1. **Palladio component Model**

Palladio Component Model (PCM) is a component-based software architecture approach, that can be used to predict and evaluate the performance and the reliability of component-based software systems at design stage. For performance prediction, it makes it possible to analyse components of the software architectures before implementation and thus it can for example detects bottlenecks, predict response time and predict throughput. For software reliability, PCM specifies metrics like detecting the probability of failures on demand.

**1.2. Palladio Models**

To allow the prediction of the Performance and the reliability of a software system, PCM defines four different roles to create four different models **Figure 1**. The Repository model, which is created by developers. Developers specify and implement components, interfaces, provided roles, required roles and signatures for the repository. In the context of PCM, a component is by definition a block of software, that can be composed, deployed and customized without requiring the understanding of its internals. Moreover, the roles in PCM specify the relation between the components and the interfaces. Developers should also specify the internal behaviour of the components in term of Service Effect Specification (SEFF). The Assembly model is created by software architects. Software architects use the existing components in repository to create the software system. The allocation model is created by system deployers. Deployers are responsible for specifying the environment resource, like Servers, CPUs, HDDs and network connection. Afterwards, they decide which assembly can be deployed in which resource. The usage model is created by domain experts. Domain experts provide information about the interaction between the system and the users. Additionally, domain experts can define the relevant critical usage scenarios and the inputs parameters values.

In the context of this thesis, we will use the information provided by SEFFs. Thus, we will explore the related concepts based on [2] and [1].

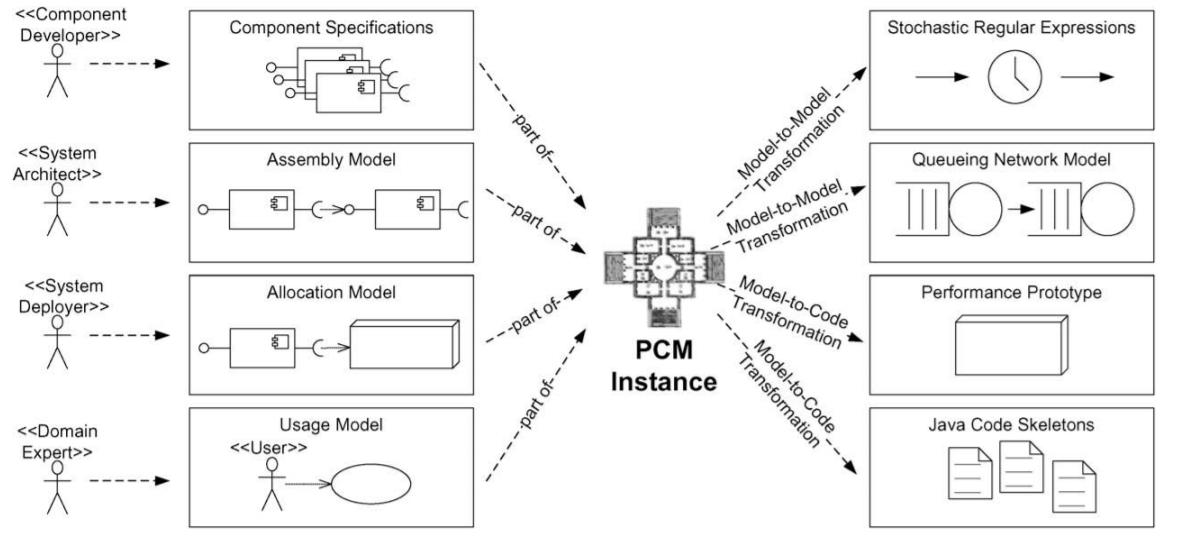
**1.2. SEFFs**

Service Effect Specification (SEFF) were firstly presented by [1], which describes like a UML Activity Diagram the control flow of component services. For each provided service, SEFF describes how services in the required interface are called in the provided service. Moreover, as mentioned before, the components in PCM can be used without understanding their internals and thus as a black box. However, SEFF turns a component to a gray box by describing the behaviour of its provided services.

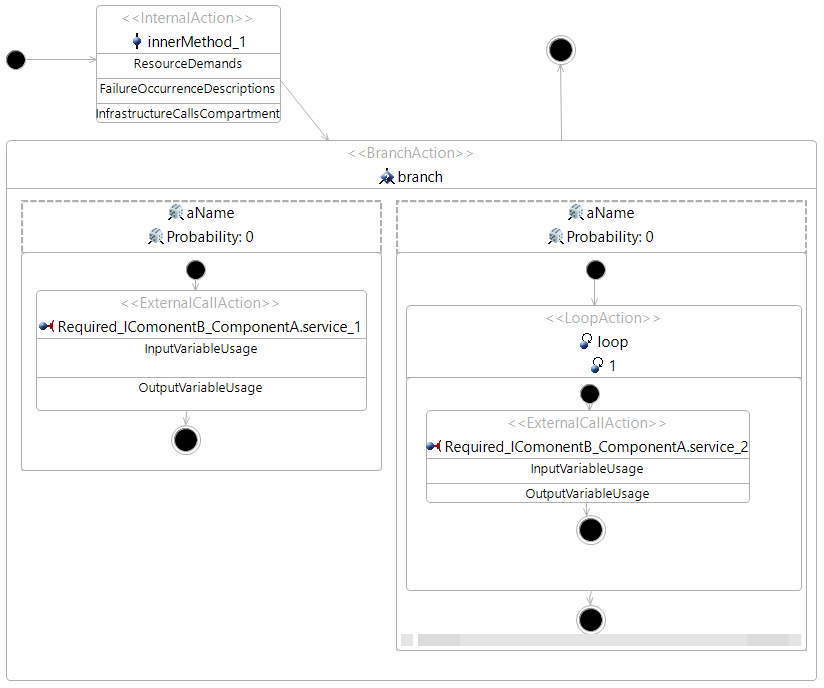
The ResourceDemandingServiceEffectSpecifaction (RDSEFF) used in PCM to predict the performance, is an extension of SEFF. RDSEFF offers the possibility to add performance inputs values associated to each activity of SEFF. Moreover, to each component provided service, developers can specify a RDSEFF in order to describe how the service uses the hardware/software resources and how it calls the component’s required services.

In the rest of this thesis, we will indicate RDSEFF as SEFF in order to avoid ambiguity. The principal elements of SEFF will be explored in the following. SEFF includes the so called *ExternalCallActions* which present the call of a required Service within the SEFF. *InternalActions* are used within SEFF to abstract the internal computation of the component. Moreover, an internal action can be defined as a set of successive instruction, that do not include any external call from other components. *LoopActions* within SEFF are specified to indicate the number of times the sub control flow within the loop is going to be executed. *BranchActions* models the branches within the control flow. The execution of a branch can be decided either on an input parameter or a probability. *InternalActionActions* refers to the internal behaviour of a component, that can be only used by the services of this component.

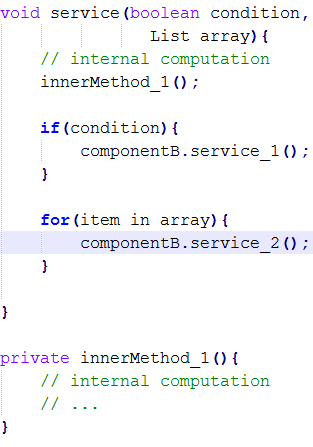
To have a better understanding of these concepts, we put them together in an example Figure 2, which depicts on the right hand the source code and on the left hand the corresponding SEFF model. The implementation of the Service *service()* starts with an internal call of an inner method. Since the inner method *innerMethod\_1()* represent an internal computation and does not make any external call, it is seen an internal action within SEFF. Within the next branch is seen as a *BranchAction*, because there is a call of the service *service\_1()* of the component *componentB*. In the second branch transition, there is loop, which makes call of the external service *service\_2()*, that’s why it’s represented as a *LoopAction* within the corresponding SEFF.



**Figure 1: PCM Models and transformations**



**Figure 2: Example of source code and SEFF**



1. **Vitruvius**

The Vitruvius approach is a view-based [7] engineering approach, which was introduced in [3, 4]. Vitruvius can be used to keep the models of a system consistent. In Model Driven Development (MDD) [5, 6], the whole system can be represented by different models, that describe different aspects of the system, even the source code of the system is seen as a model. These models can be changed separately and became inconsistent, for example if the source code of a system has changed, the architecture model of the system has to be also updated. In this case, architectures must capture the changes in the source code and update accordingly the associated architecture model. Therefore, Vitruvius performs this task automatically by defining consistency rules, that define how the changes in a model must be transformed to another model.

Vitruvius uses views to make access to the models possible. This approach reduces the complexity of dealing with the whole models, because views present only a part of the model. Therefore, developers can focus only on the relevant part of the system.

Vitruvius defines the so called Virtual Single Underlying Model (VSUM), which contains all information related to a system. VSUM contains the metamodels instances, that are used within a system. Moreover, Vitruvius provides the correspondence metamodel, which offers the possibility to map between correspondent elements of different model, like SEFFs in PCM and methods in Java source code.

To enable the process of keeping models consistent within Vitruvius, developers should provide and implement two concepts defined by Vitruvius, namely Domains and Application. A Vitruvius Domain represents a defined metamodel in the system and provides information for its use within the Vitruvius framework. Vitruvius Domains can be reused, for example the Domain of Java metamodel can always be used in systems that are implemented using java. Vitruvius Applications determine the relation between two Domains. They specify how changes in a Domains metamodel instance should be transformed to the instance of another Domains metamodel instance. Vitruvius Applications can be also reused in many environments, if the are needed.

1. **Java Model Parser and Printer**

Java Model Parser and Printer (JaMoPP) [10] is a parser and printer for the Java language. JaMoPP defines a complete metamodel for the Java language based on the metamodeling language Ecore [11]. JaMoPP parser allows to parse a Java source code into a Java model, and the JaMoPP printer allows to print a Java model into Java source code. JaMoPP can convert Java source code into an EMF model, which can be manipulated using model driven techniques, like model’s transformation. Using JaMoPP, we can for example, parse a Java file, which contains Java source code, add new Java statements after or before an existing statement and print back the changes in the Java source code file. The creation of JaMoPP is based on EMFText [12], which allows to define text syntax for languages described by an Ecore metamodel.

1. **Eclipse Modeling Framework**

Eclipse Modeling Framework (EMF) is a framework, that enable the creation of metamodels based on Ecore. Moreover, EMF provides the following facilities:

* Source code generation from metamodels.
* Generate an instance of eclipse in order to instantiate metamodels and edit them.

1. **Automated Coevolution of Source Code and Software Architecture Models**

The approach of co-evolution of source code and component-based software architecture presented in [8, 9] is based on the platform Vitruvius present above. It gives developers and architectures the possibility to keep the architecture and the source code of a software system consistent. The Consistency in this approach is kept in both directions, that means, if developers changed the source code of a system, the corresponding architecture will automatically be changed, and if architectures updated the architecture of the system, the source code will also automatically be updated. The co-evolution approach uses PCM models as architectures models. Therefore, it describes the behaviour of the source code in term of SEFF. Hence, a special objective of the co-evolution approach is to keep incrementally the behaviour of the source code consistent with the source code itself. The incremental up-to-date of the behaviour model of the source consists of only the part of the SEFF that corresponds to the changes performed in the source code. For example, if the developers update a service S of Component A, only the SEFF of the service S will be reconstructed but not the whole SEFF model of the system.

The co-evolution approach uses two different concepts to preserve different modes consistent, namely model-driven engineering and change-driven engineering. Moreover, the co-evolution approach uses the concepts of Vitruvius to keep tracks of the models, that represents the system.

The co-evolution approach considers all involved artifacts as models. Hence, the model-driven engineering is used in this approach as a main concept. In mode-driven engineering, all artifacts of the system are represented by models and the models are centric in the development process. That means, that the source code must be also represented within the co-evolution approach as a model. Therefore, in the case of Java, the co-evolution approach uses JaMoPP to parse the Java source code to a mode, on which the change-driven techniques can be applied, like model to model transformations.

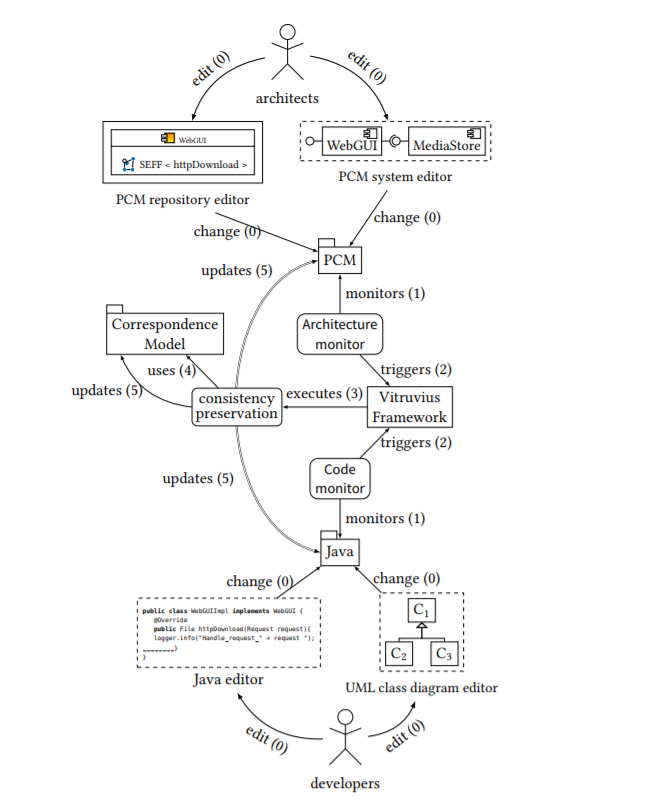
The co-evolution approach uses change-driven engineering in order to reacts on changes, that the users perform on models, especially the architecture model and the source code model. These changes are monitored, converted in a Vitruvius change model and propagated. The propagated changes can be captured and transferred using consistency preservation rules to the target model. The consistency preservation rules are bidirectional and defined between the architecture metamodel and the source code metamodel.

The co-evolution approach was applied to the Palladio Component Model (PCM) as an architecture model and Java source code. However, the concepts presented in this approach can be applied to other component-based architecture model, like UML component-diagrams, and other object-oriented languages. In the following, we will review the application of the co-evolution approach to PCM and Java source code and the steps, that are required to keep PCM models’ instances and Java source code consistent.

Figure 3 shows the steps, that are used to keep architecture models and the source code consistent:

* Step (0): users change either the PCM repository, the PCM system or the Java source code.
* Step (1): monitors capture changes on the models
* Step (2): monitors trigger Vitruvius Framework and pass it the changes
* Step (3): based on these changes, the Vitruvius framework executes the consistency preservation transformations.
* Step (4): the transformations use information from the changes and the correspondence model to execute the preservation rules.
* Step (5): the transformations update the models

When it comes to an automatic update of SEFF in step (5), the co-evolution approach executes an incremental SEFF reconstruction step instead of transformation.



**Figure 3: The steps used in co-evolution approach to keep architecture models and the source code consistent**

1. **Kieker Monitoring**

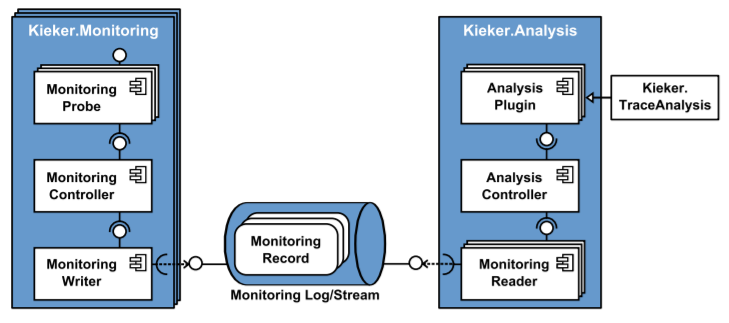
Kieker is an extensible Java-based application performance monitoring and dynamic software analysis framework [13]. Figure 4 shows the architecture of Kieker framework. It’s composed from two main components, namely the monitoring component and analysis component. The analysis component can be used to read monitoring data, analyze and visualize them for a certain purpose, like generating UML sequence diagram, dependency graphs or Markov chains.

The monitoring component is responsible for source code instrumentation, data collection and data logging. The Monitoring probes are responsible for collecting the monitoring data and send them to the monitoring controller component, which instantiates a monitoring record for every probe. The Monitoring writer component receives the monitoring records from the monitoring controller and serialize them to the monitoring log/stream.

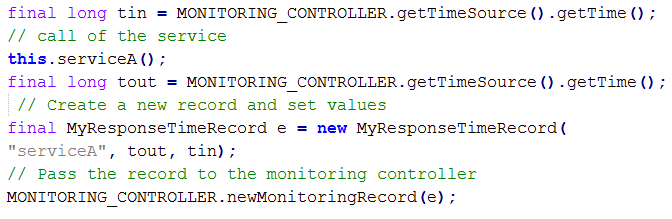
A monitoring Record represents the measurement data gathered in a single measurement. Kieker provides the possibility to store different types of records for different types of probes. Kieker offers the possibility to create customized probes. We can create new probes either manually by extending the interface *IKiekerMonitoringProbe* or automatically by using the Instrumentation Record Language (IRL) [14]. For example, in one record, we can store the signature and the response time of a method, in another record, we can persist the response time of specific number of statements inside a method.

A monitoring Probe represents the monitoring logic used to collect measurement data from the application. There are two ways to use probe within Kieker. There is the manually instrumentation, which consists of mixing the instrumentation logic with the business logic of the application. Figure 5 show an example, in which the monitoring probe is implemented by mixing monitoring logic with business logic, in this example we create a probe that logs the name of the called service, the start time and end time. Kieker includes also probes based on Aspect Oriented Programming (AOP) [15], which helps to separates the instrumentation logic from the business logic. Kieker defines AOP based monitoring probes like *OperationExecutionAspectAnnotation and OperationExecutionAspectAnnotationServlet.* Figure 6 shows how the instrumentation looks like, when using AOP for probes implementation.

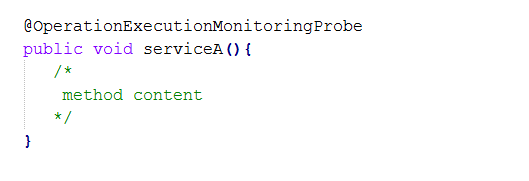
Using AOP to instrument the source code has the advantage of separating concerns. However, this technique has a limitation, when it comes to the monitoring of certain statements of inside a method, like monitoring the number of executions of a loop or the probability of a branch execution, because the annotation possibilities on these cases are out of box .



**Figure 4: overview of the Kieker's architecture**

****

**Figure 5:examle of manual instrumentation of source code**



**Figure 6: example of a Kieker Probe using AOP**

1. **Source Code Model eXtractor**

Source Code Model eXtractor (SoMoX) is a reverse engineering approach, which has been developed by Krogmann [16]. SoMoX is able to reverse-engineer software component architectures. Moreover, SoMoX can extract a PCM repository from source code and creates a PCM system derived from the repository. The repository created by SoMoX contains mainly components, interfaces, roles and SEFFs. Thus, the results of reverse engineering of SoMoX depend strongly on the project implementation to reverse-engineer which means that SoMoX delivers best results, if the analysed source code followed a component-based architecture.

In the following, we will review only the features of SoMoX that are involved in the context of this thesis.

In order to create the architecture of a software system, SoMoX reverse-engineer the software system using the following steps:

* Parse the source code into a model
* Detect components and interfaces using metrics
* Detect data types and signatures using metrics
* Reconstruct the SEFFs

For source code parsing in the first step, SoMoX uses JaMoPP to create an EMF model of the Java source code that can be manipulated. In this thesis, we will also use JaMoPP source code parsing and manipulation purpose.

For components and interfaces detection, SoMoX uses various source code metrics and combine them to determine detection strategies for architecture elements. These metrics must be given each a value between 0 and 100 by the user of SoMoX. The value of the metric tells SoMoX the impact factor, for example the value 0 of a metric means that the impact factor is low, whereas the value 100 of the metric means that the impact factor is high.

The reconstruction of SEFFs which aims to reverse-engineer the statical behaviour of the source code is done by analysing the methods of the source code. This step has been extended by Langhammer [17] and its results are used in thesis. In the following, we will explain briefly how the reconstruction of SEFFs is done within SoMoX and how it was extended by Langhammer.

**7-1. SoMoX SEFFs reconstruction**

For SoMoX SEFFs reconstruction which is done in the last step of the reverse-engineering process, SoMoX uses two models which were created in the first and the second steps. The first model is the Java source code model which was created in the first step and the second is the so-called Source Code Decorator Model (SCDM) which was created in the second step. The SCDM contains the information that map between the source code model elements and the reverse-engineered architectural model elements.

In order to create the SEFF of a method, SoMoX analyses the source code of the method which was detected as a provided method of a component. This analysis is performed in two steps which are described in the following.

In the first step, the SoMoX visits all the method calls within the analysed method and classified them in three categories. The first category contains the component-external method calls which are considered as required roles. The second category contains library calls which considered as calls to a third-party library like *java.lang*. the second category contains the component-internal calls which are calls to the inner methods of the component.

In the second step, SoMoX creates the SEFF for the method. To do so, SoMoX visit again the statements in the source code of the method in order to find the following SEFF elements if they exist: *ExternalCallActions*, *BranchActions*, *LoopActions* and InternalActions. BranchActions and LoopActions are created for branches and loops. A Branch or a Loop is considered as *BranchAction* or *LoopAction* if it has an external method call, else it will be combined with an *InternalAction*.

**7-2. Incremental SEFF reconstruction**

Langhammer has proposed in his Co-evolution approach an incremental SEFF reconstruction approach which can build the SEFFs for only the changed parts of the source code. In contrast to SoMoX, the incremental SEFF reconstruction neither require the parsing of the complete project source code nor the SCDM. Moreover, the SEFF of the smallest unit that can be currently incrementally reconstructed is the SEFF of a method.

The incremental SEFF reconstruction is integrated in the co-evolution approach that means it can use functionalities and information provides by Vitruvius. Moreover, the incremental SEFF reconstruction is done in change-driven way, that means change that happened in the source code are captured in a Vitruvius change model instance and can be treated in order to regenerate the SEFF of the method in which they belong. To classify the method calls which is necessary for SEFF reconstruction, Langhammer uses the current preservation rules and information from the Vitruvius correspondence model. More details on this step can be found in his thesis [17].

In the following we explain the steps of the incremental SEFF reconstruction based on the its implementation. Figure

Are Changes relevant

Remove all abstract actions of the old method from the correspondence model

Run SoMoX to extract SEFF of the new method

Reconnect the newly extracted SEFF elements with the old elements

Create new correspondences between the new SEFF elements and the new method

1. **Continuous Integration of Performance Model**

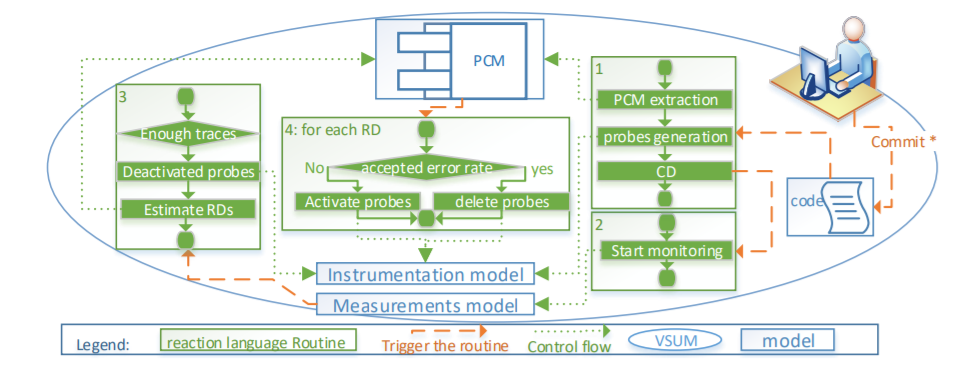
Continuous Integration of Performance Model (CIPM) is an approach proposed by Mazkatli and Koziolek [18] in order to extract incrementally and iteratively the Performance Model from the source code and enrich it by the Performance Model Parameters (PMPs). Furthermore, CIPM aims to keep the source code and the extracted Performance Model consistent during the system development. CIPM extends the Continuous Integration (CI) and the Continuous Deployment (CD) of the source code with a continuous integration of the performance model.

To achieve that Mazkatli and Koziolek used the Coevolution approach developed by Langhammer (section 5) which uses the Palladio Performance Model as a Performance Model and the Vitruvius Platform to keep incrementally and in a change-driven way the source code and the corresponding Performance Model consistent. CIPM uses likewise a change-driven way to enrich the extracted PM by the Coevolution process with PMPs. It defines the consistency rules that minimize the monitoring of the source code execution and the analysis overhead. In addition, it specifies a self-validation process that validates the estimated PMPs. To release that, CIPM automates four activities that are executed in each iteration [Figure 7]. In this section we will describe only the first two activities because we based our work on them, more details on the other activities can be found in [18].

The activities in CIPM are represented by the Reaction Language (RL) routines. RL is used in Vitruvius to describe the consistency rules and it’s based on two concepts, namely reaction and routine. A reaction specifies changes and triggers a routine that contains the consistency rules that should be executed in reaction to these changes.

The first activity in CIPM is responsible for updating the structure of the performance model, the usage model and the probes used in the activity two. The structure of the performance model is updated based on the work of Langhammer [8]. The usage model is extracted incrementally from the test cases. For the probes generation CIPM specifies an Instrumentation Metamodel (IMM) and added it to VSUM. IMM contains and manages the instrumentation points and the weaving information based on Aspect-oriented Programming (AOP) [19]. CIMP defines the consistency rules that keep the instrumentation model and the source code consistent. For example, a new probe can be added to the instrumentation model, when a part of the source code was added that corresponds to a new SEFF element.

The second activity uses the probes from the first activity, instruments the source code and creates the monitoring data. For the monitoring data, CIPM specifies a Measurement Metamodel (MMM) and added it to VISUM. MMM describes the data structures of the different monitored data as well as the consistency rules to keep it consistent with the IMM. For example, after the monitoring phase is done, if a SEFF element had enough monitoring data, this element has to be deactivated in the instrumentation model.



**Figure 7: Automated CIMP activities**

1. **Iterative Performance Model Parameter Estimation Considering Parametric Dependencies**

In his master thesis, Jan [] has created an approach that can estimate Performance Model Parameters taking into account the parametric dependencies. The approach is based on the vision presented by Mazkatli and Koziolek (section 8), it extends precisely the concepts introduced in the activities three and four in Figure 7. The approach is designed to make iteratively the estimation and thus reduce the overhead resulting from the estimation for the whole system.

Jan uses the Palladio Performance model as a Performance Model for his approach. Since the Palladio Performance Model is expressed in terms of SEFF, Jan estimates the performance model parameters for loop iteration, resource demands, branch transitions and external call arguments. He specifies diverse predictive models for estimating Performance model parameters. He uses decision threes to create a predictive model for branch transitions. For loop iterations and resource demands he uses regression analysis. The predictive models are related with service call arguments. The predictive models can be transformed into stochastic expressions in order to use them for enriching performance model.

This approach uses as inputs the monitoring data that are generated from instrumenting and executing the source code that corresponds to the performance mode for which the estimation is done. The monitoring data are structured in records that contain diverse information about diverse source code elements like loop record, branch record response time record etc. Since the execution depends strongly on the monitoring data, in order to keep the estimation iterative, the monitoring must be also done iteratively. this includes also the fine-grained, automatic, iterative instrumentation of the source code. the instrumentation has to be fine-grained because the estimation requires specific information about the program elements like the number of loop execution, the response time of a specific source code that corresponds to an internal action. This thesis provides an approach that can generate iteratively the monitoring data for performance model. These monitoring data can be used by the approach of Jan to make the estimations.

**References:**

[1] Koziolek, Heiko, Jens Happe, and Steffen Becker. "Parameter dependent performance specifications of software components." *International Conference on the Quality of Software Architectures*. Springer, Berlin, Heidelberg, 2006.

[2] Becker, Steffen, Heiko Koziolek, and Ralf Reussner. "The Palladio component model for model-driven performance prediction." *Journal of Systems and Software* 82.1 (2009): 3-22.

[3] Erik Burger. “Flexible Views for View-based Model-driven Development”. PhD thesis. Karlsruhe, Germany: Karlsruhe Institute of Technology, July 2014.

[4] Kramer, Max E., Erik Burger, and Michael Langhammer. "View-centric engineering with synchronized heterogeneous models." *Proceedings of the 1st Workshop on View-Based, Aspect-Oriented and Orthographic Software Modelling*. ACM, 2013.

[5] Thomas Stahl et al. Modellgetriebene Softwareentwicklung : Techniken, Engineering, Management. 1. Aufl. Heidelberg: dpunkt-Verl., 2005.

[6] D.C. Schmidt. “Guest Editor’s Introduction: ModelDriven Engineering”. In: Computer 39.2 (Feb. 2006), pp. 25–31

[7] Thomas Goldschmidt, Steffen Becker and Erik Burger. “View-based Modelling - A Tool Oriented Analysis”. In: Proceedings of the Modellierung 2012, Bamberg. Mar. 2012.

[8] Michael Langhammer and Klaus Krogmann. “A Co-evolution Approach for Source Code and Component-based Architecture Models”. In: 17. Workshop Software-Reengineering und-Evolution. Vol. 4. 2015

[9] Michael Langhammer. “Automated Coevolution of Source Code and Software Architecture Models”. PhD thesis. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT), 2017. 259 pp.

[10] Heidenreich, F., Johannes, J., Seifert, M., & Wende, C. (2009, October). Closing the gap between modelling and java. In *International Conference on Software Language Engineering*(pp. 374-383). Springer, Berlin, Heidelberg.

[11] Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: Eclipse Modeling Framework, 2nd Edition. Pearson Education (2008)

[12] Heidenreich, F., Johannes, J., Karol, S., Seifert, M., & Wende, C. (2009, June). Derivation and refinement of textual syntax for models. In *European Conference on Model Driven Architecture-Foundations and Applications* (pp. 114-129). Springer, Berlin, Heidelberg.

[13] van Hoorn, A., Rohr, M., Hasselbring, W., Waller, J., Ehlers, J., Frey, S., & Kieselhorst, D. (2009). Continuous monitoring of software services: Design and application of the Kieker framework.

[14] Jung, Reiner. *An instrumentation record language for kieker*. Tech. rep. Kiel University, 2013.

[15] G. Kiczales, J. Lamping, A. Menhdhekar, C. Maeda, C. Lopes, J.-M. Loingtier, and J. Irwin (1997). Aspect-oriented programming. In Proceedings of the 2007 European Conference on Object-Oriented Programming (ECOOP ’97), volume 1241 of Lecture Notes in Computer Science, pages 220–242. Springer

[16] Klaus Krogmann. Reconstruction of Software Component Architectures and Behaviour Models using Static and Dynamic Analysis. Vol. 4. The Karlsruhe Series on Software Design and Quality. KIT Scientic Publishing, 2012. doi: 10.5445/KSP/1000025617. url: [http://digbib.ubka.uni-karlsruhe.de/vo lltexte/1000025617](http://digbib.ubka.uni-karlsruhe.de/vo%20lltexte/1000025617)

[17] Langhammer, Michael. "Automated Coevolution of Source Code and Software Architecture Models." (2017)

[18] Mazkatli, Manar, and Anne Koziolek. "Continuous Integration of Performance Model." *Companion of the 2018 ACM/SPEC International Conference on Performance Engineering*. ACM, 2018.

[19] Gregor Kiczales et al. “Aspect-oriented programming”. In: ECOOP 97—Object-oriented programming (1997).