Towards a Holistic Approach for Integrating Middleware with Software Product Lines Research

Aniruddha Gokhale
ISIS, Dept. of EECS
Vanderbilt University
Nashville, TN 37235, USA
gokhale@dre.vanderbilt.edu

Akshay Dabholkar
ISIS, Dept. of EECS
Vanderbilt University
Nashville, TN 37235, USA
aky@dre.vanderbilt.edu

Sumant Tambe
ISIS, Dept. of EECS
Vanderbilt University
Nashville, TN 37235, USA
sutambe@dre.vanderbilt.edu

ABSTRACT
Prior research on software product lines (SPLs) in different domains (e.g., avionics mission computing, automotive, cellular phones) has focused primarily on managing the commonalities and variabilities among product variants at the level of application functionality. Despite the fact that the application-level SPL requirements drive the specializations (i.e., customizations and optimizations) to the middleware that host the SPL variants, middleware specialization is seldom the focus of SPL research. This results in substantial and ad hoc engineering efforts to specialize middleware in accordance with the functional and quality of service (QoS) requirements (e.g., latency, reliability) of the product lines. To overcome these problems, this paper highlights the need to unify middleware specialization issues within SPL research, and argues for new research directions in modularization and generative programming techniques that can account for the deployment and runtime issues, such as QoS and resource management. Preliminary ideas demonstrating how feature-oriented programming and model-driven development tools together can address these challenges are presented.

Categories and Subject Descriptors
D.2.8 [Software Engineering]: Metrics—complexity measures, performance measures

General Terms
Product Lines, Middleware Specializations, Modularizations

Keywords
Generative programming + Product lines, FOP/AOP + MDD

1. INTRODUCTION
Research on software product lines (SPLs) [7] has focused primarily on managing the commonalities and variabilities [8] in application-level functionality of product variants. Generative programming [9] and modularization techniques, such as feature-oriented programming (FOP) [20] and aspect-oriented programming (AOP) [12], play an important role in composing and synthesizing product variants from modularized units called features and aspects.

Middleware is an important asset for SPLs across many domains, such as avionics (e.g., Boeing’s Bold Stroke architecture [23]), telecommunications (e.g., Ericsson’s family of carrier class switches [1]) and even cell phones (e.g., Nokia or Motorola’s family of cell phones). Middleware manages the quality of service (QoS) (e.g., latency, reliability, security), and resource management (e.g., bandwidth, CPU, memory) issues in product variants of a SPL. SPL developers tend to rely on standardized, general-purpose middleware, such as not limited to J2EE, .NET Web Services, and CORBA, since these middleware provide a reliable, robust, low cost and low maintenance solution with the added benefit of feature-richness, flexibility, and high degree of configurability.

Although existing research in SPLs has significantly improved the quality of product lines, and reduced their development and maintenance costs, these research efforts have seldom addressed the challenges in effectively using middleware for SPLs. Addressing middleware challenges as part of SPL research is necessary since the feature-richness and flexibility of general-purpose middleware often becomes a source of excessive resource consumption and a lost opportunity to optimize for significant performance gains and/or energy savings in SPLs. Moreover, it is infeasible for general-purpose middleware to provide solutions to all possible domain-specific requirements since they are developed with the aim of broader applicability. Developing proprietary middleware for SPLs, however, is not a viable solution due to the excessively high development and maintenance costs.

In the current state of the art these limitations are addressed through significant but often ad hoc engineering efforts at specializing (i.e., customizing and optimizing) general-purpose middleware. To overcome these deficiencies, there is a compelling need for SPL research to consider middleware platforms as an integral part of the SPL engineering processes and methodologies. This in turn argues for new research directions in modularization and generative programming techniques that account for QoS and resource management challenges, which are inherently deployment and run-time problems, while most generative/modularization techniques are limited to design-time.

This paper proposes an integrated SPL methodology that incorporates capabilities for middleware specialization. Specialization is a process that manipulates general-purpose middleware in accordance with the commonalities and variabilities of an SPL by (a) adding custom features supplied by the application, (b) pruning unwanted features, and (c) optimizing the resulting middleware to address QoS and resource requirements of SPLs. Our approach is based on exploiting a hitherto before untapped algebraic structure of middleware by synergistically integrating (a) Origami ma-
trices [4], which provide a formal representation for feature composition, interaction and refactoring [14], (b) Aspects [12], which modularize software that exhibits crosscutting characteristics into reusable features, and (c) Generative programming [9], which promotes automation in middleware specialization.

The remainder of this paper is organized to portray our vision of middleware specialization shown in Figure 1. Section 2 determines the problem space for middleware specialization; Section 3 describes the details of our holistic approach to combining middleware specializations with SPL research; and Section 4 provides concluding remarks and discusses open research issues.

2. THE PROBLEM SPACE FOR MIDDLEWARE SPECIALIZATION

This section helps to define the problem space for middleware specialization in the context of SPLs.

2.1 Middleware System Model

The concept of middleware was born with the aim to shield applications from variabilities in lower-level artifacts, such as hardware, networks and compilers of programming languages. Years of middleware research resulted in a middleware model approximated by Figure 2. Middleware is made up of layers of software targeted to perform specific activities. At the bottom, the host infrastructure layer (e.g., a Java virtual machine or the ACE [21] middleware) shields developers from the differences in operating systems and hardware. Next, the distribution layer (e.g., CORBA or Java RMI) provides features for location transparency, request processing, and data marshaling, among others. The common services (e.g., CORBA Naming or the UDDI discovery service) include features, such as naming, transaction, fault tolerance and real-time, etc. At the top, the domain-specific middleware layer is tailored to a particular domain, such as avionics.

2.2 Survey of Related Research

Now we survey and organize related work along different dimensions that we observe to be prevalent in middleware specialization research.

- **Eliminating overhead of object-orientation**: Lohmann et. al. [15] argue that the development of fine-grained and resource-efficient system software product lines requires a means for separation of concerns [25] that does not lead to extra overhead in terms of memory and performance. The overhead of object-oriented programming (OOP), e.g., due to dynamic binding and method dispatch, is not acceptable for some embedded systems. Aspect-oriented programming (AOP) [12] is shown to eliminate this overhead. Aspects are modularized pieces of code that traditionally are scattered across application code.

- **Aspects for footprint reduction**: AOP provides a novel mechanism to reduce footprint by enabling crosscutting concerns between software modules to be encapsulated into user selectable aspects. FACET [11] identifies the core functionality of a middleware framework and then codifies all additional functionality into separate aspects. To support functionality not found in the base code, FACET provides a set of features that can be enabled and combined subject to some dependency constraints. By using AOP techniques, the code for each of these features can be woven at the appropriate place in the base code.

- **Combining modeling and aspects for refinement**: the Modelware [27] methodology adopts both the model-driven architecture (MDA) [17] and AOP. Borrowing terms from subject-oriented programming [10], the authors use the term *intrinsic* to characterize middleware architectural elements that are essential, invariant, and repeatedly used despite the variations in the application domains. They use the term *extrinsic* to denote elements that are vulnerable to refinements or can become optional when the application domains change. Modelware advocates the use of models and views to separate intrinsic functionalities of middleware from extrinsic ones. Modelware considerably reduces coding efforts in supporting the functional evolution of middleware along different application domains.

- **Combining computational reflection and aspects**: computational reflection is an efficient and simple way of inserting new functionality into reflective middleware, such as LOpenOrb [5]. It uses a meta-object protocol to abstract away the implementation details so that it is necessary only to know the components and interfaces. To conserve resources and provide dynamic adaptation, AOP can be used to specialize the reflective middleware. Aspects that are not in the application code can be dynamically inserted using a meta-object protocol.

- **Layer collapsing and bypassing**: In a typical middleware platform every request passes through each layer, whether or not the services provided by that layer are needed for that specific request. This rigid layered processing can lower overall system throughput, and reduce availability and/or increase vulnerability to security attacks [19]. For use cases where the response is a simple function of the request input parameters, bypassing middleware layers may be permissible and highly advantageous. Devanbu et. al [26, 19] have shown how AOP can be used to bypass middleware layers.

- **Importance of lifecycle stages**: Traditionally, performance problems in middleware layers have been addressed by optimizing the source code and data structures. Edicts [6] is an approach that shows how optimizations are also feasible at other application lifecycle stages, such as deployment- and run-time. Just-in-time middleware customization [28] shows how middleware can be customized after application characteristics are known. These efforts discover the configuration of the target environment and compose only the necessary modules that are best suited among alternatives and configure them in the most optimal way.
3. INTEGRATING MIDDLEWARE SPECIALIZATIONS WITH SPL METHODOLOGIES

We now describe our proposed approach to integrate middleware specialization with SPL methodologies.

3.1 A Middleware Case Study

To make the description of our proposed approach concrete we use a middleware case study. Figure 3 illustrates the CORBA middleware architecture, which is compliant with our layered middleware system model. Also shown in the figure are CORBA services, real-time CORBA (RTCORBA) [18] enhancements, and component-based abstractions. CORBA is used here only for illustration purpose, however, our approach is general.

Figure 3: CORBA Architecture

The different RTCORBA features are shown in Figure 4. RTCORBA defines standard interfaces and QoS policies that allow applications to configure and control (1) processor resources via thread pools, priority mechanisms, intra-process mutexes, and a global scheduling service, (2) communication resources via protocol properties and explicit bindings, and (3) memory resources via buffering requests in queues and bounding the size of thread pools. Applications typically specify these real-time QoS policies along with other policies when they call standard CORBA operations, such as create_POA or validate_connection. For example, the priority at which requests must be handled can be propagated from the client to the server (the CLIENT_PROPAGATED model) or declared by the server (the SERVER_DECLARED model).

3.2 Uncovering the Algebraic Structure of Middleware

Despite a rich repertoire of features, specializations including feature additions, pruning or customizations to general-purpose middleware is a hard problem due to the following challenges posed by their design and implementation:

a. fundamental restrictions and limited flexibility of programming languages such as C++ or Java do not allow interception of the control flow at arbitrary points in the control flow graph to inject required application-specific functionality or remove certain unnecessary functionality. This is currently feasible only at limited points in the code known as interception points, which is often not sufficient.

b. although object-oriented designs help develop modular middleware code, this modularity incurs a performance penalty.

c. the combinatorial complexity of the feature compositions makes it hard to find valid configurations manually because of the large number of middleware configuration options and complex semantic relationships between them.

d. deployment- and run-time specializations are even harder because feature removal and additions need to be considered simultaneously, systematically and in a semantically consistent and coordinated manner such that domain-specified requirements on performance and footprint are satisfied.

In Section 2.2 we discussed how Aspect-oriented programming (AOP) [12] is extensively used for middleware specialization (e.g., [11, 27, 19]). AOP, however, does not support any architectural model to define transformations to the structure of programs, particularly the ability to encapsulate new classes, which limits its suitability for middleware specialization. Feature-oriented programming (FOP) [20] on the other hand can represent single aspects or collections of aspects, and also can complement model-based development since both paradigms stress the importance of transformations in automated program development [3]. Moreover, FOP has better support to provide bounded (i.e., selective) quantification for feature manipulation in contrast to AOP techniques which often result in unbounded quantification.

FOP is thus a candidate approach for middleware specializations since it involves manipulation of middleware features. FOP is best suited when the underlying construct on which it operates displays a well-defined algebraic structure. FOP for middleware specializations is not straightforward, however, due to a lack of an explicit algebraic structure in the middleware design as explained above. We therefore ask ourselves whether it is possible to impose an algebraic structure on the middleware. A closer scrutiny of the middleware design reveals that if we raise the level of abstraction [24] to the level of features the middleware offers instead of focusing on source code-level details, then a strong algebraic structure unfolds wherein features can be manipulated using the FOP paradigm subject to some constraints.

We have therefore chosen the principles of AHEAD (Algebraic Hierarchical Equations for Application Design) [4], which is an implementation of FOP that uses stepwise refinement to synthesize application product lines, as the basis of the proposed approach. The notion of a feature in AHEAD is tied to basic object-oriented programming concepts, such as classes and methods. Al-
though middleware also often uses object-oriented design principles, our notion of features is at a higher level of abstraction involving patterns and frameworks that provide properties, such as real-timeliness and fault tolerance.

AHEAD starts with a small set of base capabilities and refines them by incrementally adding features. In contrast our middleware specializations start with a much larger software base pruning unwanted features and customizing the needed ones with domain-specific properties. Our goal is to enhance AHEAD and similar research to support design, deployment and run-time feature manipulation.

3.3 Exploiting the Algebraic Structure of Middleware

We now lay down our initial ideas on our proposed approach to middleware specializations based on recursive algebraic approaches such as AHEAD, however, by operating at a level of abstraction for features that is closer to patterns and frameworks, and across all stages of the application lifecycle. Table 1 depicts our attempt to capture the algebraic structure of RTCORBA capabilities as features within an Origami matrix as proposed by the AHEAD approach [4] with the difference that our level of abstraction for features is different and we consider all stages of application lifecycle. Origami is a generalization of binary decision matrices, where matrix axes define different sets of features, and matrix entries define feature interactions. Origami matrices possess a special property in that they allow folding along the rows or columns or both. We discuss how this property will be used.

The composition operator \( \bullet \) is part of a well-defined algebra [2], which to our knowledge works only for design-time feature composition. AHEAD (and hence Origami) does not support deployment- and run-time feature manipulation. We argue for new research in enhancing existing SPL research, such as AHEAD, to include deployment- and run-time phases of application lifecycle. Applying AHEAD principles to cover all the stages of application lifecycle is hard however because the level of abstraction it operates at (e.g., code level) is not suitable for feature manipulations in the deployment- and run-time stages, and it is conceivable that the existing feature algebra will be incompatible at these stages.

3.4 Feature Manipulations across Application Lifecycle Stages

The composition operator \( \bullet \) is part of a well-defined algebra [2], which to our knowledge works only for design-time feature composition. AHEAD (and hence Origami) does not support deployment- and run-time feature manipulation. We argue for new research in enhancing existing SPL research, such as AHEAD, to include deployment- and run-time phases of application lifecycle. Applying AHEAD principles to cover all the stages of application lifecycle is hard however because the level of abstraction it operates at (e.g., code level) is not suitable for feature manipulations in the deployment- and run-time stages, and it is conceivable that the existing feature algebra will be incompatible at these stages.

We make an initial attempt to enhance this theory. Imagine a third dimension added to Table 1, which defines the deployment dimension. We can visualize this scenario as comprising multiple planes each having its own Table 1, where each table corresponds to the middleware specialization for the hosted component of the product variant. Suppose that the deployment of the product variant must ensure that the middleware is specialized for the CLIENT_PROPAGATED priority model. Now suppose that one column in the original matrix. Continuing this folding along all columns and then rows (order does not matter) gives rise to a composition of features that constitutes the overall RTCORBA middleware and can be represented by Equation 1. Features are composed with each other using the composition operator \( \bullet \).

\[
\text{RTCORBA} = \text{MUTEX} \bullet \text{TPLane} \bullet \text{CLI_PROP} \bullet \text{BandConn} \bullet \text{ExtXport} \bullet \text{TPReactor} \bullet \text{PriMapper} \bullet \text{RTPOA} \bullet \text{RTORB} \quad (1)
\]

Now let us explore how such equations will help us. Our previous work [13] on handcrafted middleware specialization has showed how the RTCORBA middleware stack characterized by Equation 1, forces the software components of our avionics mission computing scenario to use all the features, many of which are sources of excess generality. We claim that an approach to prune unwanted features can follow a similar folding operations of the Origami matrix that produces an equation of features to be pruned (e.g., bypassing the request demultiplexing logic) and customized (e.g., caching requests). This can be attempted by the application developer or middleware developers who are given the requirements by domain experts. The algebraic difference between the RTCORBA equation and the equation describing the excess generality provides a formal approach to specializing middleware.

Notice how this proposed approach is no longer ad hoc unlike handcrafted specializations. This desired property stems from the significant benefit of an Origami matrix in that it can realize only valid compositions of features. Notice that erroneous compositions (e.g. folding along the diagonal) or differences are impossible due to the constraints imposed by the folding capability of the Origami matrix. A model-based tool can provide an approach to collect all the domain requirements, which then can be used to drive the Origami folding and synthesis of the different equations.

<table>
<thead>
<tr>
<th>Base</th>
<th>RT</th>
<th>BasicRT</th>
<th>Priority</th>
<th>Conc</th>
<th>Synch</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORB</td>
<td>RTORB</td>
<td>PriMapper</td>
<td>TPReactor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POA</td>
<td>RTPOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xport</td>
<td>ExtXport</td>
<td>BandConn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReqHndl</td>
<td>CLI_PROP</td>
<td>TPLane</td>
<td>MUTEX</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Origami Matrix for RTCORBA

Table 2: Folded Matrix for RTCORBA
such table uses a SERVER_DECLARED priority model for request handling instead. We need an approach by which the folding operation along the third dimension should throw an exception due to a misconfiguration in one of the matrices. Run-time issues such as adding features for, say, a coordination layer for fault management can be handled by extending the Origami matrix in the fourth dimension to cover these run-time issues.

3.5 Feature Interactions across Application Lifecycle Stages

Our discussions so far have assumed that features are independent of each other and that they can be seamlessly added or pruned. However, we cannot make such simplified assumptions in all cases. For example, Narasimhan et. al. [16] have illustrated how real-time and fault-tolerance properties of applications conflict with each other thereby requiring tradeoffs.

Figures 5 and 6 illustrates how features can interact [14] with each other at the framework level (which is our level of abstraction for features). We show two design possibilities for an RTORB that supports thread pools with lanes. The thread pool serves as an additive refinement to the RTORB (i.e., an Introduction). However, as shown in Figures 5 and 6, the request handling strategy interacts with the RTORB in different ways each with its benefits and consequences on performance.

This model simplifies the design but incurs message queuing and thread synchronization overhead. In the reactor_per_lane approach, the thread that receives the request also handles the request, which is the Leader-Follower architectural pattern [22]. This model is difficult to implement and debug for race conditions.

We showed design-time feature interactions above, which is a hard problem since our features represent patterns and frameworks. Feature interactions at other lifecycle stages are even harder to address. Simple foldings of Origami columns and rows may not suffice since the foldings have no capabilities to tradeoff one feature over the other as in the case of fault tolerance and real-time. Traditionally the tradeoff problems have been mapped to combinatorial optimization problems where heuristics are developed to find near-optimal solutions.

4. CONCLUDING REMARKS AND OPEN ISSUES

Middleware is an important asset of SPLs that operate in a distributed computing environment. In this paper we argued for extending SPL research to incorporate middleware specializations. We showed how an algebraic structure can be imposed on the middleware which in turn makes it suitable for feature manipulation. We then explored the use of generative programming and modularization techniques based on AHEAD for middleware specialization outlining how they can be extended to address deployment- and run-time issues in middleware.

A number of open issues remain unresolved as explained below.

- **Mapping higher-level feature abstractions to code:** Since the algebraic structure we consider is at a higher level of abstraction, we require a mapping from the high level artifacts to low level details such as code. Naturally, such a mapping cannot break existing code. Hence we will need out-of-band mechanisms such as source code annotations including those we developed in our preliminary work [13] or aspect definitions to refactor existing middleware into the algebraic form we require.

- **Semantics of the composition and difference operator for deployment- and run-time phases of the application lifecycle:** Is a single equation feasible that can capture the specializations to middleware by accounting all three phases of application lifecycle. An important open issue points to the algebra of these operators across the lifecycle. A number of questions must be answered: What is the associativity and precedence relationship of the operators along the lifecycle stages? Do the semantics of Origami folding change in different lifecycle stages? How are features represented at the other lifecycle stages? How can Origami folding handle distributed coordination at run-time? Can Origami capture system schedulability and performance optimizations?

- **Runtime Tradeoffs via Origami foldings:** Adaptive systems must make runtime tradeoffs among inherently conflicting system properties such as real-timeliness and fault-tolerance. Many questions must be answered if Origami abstractions are used to solve these challenges: Do individual cells encode constraints? Do foldings give rise to cost functions? How do constraints get refined during folding? What does the final equation represent? Does the composition operator encode a heuristic to solve the optimization problem? How can feature manipulations be considered simultaneously, systematically and in a semantically consistent and coordinated manner such that domain-specified requirements on QoS and footprint are satisfied?
5. REFERENCES


