Efficient Layer Activation in ContextJS

Robert Krahn*, Jens Lincke*, and Robert Hirschfeld*
Software Architecture Group
Hasso Plattner Institute
University of Potsdam, Germany
http://www.hpi.uni-potsdam.de/swa
*Email: {firstname.lastname}@hpi.uni-potsdam.de

Abstract—Context-oriented programming (COP) describes language extensions for modularizing behavioral or structural variations that are to be composed at run-time. Different COP infrastructures and implementations offer several strategies for scoping, activation, and deactivation of such compositional units. Often, the mechanisms employed cause substantial execution overhead. In this paper we present an optimization technique for ContextJS—our COP extension to JavaScript—that can significantly reduce this overhead to run context-aware code efficiently.

I. INTRODUCTION

ContextJS is a context-oriented programming (COP) extension with JavaScript as its host language. It is implemented as a library using the meta-programming facilities of JavaScript and does not require VM modifications. Similar to other COP extensions to dynamically typed host languages, ContextJS intercepts the runtime lookup mechanism for invoking behavior variations. So far, this approach comes with a significant impact on run-time performance. Micro-benchmarks [1] confirmed a performance loss of more than 99% compared to un-instrumented code.

Although this overhead does not affect the entire system but only methods made aware of behavioral variations, we experienced that extensive use of ContextJS can make an interactive system such as Lively Kernel [2], [3] noticeably slower.

In the Lively Kernel runtime and development environment we employ ContextJS for various applications. We can group those applications broadly in the following three categories:

1) Global modularization and deployment. ContextJS layers are used to both modularize and deploy/un-deploy structural and behavioral variations.

2) Object-specific behavior. ContextJS is used to selectively add experimental code on a per-object basis during development [4]. In self-sustaining systems like Lively Kernel, this allows to keep changes separate so that newly added erroneous code can be explored and removed easily. In such systems, this is especially important if programming tools like an editor might depend on the code that is being changed.

3) Tracing of system behavior. ContextJS provides infrastructure to selectively instrument and trace method activation and deactivation within the dynamic extent of a particular execution context [5]. It is often used to separate tracing target and tracer logic so that self-referentiality is avoided.

COP variations defined globally or on a per-object basis are rather static since they do not depend on propagation within dynamic extents at run-time but rather resemble partial class and method definitions.

In dynamic extend-based scoping the system's layer composition changes often, but for a single method the layer composition may be the same for consecutive calls, because the layers are (de-)activated in the same pattern. With that, layer compositions for specific objects will not change most of the time.

In this paper we present an optimization technique for layer composition and lookup in ContextJS that can significantly reduce method activation overhead. By using the meta-programming facilities of JavaScript and Lively Kernel, we reduce a composition of multiple related partial method definitions to a single method. For all compositions that will not change anymore, this compilation step removes all run-time composition machinery and with that all its run-time overhead.

This improvement does not restrict the expressibility or change the semantics of ContextJS. Especially dynamic-extent and object-based activations are optimized. To be able to react to layer composition changes, a validation check is added to the generated method. When the layer composition check indicates that the layer composition, which the method was compiled for, does not correspond to the current one, it will invalidate the current optimization leading to the generation of a new one.

The remainder of the paper is organized as follows. Section 2 gives a short overview of Context-oriented programming and the mechanisms used to define and activate behavior variations. Section 3 introduces our optimization mechanism. In Section 4 we present our implementation and discuss design decisions and implementation options. Section 5 evaluates our approach. Section 6 discusses the related work and section 7 summarizes the paper and mentions topics for future work.
II. LAYER COMPOSITION AND METHOD ACTIVATION IN COP

In this section, we give a brief introduction to COP [6] and the method lookup and activation mechanism found in ContextJS [4].

A. Activation of Behavior Variations

In the COP execution model, the program behavior depends upon the context in which it is run. For integrating behavior variations into the program COP introduces the layer construct. Layers are a modularization concept that encapsulates behaviors distributed over several objects or classes. The behavior variations are defined inside layers by means of partial method definitions. We use the term layered method for the composition of at least one partial method and a base method. Class or object slots are called layer-aware when there is at least one layer definition with a partial method for that slot, even if the layer is not activated.

The layered method compositions are based on the compositions of layers. Layer compositions are created at runtime by (de-)activating layers, e.g. in the dynamic extent of a control flow or globally. When layers are activated at runtime and a message send reaches a layer-aware slot, the object-oriented method lookup is extended by a sideways lookup. This lookup is called layer composition and gathers corresponding partial methods of active layers.

B. Layer Composition in ContextJS

ContextJS is a typical dynamic language COP implementation [1], because it is implemented only using the meta-programming facilities of its host language JavaScript.

Listing 1 shows a class definition of a Person class\(^1\). Its toString method returns a stringified representation and can be parameterized using a boolean parameter. Two layers are defined afterwards that extend the behavior of Person\(\Rightarrow\)toString. The EmploymentLayer and AddressLayer define partial methods to append to the result of preceding partial method activations.

Listing 2 shows how the class is instantiated and how the Person instances returns different results depending on the activated layers.

In ContextJS the slots of a class or object are made layer-aware at layer definition time. This means that for all layered methods and properties a wrapper method is installed. In the example the toString slot of the Person is made layer-aware when the first layer is defined. The original implementation is removed from the slot and replaced by a generic wrapper method. Upon method activation this wrapper then computes the current layer composition and performs the lookup and activation of the partial methods belonging to the currently activated layers.

Figure 1 illustrates the layer composition process for the toString message call on line 10 in listing 2. While the address layer is globally and employment layer dynamically activated, Person\(\Rightarrow\)toString is called (step 1). This activates the COP method wrapper. Upon activation the wrapper accesses the active layers and creates the corresponding layer composition, i.e. the wrapper gathers the partial methods defined in EmploymentLayer and AddressLayer as well as the base method (step 2). The partial methods are ordered accordingly to the activation order of their layers on a stack.

During the subsequent steps (3.1 - 5.1) the partial methods are activated. When a proceed call is encountered the control is handed back to the ContextJS meta level to fetch the next partial method from the stack that is activated in turn. This process continues until either a partial method does not proceed or the base method is reached. When that happens (step 6) the control is handed back to calling methods until the method wrapper is reached. The wrapper returns the result of the method activation to the call site.

\(^1\) Using the Lively Kernel class system [2]
The lookup mechanism depicted here is used on every activation of a layered method. In the next section we will discuss an approach for optimizing this process.

III. PARTIAL METHOD INLINING AND COMPOSITION CACHING

Since recorded runtime information for our COP applications showed that layer composition changes do not happen often for most objects, the default layer composition mechanism of ContextJS computes most of the time redundant compositions. In this section we will present an adapted layer composition and lookup mechanism that will only re-compute layer compositions when necessary. Furthermore, we denote what parts of the method composition and lookup mechanism can be optimized.

1. function toString(beShort) {
   if (!currentLayerCompositionIsValid())
      return inlineMethodCreator.apply(arguments);
   return function(beShort) {
      return function(beShort) {
         return beShort ?
            this.fullName : "Person( + this.fullName + ')';
         return (this, beShort) + ' (' + this.address() + ')';
      ).call(this, beShort) + ' (HPI)';
   };
   return this.fullName : "Person( + this.fullName + ')';
   return (this, beShort) + ' (' + this.address() + ')';
}).call(this, beShort) + ' (HPI)';
}

Listing 3. The result of automatically inlining the toString method with our approach.

A. Inlining

For a given layer composition, the partial methods and the necessary COP infrastructure code are combined into a
single method. Each generated method corresponds to a layer composition. After its creation the generated method is used for all subsequent activations as long as the active layer composition is similar to the method’s corresponding layer composition.

Listing 3 shows the inlining result of Person >> toString for the layer composition: EmploymentLayer, AddressLayer, base. According to the order of activated layers we inline each proceed call with the partial method that the proceed call would activate when executed. We directly embed the partial method and not just its statements so that the inlined source code has its own scope. This also makes the source code transformation simpler since the actual call to the partial method will not be changed and parameters of the inlined partial method do not need to be renamed. The normal JavaScript function call semantic applies.

B. Composition Validation

A validation check is added as the first statement of the generated method. The validation check is the only COP metadata level code that is invoked. The validation check compares the active layer composition to the corresponding layer composition of the generated method and invalidates it if necessary. Invalidation means to remove the generated method from the slot it is installed in and either generate and install a new inlined method or install a common ContextJS method wrapper.

Since the validation check is invoked for every method activation its efficiency is critical. We propose the usage of fingerprints [7] for efficient comparison of layer compositions. The fingerprint of a layer composition regarding an object slot can be computed from unique layer identifiers and version information about the partial methods and base method. Given two layer compositions and an object slot their fingerprints should be equal if both layer compositions include the same layers in the same order and, regarding the slot, the partial methods and the base method for that slot are equal. We describe in the next section a possible implementation.

C. Composition Caching

The currently active layer composition must be accessed for both the validation check and for the partial method lookup while generating an inlined method. Since this must happen at least once when activating a layered method, also this operation should be efficient.

Layers can be (de-)activated globally and within the dynamic extent of a control flow. ContextJS also allows object scoping [4], i.e. layers are activated and deactivated depending on the receiver of a message send. Since object layer activations can be completely customized, e.g. be made random, caching them is not possible without further constraints. However, when the receiver object does not implement its own layer computation, only globally and dynamically activated layers have to be considered when computing the active layer composition. When those layers are stored in order of their activation in a collection, e.g. a stack, it is possible to associate a composition cache together with that data structure. The cache can be invalidated whenever the collection changes.

IV. IMPLEMENTATION

This section presents our implementation. We will show in detail how the optimization mechanism described previously was applied to ContextJS and discuss problems we encountered.

A. Method Generation

The process of inlining partial methods of a layer composition to create generated methods as described in section 3.3 is the following:

When a partial method is defined and the slot in the refined object or class was not yet made layer-aware, install a method wrapper in that slot. The method wrapper will not handle the partial method dispatch directly but is responsible for creating an inlined function. We designate that method wrapper as inlinedMethodCreator.

The first activation triggers the inlinedMethodCreator that will then lazily build the inlined method. This involves the following steps:

1) Access the current layer composition. If the object for which the inlined method is built does customize the active layer computation in a special way, the layer composition has to be computed. If not, a cached layer composition can be used. The global collection cop.LayerStack holds all dynamically and globally activated layers. When the currently active layers have not changed since a previous composition computation then the composition is cached and can be reused. Otherwise the cop.LayerStack has to be iterated and a new layer composition has to be computed from all with and without layer specifications.

2) Compute the fingerprint for the composition.
3) Gather partial methods from methods.
4) Inline partial methods by starting with the first partial method (the partial method that is activated when the layered method is executed) and inline partial methods of the next nesting level. Repeat that until all partial methods and the base method are transformed into source code block. Add the validation check and create a new function out of the transformed source code.
5) Call new inlined method.

For each subsequent method call do:

1) Access fingerprint of current layer composition. Either using cop.LayerStack that is also used to store fingerprints together with cached compositions or compute layer composition and compute hash.
2) If the corresponding layer composition is equal to the current layer composition continue running the inlined method. Otherwise remove the inlined method from its
In JavaScript, functions bind values to free variables at definition time and are therefore closures. Slot and invoke `inlinedMethodCreator` again. This starts the composition process again.

Following method calls with the same layer composition are processed as shown in figure 3. The inlined method gets directly activated and performs the validation check. Since the activated layers have not changed, the method is still valid and the control flow continues normally.

B. Dealing with Closures

Rewriting source code in JavaScript is an interesting challenge. Rewriting means that that existing functions are converted into a string representation that is then parsed and changed. Recompiling can be done by evaluating the function source code so that a function object is the result.

In JavaScript, functions bind values to free variables at definition time and are therefore closures. Once bound, JavaScript provides no means for introspection of the sub-method level\(^4\). When transforming source code this becomes a problem since bound variables cannot be transferred to rewritten code.

\(^4\) Instrumenting a JavaScript VM would solve this problem but would make our optimization mechanism less general

To provide a solution that allows code rewriting and closures at the same time, we support first class closure objects. These objects can be created alongside functions as is shown in listing 4.

This mechanism allows to rewrite functions even with bound variables. However, this burdens users to explicitly define closures. The Lively Kernel is implemented using objects and classes; closures on a per method level are rarely used so that this was not a drawback up to this point.

C. Source Code Transformations

For the actual source code modifications we used OMeta-based \([8]\) JavaScript parser and AST transformer. However, this approach introduced a lot of overhead for runtime layer inlining.

Since the code transformations are simple — finding and replacing the proceed statement, injecting a statement at the method start — our current implementation uses regular expressions for parsing and rewriting. A special case that can appear while transforming and that is currently not covered by our implementation is that inlined partial methods with unbound variables can lexically capture variables defined in an outer function block. Listing 5 shows an example. The method `m` of `spyObj` and `SecretLayer` are inlined. In the generated method the inlined base method is able to access the variable `secret`.

Solutions to that problem include applying mechanisms that can be found in hygienic macros \([9]\) or code generators \([10]\). Alternatively the inline mechanism could be adapted so that it would not directly inline the source code but would replace the `cop.proceed` with hard wired method calls. Thus, partial methods could no longer accidentally access the lexical scope of the generated method.

D. Composition Validation

The efficiency of the validation check is critical for the overall performance characteristics of method inlining since this check is invoked on every call.

Our current approach is the following: Each layer has a unique identifier that consists of its name and a timestamp of its last change, i.e. the point of time one of its partial method were changed, added, or removed. These identifiers are then concatenated for computing the fingerprint of the whole composition. On method inline creation the fingerprint layer composition is stored and later used for comparison when subsequent calls activate the inlined method again.

We use string comparison for the validation check. To improve its efficiency fingerprints should use better hashing or fingerprint mechanism like Rabin fingerprints \([7]\), \([11]\).

V. Evaluation

To evaluate the performance of the new layer inlining approach we run benchmarks used to compare several COP implementations \([1]\). The benchmarks measure the relative performance of a layered method lookup compared to a non layered method lookup of the host language. The results of
the micro-benchmarks\textsuperscript{5} in Figure 4 show that the inlining approach is up to 10 times faster when compared with dynamic dispatch. However, relative to the not instrumented code the overhead is still very significant since only 1–3\% of its performance is reached. Profiling shows that most of the time is spend accessing layer compositions even though compositions are cached. Further optimizations are needed here to get better results.

The relative performance characteristics of running an layer-aware method without any layer activations, compared to 1–

\textsuperscript{5}All benchmarks were executed with Google Chrome 12.0.742.91 on a MacBook Pro, Intel Core i7 2 GHz 4 Cores, 4 GB RAM.
mark in Figure 5, we measured how the performance can degenerate if the layer composition changes on every method call. In this extreme case the standard dynamic lookup outperforms the inlining as shown in Listing 6. Because of source code transformations and function generation computation, costs of method inlining are very high. When the layer composition changes every time the dynamic lookup is around 200 times faster. One approach to solve this problem is the introduction of an inlined method cache. The cache stores generated inlined methods that were installed before but invalidated later. When the object is activated with the same layer composition the old methods are fetched from the cache and re-installed so that source code transformations and function creation becomes unnecessary. We implemented that optimization prototypically but have not yet adopted it since layer composition changes do not happen often and had not yet a noticeable performance impact.

VI. RELATED WORK
A common implementation approach to extending object-oriented programming languages with COP concepts is the extension of method dispatch by yet another dimension [6]. Depending on the particular implementation strategy, the performance overhead can be significant. In a benchmark using the popular “figure editor” from the AOP community as an example [12], ContextL [13] has shown that COP extensions can perform competitively [14]. But due to other implementation constraints not all COP extensions can take advantage of this particular optimization.

There are several COP implementations for Java that are each implemented with different approaches. ContextJ [15] uses compiler and preprocessor techniques. This allows for example to directly compile the lookup code into the base method. Other implementations like ContextLogicAj [16], and ContextJ* [6], an Java5 library based implementation of COP concepts, are more restricted in their implementation possibilities.

The \textit{cj} prototype [17], [18] is an atypical COP implementation. It implements a minimal subset of ContextJ [14], [19] to demonstrate a machine-model of multi-dimensional separation of concerns [20]. In \textit{cj} a layer activation directly changes the method dictionaries of all affected classes. This is a very expensive operation, but it eliminates layer composition at message dispatch time. This approach is slow when layers get often (de-)activated and when a layer refines methods in many places, but it is fast if this does not happen often. In our approach a layer composition change also invalidates all inlined partial method compositions, but this invalidation is only detected and handled lazily, making the actual layer activation fast, but the next layered method executions slow.

Aspect-oriented programming (AOP) languages like AspectJ [21] also change the method dispatch in object-oriented programming. They are implemented on a source code, byte code or the virtual machine level [20]. AspectScript [22] is an AOP implementation for JavaScript. Since it allows for very expressive aspects, the implementation has to instrument every method in the system. This is done by parsing and manipulating the JavaScript source code of the base system. This slows down the whole system by 5-6 times even when no aspects are active. Contrastingly, in COP it can be statically determined which methods are refined by layers. Therefore ContextJS only needs to change those methods. Methods that are not refined run unaffected and with their normal performance.

Inlining of methods is a well-known compiler optimization for statically and dynamically compiled Object-oriented languages. For dynamic inlining in Self [23] it is argued that with the help of type feedback the inlining of messages leads to significant performance improvements because in pure OO languages methods are typically very small and are called very frequently [24]. In Self the inlining of methods is also done at run-time but on the level of the virtual machine (VM). Since our COP message dispatch is handled on the host language level the approach cannot directly used in our implementation. But since inlining of layer compositions in ContextJS makes the execution of layered methods less dynamic, we expect that optimizations in JavaScript VMs can be more often applied as could be with a dynamic dispatch mechanism.
VII. SUMMARY AND OUTLOOK

COP language extensions enable behavior variations to be modularized, dynamically activated, and scoped. Our language extension ContextJS implements the COP paradigm in JavaScript and is used in the Lively Kernel environment for various applications. Regular ContextJS dispatch and composition mechanisms introduced a significant overhead for layer-aware code at run-time.

To make ContextJS more efficient, we present an optimization technique for ContextJS to improve run-time layer lookup and composition performance. Based on our observation that layer composition changes per object happen rarely, the optimizations we implemented are targeted at making execution traces fast that rarely include layer composition changes.

We propose the generation of a single method that has all contributing partial and base methods inlined. This removes all meta-level composition code apart from the validation checks in layered methods. Since a validation check still has to be included, we also presented techniques to optimize layer composition access and checks for composition change.

Our evaluation shows noticeable performance improvements. However, there is still a big performance gap to code not instrumented that comes from the validation check that is necessary to re-compose layered methods on layer composition changes.

A macro benchmark suite for COP implementations would be a macro help to guide further optimization efforts.

REFERENCES


