AN EXPANDABLE EXTRACTION FRAMEWORK FOR ARCHITECTURAL PERFORMANCE MODELS

Jürgen Walter  
University of Würzburg  
97074 Würzburg Germany

Christian Stier  
FZI Research Center for Information Technology  
76131 Karlsruhe Germany

Heiko Koziolek  
ABB Corporate Research  
68526 Ladenburg Germany

Samuel Kounev  
University of Würzburg  
97074 Würzburg Germany

ABSTRACT

Providing users with Quality of Service (QoS) guarantees and the prevention of performance problems are challenging tasks for software systems. Architectural performance models can be applied to explore performance properties of a software system at design time and run time. At design time, architectural performance models support reasoning on effects of design decisions. At run time, they enable automatic reconfigurations by reasoning on the effects of changing user behavior. In this paper, we present a framework for the extraction of architectural performance models based on monitoring log files generalizing over the targeted architectural modeling language. Using the presented framework, the creation of a performance model extraction tool for a specific modeling formalism requires only the implementation of a key set of object creation routines specific to the formalism. Our framework integrates them with extraction techniques that apply to many architectural performance models, e.g., resource demand estimation techniques. This lowers the effort to implement performance model extraction tools tremendously through a high level of reuse. We evaluate our framework presenting builders for the Descartes Modeling Language (DML) and the Palladio Component Model (PCM). For the extracted models we compare simulation results with measurements receiving accurate results.

1. INTRODUCTION

During the life-cycle of a software system, performance analysts continuously need to provide answers to and act on performance-relevant questions about response times, resource utilization, bottlenecks, trends, anomalies, etc. It is a common approach to evaluate systems using model-based predictions in addition to measurement-based approaches. Model-based predictions allow for exploration of alternative deployments, architectures, and configurations without the need to test them in a live system. Architectural performance models, a subcategory of quality-aware architecture description languages, can be applied to explore performance properties of a software system for design time and runtime scenarios. At design time models can be applied to reason on effects of design decisions when implementation not fully available yet. At runtime they can be used to reason on effects of changing user behavior on performance to avoid contention via reconfiguration. The main advantage of architectural models, compared to purely predictive models, is that they preserve context information alongside the performance information. Thereby, architectural performance models provide actionable knowledge for automated or manual architectural design decisions. There exist many architectural performance modeling languages, e.g., Descartes Modeling Language (DML) [10], Palladio Component Model (PCM) [4], CACTOS [1], UML MARTE [12], and ACME [17]. Even though the mentioned architectural models focus on different application domain, they have a high semantic overlap. Key concepts such as compositionality, component interfaces, and interface providing roles can be found in all of the mentioned languages. Each of the languages focuses on a different application scenario and is supported by different tool chains for model analysis.

The manual creation of accurate performance models for large scale systems requires extensive effort and knowledge of the architectural modeling language. Existing performance model extraction tooling focuses on the extraction of models for a single language. This requires the reimplementation and maintenance of extraction tooling for each architecture language.

In this work we provide an expandable approach for automated extraction of architectural performance models. Our approach isolates the extraction of shared concepts from language specific implementations. Developers adopting our framework only need to implement a builder interface covering the language specific mapping of common concepts, instead of implementing the entire extraction code. The implementation of our approach, called Performance Model Extractor (PMX), can be reused at different development stages to create performance models of different modeling languages. The remainder of this paper is organized as follows: Section 2 motivates our approach by means of a pro-
2. PROBLEM STATEMENT

Architectural performance models can be applied for various purposes and at different development stages [7]. Architectural performance models have been applied for runtime optimization, as well as evolution scenarios. Further, architectural performance models have been applied to evaluate different concerns. Analysis approaches include, e.g., SLA evaluation, or extensions to support energy cost prediction. To enable the application of previously described in an integrated approach, there is very limited tool support. Most analysis techniques are only supported by a single analysis tool chain. An integrated approach would enable the following aspects:

- Analysis Tool chain Parallelism “As a user, I would like to use different tool chains in parallel.” Advanced Analysis approaches only supported for a limited set of languages. Analysis tool chain parallelism would enable an extended range of analysis techniques.
- Analysis Toolchain Flexibility “As a user, I prefer not to be forced to decide about the toochain in advance.” Easy change of modeling formalism.
- Extraction Toolchain Reuse “As a user, I would like to include performance model extraction for newly emerging formalisms without bothering about extraction complexity.” Currently there is a reinvention of extraction methodology for each modeling formalism. The integration and composition of libraries for the extraction causes huge efforts.

We see a need for an integrated approach that may reuse the same monitoring data for different concerns. However, performance model extraction tooling so far allows only for the extraction of a single language.

3. APPROACH

The goal of our approach is to separate the extraction of shared concepts and concerns found in many architectural performance models, from language specific construction and mapping routines. A key concern shared when creating architectural performance models is to determine resource demands. Resource demands represent the computational demand caused when users issue calls to the services of the modeled system. Resource demand estimation techniques leverage measurement logs to estimate the resource demands of individual service calls [14]. To provide resource demand estimation techniques with sufficient information, the compositional structure and call dependencies between system components need to be extracted. Kieler is an example of monitoring tooling that supports this [20].

In this work we present a framework, called PMX, that provides developers with a solution that integrates established tooling for monitoring [20], log processing [19], and resource demand estimation [14]. To leverage PMX for model construction, developers only have to implement a model builder interface that maps language independent concepts to language specific representations.

PMX employs the builder [8] pattern to decouple language specific mappings from common model extraction concerns. Figure 1 shows the coarse grained architecture including two builder implementations. The intent of the builder design pattern is to separate the construction of a complex composed entity from its representation. Through this, the same construction process can create different representations. The construction complexity of architectural performance models is caused by control flow extraction, the high degree of interconnection between concepts [18] and by the choice of resource demand extraction techniques. By decoupling resource demand estimation and modeling extraction methods for the concepts shared among architectural performance models, PMX reduces the effort for implementing automated performance model extraction.

4. BUILDER INTERFACE

Our approach builds upon the basic assumption that architectural performance modeling languages share equivalence classes of elements. For identification we orient at model-driven software development (MDSD) community. Table 1 shows the core concepts of architectural performance models we identified. They include application architecture concepts like e.g. role, interface, signature, components, service behavior and resource demands. Those concepts occur or correspond with entities in many quality-aware architectural description languages like, e.g., DML, PCM, CACTOS [1], UML MARTE [12], and ACME [17]. The terminology of PMX builds upon the terminology used in DML, PCM and CACTOS. The following outlines the correspondence between the chosen terminology and ADL, if it deviates. For UML Marte (AADL) for Embedded Systems, ClientServer-Ports with kind = provided and kind = required correspond to provided and required roles, respectively. WorkloadBehavior from the package PAM_Workload has the role of a service behavior. In ACME [9], ports subsume both required and provided roles of a component. The role of a port in the connector between two components identifies a port as either required or provided. The connectors correspond with assembly connectors. Service behavior in a set of component properties that can be parameterized. The identified core concepts lead to the builder interface we present in the following.

```java
public interface IModelBuilder {
    public void createWorkload(HashMap<String, List<Double>> workload);
    public void createResourceDemand(String service); //
    public void createServiceBehavior(String componentName, String interfaceName, double meanResourceDemand); //
    public void createAssembly(String assemblyName, String componentName); //
    public void createRequiredRole(String componentName, String kind = required, String componentName); //
    public void createProvidedRole(String componentName, String kind = provided, String componentName); //
    public void createAllocation(String componentName, String componentName); //
    public void createHost(String hostName, int numberOCores); //
    public void createInterface(String componentName, String InterfaceName); //
    public void createMethod(String interfaceName, String componentName); //
    public void createAssembly(String componentName, String componentName); //
    public void createServiceBehavior(String componentName, String methodName, List<ExternalCall> externalCalls, String componentName); //
    public void createRequiredRole(String componentName, String componentName); //
    public void createProvidedRole(String componentName, String componentName); //
    public void createAllocation(String componentName, List<Double> workload); //
}
```
The creation methods of the interface, presented above, represent creation and/or connection functionality. The created objects are referenced by other interface methods during model creation. The implementation of creation methods for host, component, and interface require only object instantiation. The implementation of such creation methods, for a concrete language, is straightforward. These basic elements are referred to in the performance model composition process at multiple stages. For example, createMethod references interface to append signatures, createProvide-dRole, createRequiredRole are appended to existing components and reference the previously created interface, createResourceDemand enriches service with internal resource demand. The resource demand is not a parameter as it can be taken from resource demand HashMap using identifier (c.f. Algorithm 1). Other connections include, for example, connectAssemblies requires to add references to connected assemblies to the connection element. The function addCom-ponentToAssembly sets the component for an assembly.

```java
public EObject connectAssemblies(String providingAssemblyName, String requiringAssemblyName);
public void addComponentToAssembly(String assemblyName, String componentName);
```

To relieve the developer of a builder implementation from implementing all getter functions, we introduced Abstract-ModelBuilder which implements all getter functions of the IModelBuilder interface. It stores created elements in hash maps so that they can be referenced within the extraction algorithm. Hence, builder implementations also have to extend AbstractModelBuilder to be compatible with the framework.

### 5. FRAMEWORK IMPLEMENTATION

We build our framework to derive the core aspects of architectural performance models like control flow, resource demands, and workload. Then PMX uses them to trigger the methods of the builder interface to create an architectural performance model. Algorithm 1 represents the performance model extraction receiving the path to monitoring logs and a builder instance. Lines 2-7 extract information
Algorithm 1 Model Extraction Using Generic Builder

```
1: function CONSTRUCT(Path path, Builder builder)
2:   logs ← readLogFile(path)
3:   analyzer ← compose analysis filters
4:   analyzer.analyze(logs)
5:   operationGraph ← analyzer.getOperationGraph()
6:   rds ← analyzer.getResourceDemands()
7:   workload ← analyzer.getWorkload()
8:   buildModel(operationGraph, rds, workload, builder);
9:   builder.save()
10: end function
```

independent of targeted language description which is used in Line 8 to trigger the `construct` method of the builder pattern that triggers calls to builder methods. The extraction of basic information is based on filters that are connected using pipes-and-filter architecture of Kieker [20]. The processing of monitoring logs is triggered in Line 4. The extraction of control flow is based on Kieker filters that extract operation call graph including calls weights that allow to derive call probabilities (for details see [? ]). The resource demands extraction filters for method call times and resource logs. After processing of all logs, the framework triggers the Library for Resource Demand Estimation (LibReDE) [15, 14] 1 to estimate resource demands. In case resource utilization information is available, estimation is based on service demand law [6]. Otherwise, estimation uses a response time approximation approach. To describe the workload, PMX stores for each interface/role at the system border the arrival times. This allows for the creation of empirical workload models as same as for aggregated probabilistic ones.

Algorithm 2 refines access of builder. Lines 2-5 apply Kiekers systemModel containing static system properties (names of hosts, components, interfaces, allocations) which can be received counting divergent identifiers. Each of the named lines iterates over all elements for the type. For example, Line 2 triggers the `createHost interface method for each host. Interaction element creation happens using call graph processing. Vertices represent methods (including information about component and host). Edges represent calls to other methods. Lines 6-23 process all call graph vertices. Lines 7 and 8 create per edge an assembly and adds component to assembly. method executions. Line 22, uses information that is For each outgoing edge, a service call is created.

6. EXPANDABILITY OF FRAMEWORK

Even our framework covers a full extraction story, it not yet covers all modeling techniques and possible extraction techniques. Our implementation is limited by the builder interface (shared concepts) and by the available input information. In the following we sketch how to expand:

Information retrieval methods To include new information retrieval methods without introducing new elements is possible to extend the framework without changing the builder interface. For example, when the monitoring framework enables to measure resource demands per request, it could be included to replace estimation methods. Further, there exist some limitations using standard monitoring log informations. The model creation of PMX builds upon

```
Algorithm 2 Application of builder for Performance Model Generation

1: function BUILDMODEL(systemModel, operationGraph, resourceDemands, workload, builder)
2:   createHosts(systemModel, builder);
3:   createComponents(systemModel, builder);
4:   createInterfaces(systemModel, builder);
5:   createAllocations(systemModel, builder);
6:   for all source : operationGraph.vertices do
7:     component ← source.component.name
8:     host ← source.host.name
9:     assembly ← component + host
10:    builder.addAssembly(assembly);
11:   builder.assign(assembly, component);
12:   for all edge : source.outgoingEdges do
13:     target ← edge.getTargetVertex;
14:     tComponent ← target.component.name
15:     tHost ← target.host.name
16:     tAssembly ← tComponent + tHost
17:     builder.assign(Assembly, tComponent);
18:     builder.connect(assembly, tAssembly);
19:     calls ← outgoing.getExternalCalls();
20: end for
21:   rd ← resourceDemands.get(signature)
22:   builder.addBehavior(component, signature, calls, host, rd);
23: end for
24: end function
```

a probabilistic call graph introducing some limitations. For example, it cannot be said whether a loop behavior has been created using a "for" or "while"-operator. Hence, our framework does neither. Moreover, component relations can be extracted while containments cannot be derived from available measurement information. The measurement logs do not uncover whether a method call has been triggered by an interface call or triggered by an event listener. Additional use of source code information could improve the extracted model and retract limitations.

Expanding information sources, additional runtime information, e.g. garbage collection [23] could be included. Some events, like garbage collection, may occur rarely. It depends on chance if such information is included in measurements. In the direction of rare events, the framework could be extended with outlier detection mechanisms.

New features It might be required for some applications to expand PMX to extract additional language features available only in a subset of formalisms. This requires to expand the builder interface. We propose to use template method [8] extending the skeleton of the `CONSTRUCT method, deferring building again to builders. It is important to provide an empty default implementation to not break other builders and remain downward compatible. Extending the common set of core modeling techniques, there exist concepts which are integrated differently. For example, parametric dependencies (e.g. parametric resource demands and branching probabilities) have been integrated differently. Hence, extensions in this direction should not be required for every builder.

Modeling alternatives Some languages offer to model the same systems properties in various ways. For example, DML offers different granularities (black-box, coarse-
7. STATE OF THE ART

There exist various quality-aware architecture description languages, often focused on performance. A comparison of architectural performance modeling languages has been performed e.g. in [3]. The extraction of such models can be grouped by aspects (e.g., control flow and resource demands), each providing different possibilities for their extraction [21]. For example, the options for the extraction of resource demands include direct measurement and many estimation techniques [24, 14]. In [16] an agent-based model update for online scenarios has been proposed, that updates parts of the model in different frequency and can be applied supplementary to the presented approach. The approach in this paper compares mainly to performance model extraction approaches targeting architectural models. Examples include e.g. [22], [25]. Compared to our contribution, the named approaches are limited to a single modeling language and rely on commercial monitoring infrastructure. In addition to measurement-based extraction, there exist approaches that leverage both static code analysis and measurements to extract architectural performance models [11]. The static analysis allows for the detection of architectural information from source code, e.g., control flow causality for loop counts. However, they require an active profiling of the application to detect these causalities.

8. EVALUATION

To evaluate our framework, we developed two builder implementations of the builder interface. We selected PCM and DML. Both have been applied at design time and runtime scenarios. While the first is more related to design time, the second is more a run time architectural model.

Setting To evaluate our approach, we selected the Pet Clinic application representing a portal for vet appointments. We deployed it on a Dell Power Edge R815 with 48 cores, each core equipped with an Opteron 6174 CPU 2.6 GHz. The application was running on an Ubuntu 14.04.5 VM equipped with 16 GB RAM (to be no bottleneck) and an assignment of 42 cores. We modified the "browse and edit" workload shipped with the application removing the "edit" operations to avoid database contention (due to locking). The resulting workload is open with an exponentially distributed arrival rate. Each vet customer calls the following sequence of interfaces: The vet visits the front page welcomeGET, looks at all vets showVetListGet, then searches pet owners initFindFormGet, processFindFormGet and displays all pet owners using showOwnerGET. Then the vet triggers two times a specific pet owner page processFindFormGet. Besides workload, we modified the application to cache the vet catalog once at startup to improve performance.

The load driver has been deployed at the application machine to avoid a network bottleneck that manifested itself for transaction rates above 400 workload requests per second. We employed a JMeter instance as a load driver and deployed it on a separate VM which received the remaining six cores.

Results For the described setting, we measured a calibration workload at low utilization used to trigger performance model extraction applying the builder implementations for PCM and DML. Then we performed benchmark measurements using the Kieker monitoring framework for different load scenarios and compared them to simulation results of the extracted models. We derived performance metrics in the following way: Kieker log files where filtered using the Descartes Query Language [5] to derive response times. Utilization has been measured using the Linux top command.

DML Analysis used a transformation to Queueing Petri Nets (QPNs) [13] and simulation using SimQPN simulation engine. PCM analysis was performed using Simulizar [2] simulation engine. Table 2 compares the measured and predicted response times and utilizations. For the evaluated scenarios we receive accurate model-based predictions. The deviation for utilization is below 2% and below 10% for response times.

9. CONCLUSION

In this paper, we present a framework for the extraction of architectural performance models generalizing over the target modeling language. Using the presented approach, the user only has to implement our builder interface to create a performance model generation tool for a specific modeling language. This lowers the effort for such a tool tremendously through a high level of reuse. Our approach enables an easy comparison of architectural performance modeling languages and access to different tool chains. Our evaluation presents accurate prediction results for the extracted models. Source code as well as compiled eclipse plugins for the framework as well as model specific extraction tools (including concrete builder implementations) have been made available online. 3

Acknowledgments

This work was co-funded by the German Research Foundation (DFG) under grant No. KO 3445/11-1.

3http://descartes.tools/pmx

---

Table 2: Evaluation Results Pet Clinic Case Study.

<table>
<thead>
<tr>
<th>Workload in requests per second</th>
<th>CPU utilization (average)</th>
<th>session response time in ms (average)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
<td>DML</td>
</tr>
<tr>
<td>1(calibration)</td>
<td>0.33%</td>
<td>0.35%</td>
</tr>
<tr>
<td>732</td>
<td>25.22%</td>
<td>24.64%</td>
</tr>
<tr>
<td>940</td>
<td>33.12%</td>
<td>31.64%</td>
</tr>
</tbody>
</table>
References


