require this, in practice programmers also need to restrict themselves to achieve maximum performance. Realizing how these self-imposed restrictions result in the creation of vaguely defined language subsets, and how restricting dynamism seems unavoidable, we asked ourselves: what if we accepted the restrictions and defined a new programming language based on them? By focusing on this "well behaved" subset and making it explicit, instead of trying to optimize or type a dynamic language in its full generality, we would be able to drastically simplify the type system and the compiler.

To study this question, we are developing the Pallene programming language, a statically typed companion to Lua. Pallene is intended to act as a system language counterpart to Lua's scripting, but with better interoperability than existing static languages. To avoid the complexity of a JIT compilation, Pallene should be amenable to standard ahead-of-time compiler optimization techniques. To minimize the run-time mismatch, Pallene should use Lua's data representation and garbage collector. To minimize the conceptual mismatch, Pallene should be familiar to Lua programmers, syntactically and semantically.

In the next section of this paper, we overview the main approaches currently used to tackle the performance problems of

[1] Gualandi [9] presents a historical review of type systems for dynamic languages.

SBLP, September 2018, São Carlos, São Paulo, Brazil

ACM ISBN 978-1-4503-6480-5.

dynamic languages. In Section 3, we describe how we designed Pallene, aiming to combine desirable properties from those approaches. In Section 4, we discuss how our goals for Pallene affected its implementation. In Section 5, we evaluate the performance of our prototype implementation of Pallene on a set of micro benchmarks, comparing it with the reference Lua implementation and with LuaJIT [20], a state of the art JIT compiler for Lua. This evaluation suggests that it is possible to produce efficient executable code for Pallene programs. In the last two sections of this paper we compare Pallene with related research in type systems and optimization for dynamic languages, and we discuss avenues for future work.

2 OPTIMIZING SCRIPTING LANGUAGES

In this section we discuss the existing approaches to optimizing dynamic languages that we mentioned in the introduction.

2.1 Scripting

One way to overcome the slowness of dynamic languages is to avoid them for performance-sensitive code. Dynamic scripting languages are often well-suited for a multi-language architecture, where a statically typed low-level *system language* is combined with a flexible dynamically typed *scripting language*, a style of programming that has been championed by John Ousterhout [19].

Lua has been designed from the start with scripting in mind [13] and many applications that use Lua follow this approach. For instance, a computer game like Grim Fandango has a basic engine, written in C++, that performs physical simulations, graphics rendering, and other machine intensive tasks. The game designers, who are not professional programmers, wrote all the game logic in Lua [18].

The main advantages of this scripting architecture are its pragmatism and its predictability. Each language is used where it is more adequate and the software architect can be relatively confident that the parts written in the system language will have good performance. The main downside is the conceptual mismatch between the languages.

Rewriting modules from one language to the other is difficult. A common piece of advice when a Lua programmer seeks better performance is to "rewrite it in C", but this is easier said than done. In practice, programmers only follow this advice when the code is mainly about low-level operations that are easy to express in C, such as doing arithmetic and calling external libraries. Another obstacle to this suggestion is that it is hard to estimate in advance both the costs of rewriting the code and the performance benefits to be achieved by the change. Often, the gain in performance is not what one would expect: As we will see in Section 5, the overhead of converting data from one runtime to the other can cancel out the inherent gains of switching to a static language.

2.2 JIT Compilers

Just-in-time (JIT) compilers are the state of the art in dynamic language optimization. A JIT compiler initially executes the program without any optimization, observes its behavior at runtime, and then, based on this, generates highly specialized and optimized executable code. For example, if it observes that some code is always operating on values of type *double*, the compiler will optimistically compile a version of this code that is specialized for that type. It will also insert tests (guards) in the beginning of the code that jump back to a less optimized generic version in case some value is not of type *double* as expected.

JIT compilers are broadly classified as either method-based or trace-based [7], according to their main unit of compilation. In method-based JITs, the unit of compilation is one function or subroutine. In trace-based JITs, the unit of compilation is a linear trace of the program execution, which may cross over function boundaries. Trace compilation allows for a more embedable implementation and is better at compiling across abstraction boundaries. However, it has trouble optimizing programs which contain unpredictable branch statements. For this reason, most JIT compilers now tend to use the method-based approach, with the notable exceptions of LuaJIT [20] and the RPython Framework [5].

JIT compilers detect types at run-time because inferring types at compile type is very hard and usually produces less specific results. Additionally, in many dynamic languages data types are created at run-time and there are no data definition declarations or type annotations for the compiler to take advantage from. As an example, the PyPy authors explicitly mentioned this as one of their main motivations for using JIT technology [23].

Implementing a JIT compiler can be challenging. The most performant JITs depend heavily on non-portable low-level code and are architected around language-specific heuristics. High level JIT development frameworks are still an active research area. There are various promising approaches, such as the *metatracing* of the RPython framework [5] and the partial evaluation strategy of Truffle [29], but so far these have not been able to compete in terms of performance and resource usage with hand-written JITs such as LuaJIT [20], V8 [26] and HHVM [1].

From the point of view of the software developer, the most attractive feature of JIT compilers is that they promise increased performance without needing to modify the original dynamically typed program. However, these gains are not always easy to achieve, because the effectiveness of JIT compiler optimizations can be hard to predict. Certain code patterns, known as *optimization killers*, may cause the whole section they are in to be de-optimized, resulting in a dramatic performance impact. Programmers who seek performance must carefully avoid the optimization killers for the JIT engines they are targeting, by following advice from the official documentation or from folk knowledge [2, 22].

Since there may be an order of magnitude difference in performance between JIT-optimized and unoptimized code, programmers have an incentive to write their programs in a style that is more amenable to optimization. This leads to idioms that are not always intuitive. For example, the LuaJIT documentation recommends caching Lua functions from other modules in a local variable before calling them [21], as is shown in Figure 1. However, for C functions accessed via the foreign function interface the rule is the other way around: functions from the `C` namespace should not be cached in local variables, as shown in Figure 2.

Another example from LuaJIT is the function in Figure 3, which runs into several LuaJIT optimization killers (which the LuaJIT documentation calls "Not Yet Implemented" features). As of LuaJIT 2.1, traces that call string pattern-matching methods such as `gsub` are

```
-- Bad
local function mytan(x)
  return math.sin(x) / math.cos(x)
end

-- Good
local sin, cos = math.sin, math.cos
local function mytan(x)
  return sin(x) / cos(x)
end
```

Figure 1: LuaJIT encourages programmers to cache imported Lua functions in local variables.

```
-- Good (!)
local function hello()
  C.printf("Hello, world!")
end

-- Bad (!)
local printf = C.printf
local function hello()
  printf("Hello, world!")
end
```

Figure 2: Surprisingly, LuaJIT encourages programmers not to cache C functions called through the foreign function interface.

```
function increment_numbers(text)
  return (text:gsub("[0-9]+", function(s)
    return tostring(tonumber(s) + 1)
  end))
end
```

Figure 3: This function cannot be optimized by LuaJIT because it calls the gsub method and because it uses an anonymous callback function.

not compiled into machine code by the JIT. The same is true for traces that create closures or define anonymous functions, even if the anonymous function does not close over any outer variables.

The different coding style that JITs encourage is not the only way they affect the software development process. Programmers also monitor the performance of their programs to verify whether the JIT compiler is actually optimizing their code. When it is not, they resort to specialized debugging tools to discover which optimization killer is the culprit [11]. This may require reasoning at a low level of abstraction, involving the intermediate representation of the JIT compiler or its generated machine code [12].

Another aspect of JIT compilers is that before they can start optimizing they must run the program for many iterations, collecting run-time information. During this initial warmup period the JIT will run only as fast or even slower than a non-JIT implementation. In some JIT compilers the warmup time can also be erratic, or even cyclic, as observed by Barrett et al [3].

2.3 Optional Types

Static types serve several purposes. They are useful for error detection, as a lightweight documentation, and they facilitate efficient code generation. Therefore there are many projects aiming to combine the benefits of static and dynamic typing in a single language.

A recurring idea to help the compiler produce more efficient code is to allow the programmer to add optional type annotations to the program. Compared with a more traditional scripting approach, optional typing promises a single language instead of two different ones, which makes it easier for the static and dynamic parts of the program to interact with each other. The pros and cons of these optional type system approaches vary from case to case, since each type system is designed for a different purpose. For example, the optional type annotations of Common LISP allow the compiler to generate extremely efficient code, but without any safeguards [8].

A research area deserving special attention is Gradual Typing [24], which aims to provide a solid theoretical framework for designing type systems that integrate static and dynamic typing in a single language. However, gradual type systems still face difficulties when it comes to run-time performance. On the one hand, systems that check types as they cross the boundary between the static and dynamic parts of the code are often plagued with a high verification overhead cost [25]. On the other hand, type systems that do not perform this verification give up on being able to optimize the static parts of the program.

One problem with optional types is that, to embrace the ideal of smooth transition between the typed and untyped worlds, the static type system should support common idioms from the dynamic language. This requirement for flexibility usually leads to a more complex type system, making it more difficult to use and, more importantly to us, to optimize. For example, in Typed Lua [16] all arrays of integers are actually arrays of nullable integers. In Lua, out of bound accesses result in `nil`; moreover, to remove an element from a list one has to assign `nil` to its position. Both cases require the type system to accept `nil` as a valid element of the list.

3 THE PALLENE PROGRAMMING LANGUAGE

Although JIT compilers and optional type systems are said to be designed to cover all aspects of their dynamic languages, this is not the case in practice. Normally there is a "well behaved" subset of the language that is more suitable to the optimizer or the type system, and programmers will restrict themselves to this subset to better take advantage of their tools.

For example, programmers targeting LuaJIT will tend to restrict themselves to the subset of Lua that LuaJIT can optimize. Similarly, those using TypeScript (an optional type system for Javascript) will prefer to write programs that can fit inside TypeScript's type system. From a certain point of view, we could say that these programmers are no longer programming in Lua or Javascript, at least as these languages are normally used.

As we already mentioned in the introduction, the realization that programmers naturally restrict themselves to a subset of the

language in the search for better performance led us to think about a new language that makes these restrictions explicit. Our hypothesis is that programmers would be willing to accept these restrictions in exchange for guarantees from the compiler that it will be able to generate good code.

Pallene is intended to be that language. As we described before, our plan for Pallene is that it should be amenable to standard ahead-of-time compiler optimization techniques and be compatible with Lua, not only in terms of the run time and data structures, but also in terms of language semantics and familiarity. In the following paragraphs we describe how these goals affected the design of Pallene.

Pallene should be amenable to standard ahead-of-time compiler optimization techniques. This goal led us not only to make Pallene statically typed, but also statically typed with a simple and conventional type system. The compile-time guarantees afforded by such type system can make it much easier for a compiler to produce efficient code. Naturally, we should expect that this same type-system rigidity that aids the compiler will also restrict the programming idioms available in Pallene, which is why Pallene is designed from the start to be used in conjunction with Lua, following a traditional scripting architecture. Since Lua is available when more flexibility is desired, our intention is that Pallene's type system will only support idioms that can be compiled to efficient code. This includes primitive types like floats and integers, arrays, and records. It excludes dynamic language features such as ad-hoc polymorphism. This design should allow programmers to trust that their Pallene programs will have good performance, which is not the case for compilers that rely on speculative optimizations, as is the case with JIT compilers.

Pallene should share the Lua runtime. To allow for seamless interoperability with Lua, Pallene can directly manipulate Lua data structures and also shares Lua's garbage collector. This should reduce the run-time overhead of communicating with Lua. Other languages must use the Lua–C API to manipulate Lua values and can only reference these values indirectly, through the API's stack and registry [14].

C code, when manipulating Lua values, does not keep direct pointers to Lua objects. Instead, Lua exposes them to C through a stack, known as *the Lua stack*, and C functions refer to Lua objects by integer indexes into the stack. This scheme facilitates accurate garbage collection (live objects are rooted in the stack) and dynamic typing (stack slots can contain Lua values of any type) but introduces some overhead.

Pallene should be familiar to Lua programmers, syntactically and semantically. To make it easier to combine Lua and Pallene in a single system, Pallene is very close to a typed subset of Lua, inspired by optional and gradual typing. For example, Figure 4 shows a Pallene function that computes the sum of an array of floating-point numbers. Other than the type annotations, it is valid Lua code. Semantically, it also behaves exactly like the Lua version, except that this Pallene function will raise a run-time error if Lua calls it with an argument of the wrong type (for example, an array of integers).

```
function sum(xs: {float}): float
  local s: float = 0.0
  for i = 1, #xs do
    s = s + xs[i]
  end
  return s
end
```

Figure 4: A Pallene function for summing numbers in a Lua array.

```
-- Pallene Code:
function add(x: float, y:float): float
  return x + y
end

-- Lua Code:
local big = 18014398509481984
print(add(big, 1) == big)
```

Figure 5: An example illustrating why Pallene avoids automatic type coercions.

This follows the idea behind the Gradual Guarantee [24] of gradual type systems, which states that adding type annotations to a program should not change its behavior except for perhaps introducing run-time type errors. This guarantee means that programmers can rely on their knowledge about Lua when programming in Pallene.

Syntactically speaking, Pallene is almost the same as Lua: all of Lua's control-flow statements are present and work the same in Pallene. Whenever reasonable, we try to ensure that the semantics of Pallene are the same as Lua's, following the Gradual Guarantee. For instance, Pallene does not perform automatic coercions between numeric types, unlike most statically typed languages. Consider the example in Figure 5. In Lua, as in many dynamic languages, the addition of two integers produces an integer while the addition of two floating-point numbers produces another floating point number. If we remove the type annotations from the Pallene function add, and treat it as Lua code, Lua will perform an integer addition and the program will print **`false`**. On the other hand, if Pallene automatically coerced the integer arguments to match the floating-point type annotations, it would perform a floating-point addition and the program would print **`true`**: double-precision floating-point numbers cannot accurately represent $2^{54} + 1$. To avoid this inconsistency, Pallene instead raises a run-time type error, complaining that an integer was passed where a floating-point value was expected.

Pallene's standard library is similar to Lua's but not exactly the same. The most noticeable difference is that some functions are missing, such as those that would require polymorphic types. Other functions are more restricted: for instance, string pattern-matching functions only accept literal (constant) patterns. Furthermore, the Pallene library is immutable and does not support monkey patching, unlike regular Lua libraries. If some Pallene code calls **`math.sin`**,

the compiler knows that the function has not been redefined and generates code that directly calls the sine function from the C standard library. It is worth noticing that programmers concerned with performance most likely already avoid monkey patching.

Summing up, the main difference between Pallene and Lua is that Pallene is statically typed with a simple type system. This mere change restricts several common Lua programming practices, such as functions that return a different number of results depending on their arguments, ad-hoc polymorphism, and heterogeneous collections. On the other hand, this makes Pallene an ordinary imperative language, amenable to standard compiler optimization techniques. The precise details of the type system are not yet fully defined, as Pallene is still evolving. Nevertheless, they would not affect the results presented here.

4 IMPLEMENTING PALLENE

Pallene's compiler is quite conventional. It traverses the syntax tree and emits C code, which is then passed to a C compiler (such as gcc) to produce the final executable. This binary complies with Lua's Application Binary Interface (ABI) so that it can be dynamically loaded by Lua just like other modules written in C.

Leveraging an existing high quality compiler backend keeps the Pallene compiler simple and makes Pallene portable to many architectures. Using C also allows us to reference datatypes and macros defined in Lua's C header files.

The main peculiarity of Pallene compilation is that the code it generates for accessing data-structure fields directly manipulates the Lua data structures, which is not allowed for regular C code. This allows better performance than what is currently possible with C Lua modules as we will show. This style of programming would be dangerous if exposed to C programmers, but in Pallene the compiler is able to guarantee that the invariants of the Lua interpreter are respected.

Because of static typing, Pallene can optimize this code much more than the interpreter. As a striking example, when we write `xs[i]` in Pallene (as happens in Figure 4), the compiler knows that `xs` is an array and `i` is an integer, and generates code accordingly. The equivalent array access in Lua (or in C code using the C–API) would need to account for the possibility that `xs` or `i` could have different types or even that `xs` might actually be an array-like object with an `__index` metamethod.

From the point of view of garbage collection, Pallene's direct-manipulation approach is closer to the APIs of Python and Ruby, which expose pointers to objects from the scripting language. However, since Pallene is a high level programming language, it can come with an accurate garbage collector. In Python the programmer is tasked with manually keeping track of reference counts and in Ruby the garbage collector is conservative regarding local variables in the C stack.

One of the few points where the resulting C code is not a direct translation of the corresponding Pallene code regards garbage collection. Pallene currently uses *lazy pointer stacks* [15] to interact with the garbage collector. Collectable Lua values in Pallene are represented as regular C pointers, which at runtime will be stored in machine registers and the C stack, with low overhead. At locations in the program where the garbage collector is invoked, Pallene saves all the necessary Lua pointers in the Lua stack, so that the garbage collector can see them.

Another situation where Pallene does not directly match a conventional static language is run-time type checking. Pallene must insert run-time type checks in the frontier between its statically typed code and Lua's dynamically typed code, in order to guarantee type safety. Currently, this means function calls (when a Lua function calls a Pallene one, or when a Lua function returns to Pallene) and data-structure reads (Lua might have written values of the wrong type to the field). Note that these type checks can only raise errors and do not affect the semantics and the code in any other way.

We avoided using wrappers to implement our type tests because obtaining good performance in a system with wrappers is challenging. In the worst case they can slow down a program by a factor of more than ten, as shown by Takikawa et al [25] and avoiding this overhead is still an open problem. Another problem with wrappers is that they interact poorly with reference equality (object identity). Instead of wrappers, we use a more lazy system of type checks, similar to the tag checks that JITs insert to guard their specialized code or to the *transient semantics* for Reticulated Python [28].

5 PERFORMANCE EVALUATION

We want to verify whether Pallene is competitive with JIT compilers and we want to verify whether bypassing the Lua–C API can lead to relevant performance gains. We are not looking to prove that Pallene can beat JITs—we are primarily concerned with whether Pallene is a viable alternative to JIT compilation for the optimization of Lua programs.

We prepared a small suite of benchmarks to evaluate the performance of Pallene. The first benchmark is a prime sieve algorithm. The second is a matrix multiplication (using Lua arrays). The third one solves the N-queens problem. The fourth one is a microbenchmark that simulates a binary search. The fifth one is a cellular automaton simulation for Conway's Game of Life.

All the benchmarks were prepared specifically for this study. For the algorithms with at least quadratic running time we ran a single iteration with a suitably large N. For the prime sieve and binary search, we repeated the computation inside a loop to obtain a measurable time. As expected, all benchmarks run the same in Pallene and Lua: the source code is identical, except for type annotations, and they produce the same results. Most of the code is also what one would naturally write in Lua. One exception was in the code for the N-queens benchmark, where we used if-then-else statements in places where in Lua it would be more idiomatic to use the idiom `x` **`and`** `y` **`or`** `z` as a ternary operator. (Currently Pallene only supports using **`and`** and **`or`** with boolean operands.) For the matrix multiplication and the cellular automaton benchmarks, we manually hoisted some loop-invariant array operations. LuaJIT already implements this optimization, but Pallene does not yet. By optimizing it manually, we could obtain a more accurate comparison of the costs of array reads and writes.

We cannot rely on the C compiler for all optimizations, such as the loop-invariant code motion we just mentioned. Pallene array operations may call Lua runtime functions (e.g. to grow the array) and these function calls inhibit several optimizations at the C level.

Unlike the C compiler, the Pallene compiler is aware of Pallene's semantics in general and array operations in particular; therefore it is better suited to perform these optimizations.

We also implemented all these benchmarks in C, but using Lua data structures manipulated through the Lua–C API. This allowed us to compare how the standard scripting approach with C compares to Pallene, which bypasses the API.

We ran our experiments in a 3.10 GHz Intel Core i5-4440 with 8 GB of RAM and normalized the execution times to the ones of the reference Lua interpreter (PUC-Lua). We used Lua version 5.4-work1 for the benchmarks.[2]

We compared Pallene programs with their Lua equivalents, ran under both the reference Lua implementation and under LuaJIT. In all benchmarks the only difference between the Lua and Pallene programs is the presence of type annotations, since we restricted the Lua programs to the language subset supported by Pallene. The results are shown as a table in Figure 6 and normalized by the Lua running time in Figure 7.

The first benchmarks—prime sieve and matrix multiplication—heavily feature array operations and arithmetic. In both, Pallene and LuaJIT achieved similar performance, with an order of magnitude improvement when compared to the reference Lua implementation. The Lua–C API implementation had much smaller gains. This shows the costs of using the Lua–C API at this level of granularity (single access to array elements). The small gain is probably due to arithmetic being performed in C instead of in dynamically-typed Lua.

The third benchmark—the N-queens problem—also features array operations and arithmetic, but has a larger proportion of arithmetic compared to array operations. Pallene and LuaJIT performed as well as they did in the previous two benchmarks. The C-API did better, due to the heavier weight of arithmetic operations.

In the fourth benchmark—binary search—Pallene ran more than three times faster than LuaJIT and the Lua–C API implementation. We can see that the performance of Pallene was similar to the other benchmarks which indicates that the difference is due to LuaJIT doing worse on this particular benchmark. The binary search does many unpredictable branches, which is very bad for trace-based JIT compilers. This illustrates the unpredictability of JIT compilation in general and trace-based JIT compilation in particular.

The final benchmark—Conway's game for life—spends much time doing string operations and generates a lot of garbage. The C-API implementation could not obtain a speedup compared to the reference interpreter and LuaJIT could barely beat it. We suspect that this is due to recent improvements in the PUC-Lua garbage collector. (Lua 5.3 takes 150% longer to run than Lua 5.4 .)[3] Given that this benchmark spends so much time in the garbage collector, Pallene's 2x speedup seems respectable.

Benchmark	Lua	Lua–C API	LuaJIT	Pallene
Sieve	6.641	4.522	1.079	1.054
Matmul	2.502	2.200	0.212	0.263
N Queens	14.276	3.850	1.610	1.500
Binary Search	8.816	2.694	2.991	0.843
Game of Life	2.133	1.990	2.459	1.151

Figure 6: Exact running times for the benchmarks, in seconds.

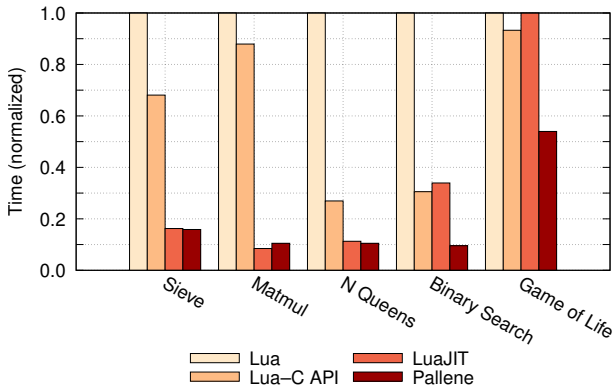

Figure 7: Comparison of Pallene's performance against Lua, LuaJIT and the Lua–C API. Times are normalized by the Lua result.

6 RELATED WORK

JIT compilers answer a popular demand to speed up dynamic programs without the need to rewrite them in another language. However, they do not evenly optimize all idioms of the language, which in practice affects programming style, encouraging programmers to restrict themselves to the optimized subset of the language. Pallene has chosen to be more transparent about what can be optimized, and made these restrictions a part of the language.

Like Gradual Typing systems, Pallene recognizes that adding static types to a dynamic language provides many benefits, such as safety, documentation, and performance. Pallene is also inspired by the Gradual Guarantee, which states that the typed subset of the language should behave exactly as the dynamic language, for improved interoperability. Unlike many gradually typed systems, Pallene can only statically type a restricted subset of Lua. This avoids the complexity and performance challenges that are common in many gradually typed systems.

Common Lisp is another language that has used optional type annotations to provide better performance. As said by Paul Graham in his ANSI Common Lisp Book [8], "Lisp is really two languages: a language for writing fast programs and a language for writing programs fast". Pallene and Common Lisp differ in how their sublanguages are connected. In Common Lisp, they live together under the Lisp umbrella, while in Pallene they are segregated, under the assumption that modules can be written in different languages.

Cython [4] is an extension of Python with C datatypes. It is well suited for interfacing with C libraries and for numerical computation, but its type system cannot describe Python types. Cython is

[2] The source code for the Pallene compiler and the benchmarks can be found at https://github.com/pallene-lang/pallene/releases/tag/sblp2018

[3] LuaJIT had plans to update its garbage collector but that still hasn't happened.

unable to provide large speedups for programs that spend most of their time operating on Python data structures.

Terra is a low-level system language that is embedded in and meta-programmed by Lua. Similarly to Pallene, Terra is also focused on performance and has a syntax that is very similar to Lua, to make the combination of languages more pleasant to use. However, while Pallene is intended for applications that use Lua as a scripting language, Terra is a stage-programming tool. The Terra system uses Lua to generate Terra programs aimed at high-performance numerical computation. Once produced, these programs run independently of Lua. Terra uses manual memory management and features low-level C-like datatypes. There are no language features to aid in interacting with a scripting language at run-time.

7 CONCLUSION AND FUTURE WORK

Our initial results suggest that Pallene's approach of providing a statically typed companion language for a dynamically typed scripting language is a promising approach for applications and libraries that seek good performance. It apparently will be able to compete with LuaJIT as an alternative to speeding up Lua applications. As we expected, Pallene performs better than the scripting approach. Moreover, porting a piece of code from Lua to Pallene seems much easier than porting it to C.

We also want to study the impact of implementing traditional compiler optimizations such as common sub-expression elimination and loop invariant code motion. Our current implementation relies on an underlying C compiler for almost all optimizations, and our work suggests that implementing some optimizations at the Pallene level might lead to significant improvements. The C compiled cannot be expected to understand the Lua-level abstractions needed to perform these optimizations.

One question we wish to answer in the future is whether the type system simplicity and the good performance results we achieved in the array-based benchmarks will be preserved as we add more language features, such as records, objects, modules and a foreign function interface.

ACKNOWLEDGMENTS

We would like to thank Hisham Muhammad, Gabriel Ligneul, Fábio Mascarenhas, and André Maidl for useful discussions about the initial ideas behind Pallene. We would also like to thank Sérgio Medeiros for assisting us with the developement of Pallene's scanner and parser.

Pallene was born as a fork of the Titan [17] programming language, with a focus on researching the performance aspects of dynamic programming language implementation. Gabriel Ligneul heavily contributed to current implementation of the Pallene compiler.

REFERENCES

[1] Keith Adams, Jason Evans, Bertrand Maher, Guilherme Ottoni, Andrew Paroski, Brett Simmers, Edwin Smith, and Owen Yamauchi. 2014. The Hiphop Virtual Machine. In *Proceedings of the 2014 ACM International Conference on Object Oriented Programming Systems Languages & Applications (OOPSLA '14)*. 777–790. DOI: http://dx.doi.org/10.1145/2660193.2660199

[2] Petka Antonov and others. 2013. V8 Optimization Killers. (2013). https://github.com/petkaantonov/bluebird/wiki/Optimization-killers Retrieved in 2017-01-08. Full author list available at https://github.com/petkaantonov/bluebird/wiki/Optimization-killers/_history.

[3] Edd Barrett, Carl Friedrich Bolz-Tereick, Rebecca Killick, Sarah Mount, and Laurence Tratt. 2017. Virtual Machine Warmup Blows Hot and Cold. In *Proceedings of the 32nd Annual Conference on Object-oriented Programming Systems, Languages, and Applications (OOPSLA '17)*. DOI: http://dx.doi.org/10.1145/3133876

[4] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn, and K. Smith. 2011. Cython: The Best of Both Worlds. *Computing in Science Engineering* 13, 2 (March 2011), 31–39. DOI: http://dx.doi.org/10.1109/MCSE.2010.118

[5] Carl Friedrich Bolz, Antonio Cuni, Maciej Fijalkowski, and Armin Rigo. 2009. Tracing the Meta-level: PyPy's Tracing JIT Compiler. In *Proceedings of the 4th Workshop on the Implementation, Compilation, Optimization of Object-Oriented Languages and Programming Systems (ICOOOLPS '09)*. ACM, 18–25. DOI: http://dx.doi.org/10.1145/1565824.1565827

[6] L. Peter Deutsch and Allan M. Schiffman. 1984. Efficient Implementation of the Smalltalk-80 System. In *Proceedings of the 11th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '84)*. ACM, New York, NY, USA, 297–302. DOI: http://dx.doi.org/10.1145/800017.800542

[7] Andreas Gal, Christian W. Probst, and Michael Franz. 2006. HotpathVM: An Effective JIT Compiler for Resource-constrained Devices. In *Proceedings of the 2nd International Conference on Virtual Execution Environments (VEE '06)*. New York, NY, USA, 144–153. DOI: http://dx.doi.org/10.1145/1134760.1134780

[8] Paul Graham. 1996. *ANSI Common LISP*. Apt, Alan R. http://www.paulgraham.com/acl.html

[9] Hugo Musso Gualandi. 2015. *Typing Dynamic Languages – a Review*. M.S. thesis, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

[10] Javier Guerra. 2017. LuaJIT Hacking: Getting next() out of the NYI list. CloudFare Blog. (Feb. 2017). https://blog.cloudflare.com/luajit-hacking-getting-next-out-of-the-nyi-list/.

[11] Javier Guerra Giraldez. 2016. LOOM - A LuaJIT performance visualizer. (2016). https://github.com/cloudflare/loom

[12] Javier Guerra Giraldez. 2017. The Rocky Road to MCode. Talk at Lua Moscow conference, 2017. (2017). https://www.youtube.com/watch?v=sz2CuDpltmM

[13] Roberto Ierusalimschy, Luiz Henrique de Figueiredo, and Waldemar Celes. 2007. The Evolution of Lua. In *Proceedings of the Third ACM SIGPLAN Conference on History of Programming Languages (HOPL III)*. ACM, New York, NY, USA, 2–1–2–26. DOI: http://dx.doi.org/10.1145/1238844.1238846

[14] Roberto Ierusalimschy, Luiz Henrique De Figueiredo, and Waldemar Celes. 2011. Passing a Language Through the Eye of a Needle. *Commun. ACM* 54, 7 (July 2011), 38–43. DOI: http://dx.doi.org/10.1145/1965724.1965739

[15] Baker J., Cunei A., Kalibera T., Pizlo F., and Vitek J. 2009. Accurate garbage collection in uncooperative environments revisited. *Concurrency and Computation: Practice and Experience* 21, 12 (2009), 1572–1606. DOI: http://dx.doi.org/10.1002/cpe.1391

[16] André Murbach Maidl, Fabio Mascarenhas, and Roberto Ierusalimschy. 2015. A Formalization of Typed Lua. In *Proceedings of the 11th Symposium on Dynamic Languages (DLS 2015)*. ACM, New York, NY, USA, 13–25. DOI: http://dx.doi.org/10.1145/2816707.2816709

[17] André Murbach Maidl, Fábio Mascarenhas, Gabriel Ligneul, Hisham Muhammad, and Hugo Musso Gualandi. 2018. Source code repository for the Titan programming language. (2018). https://github.com/titan-lang/titan

[18] Bret Mogilefsky. 1999. Lua in Grim Fandango. Grim Fandango Network. (May 1999). https://www.grimfandango.net/features/articles/lua-in-grim-fandango.

[19] John K. Ousterhout. 1998. Scripting: Higher-Level Programming for the 21st Century. *Computer* 31, 3 (March 1998), 23–30. DOI: http://dx.doi.org/10.1109/2.660187

[20] Mike Pall. 2005. LuaJIT, a Just-In-Time Compiler for Lua. (2005). http://luajit.org/luajit.html http://luajit.org/luajit.html.

[21] Mike Pall. 2012. LUAJIT performance tips. lua-l mailing list. (nov 2012). http://wiki.luajit.org/Numerical-Computing-Performance-Guide http://wiki.luajit.org/Numerical-Computing-Performance-Guide.

[22] Mike Pall and others. 2014. Not Yet Implemented operations in LuaJIT. LuaJIT documentation Wiki. (2014). http://wiki.luajit.org/NYI Retrieved 2017-01-08. Full author list available at http://wiki.luajit.org/history/NYI.

[23] Armin Rigo, Michael Hudson, and Samuele Pedroni. 2005. *Compiling Dynamic Language Implementations*. Tech. rep., Heinrich-Heine-Universität Düsseldorf. https://bitbucket.org/pypy/extradoc/raw/tip/eu-report/D05.1_Publish_on_translating_a_very-high-level_description.pdf

[24] Jeremy G. Siek, Michael M. Vitousek, Matteo Cimini, and John Tang Boyland. 2015. Refined Criteria for Gradual Typing. In *1st Summit on Advances in Programming Languages (SNAPL '2015)*. Asilomar, California, USA, 274–293. DOI: http://dx.doi.org/10.4230/LIPIcs.SNAPL.2015.274

[25] Asumu Takikawa, Daniel Feltey, Ben Greenman, Max S. New, Jan Vitek, and Matthias Felleisen. 2016. Is Sound Gradual Typing Dead?. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '16)*. 456–468. DOI: http://dx.doi.org/10.1145/2837614.2837630

[26] The Chromium Project. 2008. The Chrome V8 Engine. (2008). https://developers.google.com/v8/ Retrieved 2017-01-08.

[27] Sam Tobin-Hochstadt and Matthias Felleisen. 2006. Interlanguage Migration:

From Scripts to Programs. In *Companion to the 21st ACM SIGPLAN Symposium on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA '06)*. ACM, New York, NY, USA, 964–974. DOI: http://dx.doi.org/10.1145/1176617.1176755

[28] Michael M. Vitousek, Andrew M. Kent, Jeremy G. Siek, and Jim Baker. 2014. Design and Evaluation of Gradual Typing for Python. In *Proceedings of the 10th ACM Symposium on Dynamic Languages (DLS '14)*. ACM, New York, NY, USA, 45–56. DOI: http://dx.doi.org/10.1145/2661088.2661101

[29] Thomas Würthinger, Christian Wimmer, Andreas Wöß, Lukas Stadler, Gilles Duboscq, Christian Humer, Gregor Richards, Doug Simon, and Mario Wolczko. 2013. One VM to Rule Them All. In *Proceedings of the 2013 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software (Onward! 2013)*. 187–204. DOI: http://dx.doi.org/10.1145/2509578.2509581