IBM Research Report

Refactoring Support for Class Library Migration

Ittai Balaban
New York University
251 Mercer Street
New York, NY 10012

Frank Tip, Robert Fuhrer
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
Refactoring Support for Class Library Migration

Ittai Balaban  
New York University  
251 Mercer St., New York, NY 10012  
balaban@cs.nyu.edu

Frank Tip  
Robert Fuhrer  
IBM T.J. Watson Research Center  
P.O. Box 704, Yorktown Heights, NY 10598  
(ftip,rfuhrer)@us.ibm.com

ABSTRACT

As object-oriented class libraries evolve, classes are occasionally deprecated in favor of others with roughly the same functionality. In Java’s standard libraries, for example, class Hashtable has been superseded by HashMap, and Iterator is now preferred over Enumeration. Migrating client applications to use the new idioms is often desirable, but making the required changes to declarations and allocation sites can be quite labor-intensive. Moreover, migration becomes complicated—and sometimes impossible—if an application interacts with external components, if a legacy class is not completely equivalent to its replacement, or if multiple interdependent classes must be migrated simultaneously. We present an approach in which mappings between legacy classes and their replacements are specified by the programmer. Then, an analysis based on type constraints determines where declarations and allocation sites can be updated. The method was implemented in Eclipse, and evaluated on a number of Java applications. On average, our tool could migrate more than 90% of the references to legacy classes.

1. INTRODUCTION

Class libraries evolve because it is difficult to anticipate what functionality is needed, and new usage patterns may arise after a library’s initial deployment. The incorporation of additional functionality in a class library can often be accomplished in backward-compatible ways by creating additional interfaces, and/or adding methods to existing classes. However, when an existing library class provides insufficient flexibility because of an unfortunate design decision, it is often undesirable to make incompatible changes that would break existing library clients. As a result, classes are occasionally deprecated in favor of others with roughly the same functionality. In such cases, legacy classes are typically not removed from a library for backward compatibility reasons.

Several examples of the scenario described above can be found in the java.util collection libraries. Class Hashtable has been superseded by a class HashMap that is similar but allows unsynchronized\(^1\) access to a hash-table’s elements as well as the use of null keys and values. Class ArrayList is similar to class Vector except for the fact that it allows unsynchronized access to the list’s elements. Furthermore, the Iterator interface, now favored over the similar Enumeration interface, uses more concise method names and provides an additional operation remove(). Outside of the standard collection classes, class java.io.PrintWriter is now preferred instead of the similar class java.io.PrintStream [5]. Moreover, Java 1.5 supports a class java.io.StringBuffer that provides the same functionality as java.io.StringBuffer but is not thread-safe. Migrating a client application to the new idioms is often desirable, but can be labor-intensive as it may involve updating many declarations, allocation sites, and call sites. Moreover, migration becomes complicated—and sometimes impossible—if an application interacts with external components, if a legacy class is not completely equivalent to its replacement, or if multiple interdependent classes must be migrated simultaneously.

The contribution of this paper is a technique for automatic migration of applications that use legacy library classes. In this approach, a migration specification defines how uses of legacy library classes are mapped to uses of their replacement classes. Each such mapping needs to be defined only once, and can then be used to migrate any number of applications. We use an analysis based on type constraints [25] to determine where, for a given migration specification, it is possible to migrate uses of legacy classes without affecting the program’s type-correctness or behavior. In addition, an escape analysis (see e.g., [6, 1, 27]) serves to determine where it is safe to replace a synchronized legacy class with an unsynchronized replacement class\(^2\). In cases where it cannot be shown that synchronization can be removed safely for a given object o, our approach is to insert a synchronization wrapper around o. This is accomplished using an instance of the DECORATOR design pattern [15] where a collection type is implemented by delegating all operations to o, and where thread-safety is added by making all the forwarding methods synchronized. The synchronization wrap-

\(^1\)In new Java container classes, synchronization is decoupled from a collection’s functionality. Class java.util.Collections contains “synchronization wrapper” methods for transforming any Map or List into a synchronized Map or List, respectively.

\(^2\)For example, consider the migration from the synchronized legacy class Hashtable to its unsynchronized replacement HashMap.
Our work was implemented as a refactoring [13, 22] in Eclipse (see www.eclipse.org), a widely used open-source development environment for Java, using an existing refactoring infrastructure [2]. Thus far, our experiments have concentrated on migrations involving classes from the Java Standard Collections Framework, but our techniques can also be used for migrations involving other library classes, and for migrations between user-defined classes. At present, the framework requires that the migration of method calls does not change the exceptions that may be thrown, and migrations are simply disallowed if this is not the case. This has not posed any problems in practice, because we did not encounter situations that required nontrivial migrations of exception types while conducting our experiments. Section 2 presents a simple workaround mechanism that is sufficient to handle most cases of migrations between methods that differ in the thrown exceptions. In the long term, it does not seem especially difficult to support the migration of exception types, as is discussed in Section 7. Furthermore, we expect that handling other statically typed object-oriented languages (e.g., C#) would only require minor modifications of the framework.

We evaluated our technique on a number of moderately-sized Java applications in which we migrated a number of heavily used legacy classes from the standard collection libraries. Specifically, we considered migration from Vector to ArrayList, HashTable to HashMap, and Enumeration to Iterator. In these benchmarks, our tool could automatically migrate an average of 90.4% of the declarations that refer to legacy classes, 96.6% of the call sites, and 92.0% of the reference variables.

Figure 1: (a) Example program that uses the legacy classes Vector, HashTable, and Enumeration. (b) The example program after migrating from Vector to ArrayList, HashTable to HashMap, and Enumeration to Iterator. Underlining indicateschanged code fragments. A1, A2, and A3 denote allocation sites in the program.

```
(1) public class Wizard {
(2)  private static JComboBox selectionBox;
(3)  private static Hashtable h1 = new Hashtable(); /* A1 */
(4)  private static Vector v2;
(5)  private static void main(String[] args) {
(6)    for (int count = 0; count < 5; count++)
(7)      h1.put(new Integer(count),
(8)          "item " + Integer.toString(count));
(9)    Vector v1 = new Vector(); /* A2 */
(10)   retrieveData(v1);
(11)   v2 = runWizard(v1);
(12)   Object[] dbData = new Object[v2.size()];
(13)   v2.copyInto(dbData);
(14)   writeToDatabase(dbData);
(15) }
(16) private static Vector runWizard(Vector v3) { /* A3 */
(17)   Vector v4 = new Vector();
(18)   for (int sel = showWizardDialog(v3);
(19)     sel >= 0;)
(20)     v4.addElement(v3.elementAt(sel));
(21)   } sel = showWizardDialog(v3);
(22) }
(23) System.out.println("User Selections:");
(24) printElements(v4.elements());
(25) return v4;
(26) }
(27) private static void printElements(Enumeration e1) {
(28)   while (e1.hasMoreElements())
(29)     System.out.println(e1.nextElement());
(30) }
(31) private static void retrieveData(Vector v6) {
(32)   printElements(h1.elements());
(33)   for (Enumeration e2 = h1.elements();
(34)     e2.hasMoreElements();
(35)     v6.addElement(e2.nextElement()));
(36) }
(37) private static int showWizardDialog(Vector v6) {
(38)   selectionBox = new JComboBox(v6);
(39)   selectionBox.
(40)   addActionListener(new MyActionListener(v2));
(41)   }
(42) static void writeToDatabase(Object[] data) { /* ... */
(43) }
(44) public class MyActionListener implements ActionListener {
(45)  private static Vector v7;
(46)  public MyActionListener(Vector v6) { v7 = v6;}
(47)  }
(48) }
```

```
(1) public class Wizard {
(2)  private static JComboBox selectionBox;
(3)  private static HashMap h1 = new HashMap();
(4)  private static List v2;
(5)  private static void main(String[] args) {
(6)    for (int count = 0; count < 5; count++)
(7)      h1.put(new Integer(count),
(8)          "item " + Integer.toString(count));
(9)    Vector v1 = new Vector();
(10)   retrieveData(v1);
(11)   v2 = runWizard(v1);
(12)   Object[] dbData = new Object[v2.size()];
(13)   v2.copyInto(dbData);
(14)   writeToDatabase(dbData);
(15) }
(16) private static List runWizard(Vector v3) { /* ... */
(17)   List v4 = Collections.synchronizedList(new ArrayList());
(18)   for (int sel = showWizardDialog(v3);
(19)     sel >= 0;)
(20)     v4.add(v3.elementAt(sel));
(21)   } sel = showWizardDialog(v3);
(22) }
(23) System.out.println("User Selections:");
(24) printElements(v4.iterator());
(25) return v4;
(26) }
(27) private static void printElements(Iterator e1) {
(28)   while (e1.hasNext())
(29)     System.out.println(e1.next());
(30) }
(31) private static void retrieveData(Vector v6) {
(32)   printElements(h1.values());
(33)   for (Iterator e2 = h1.values().iterator();
(34)     e2.hasNext();
(35)     v6.addAll(e2.next());
(36) }
(37) private static int showWizardDialog(Vector v6) {
(38)   selectionBox = new JComboBox(v6);
(39)   selectionBox.
(40)   addActionListener(new MyActionListener(v2));
(41)   }
(42) static void writeToDatabase(Object[] data) { /* ... */
(43) }
(44) public class MyActionListener implements ActionListener {
(45)  private static List v7;
(46)  public MyActionListener(List v6) { v7 = v6;}
(47)  }
```

```
(a)

(b)

pers that we use are provided by the standard library class java.util.Collections exactly for this purpose.

...
Java programs may incur significant performance overhead via a call to the method Swing.

Figure 1(a) shows an example program that implements a simple "wizard" using a number of components from the Swing class library. The program is implemented in a drop-down menu, and prompts the user to select one or more of these. Then, the selected values are written to a database via a call to the method `writeToDatabase(Object[]).` The definition of which is omitted. Our goal is to transform this program so that it uses `ArrayList` instead of the legacy class `Vector`, `HashMap` instead of `Hashtable`, and `Iterator` instead of `Enumeration`. Figure 1(b) shows the desired refactored program.

The remainder of this paper is organized as follows. Section 2 presents a motivating example that illustrates our approach. Section 3 presents the model of type constraints that will be used to determine where migrations are allowed, and Section 4 describes how type constraints are solved. Section 5 discusses the evaluation of our approach on a set of moderate-sized benchmarks. Related work is discussed in Section 6. Finally, conclusions and directions for future work are presented in Section 7.

2. MOTIVATING EXAMPLE

Figure 1(a) shows an example program that implements a simple "wizard" using a number of components from the Swing class library. The program displays a list of values in a drop-down menu, and prompts the user to select one or more of these. Then, the selected values are written to a database via a call to the method `writeToDatabase(Object[]),` the definition of which is omitted. Our goal is to transform this program so that it uses `ArrayList` instead of the legacy class `Vector`, `HashMap` instead of `Hashtable`, and `Iterator` instead of `Enumeration`. Figure 1(b) shows the desired refactored program.

`ArrayList` was introduced in the standard libraries to replace `Vector`, and is considered preferable because its interface is minimal and matches the functionality of the `List` interface. More importantly, it allows for unsynchronized access to a list's elements whereas all of `Vector`'s methods are `synchronized` which results in unnecessary overhead when `Vectors` are used by only one thread. Similarly, `HashMap` was introduced to replace `Hashtable` and allows for unsynchronized access to a hash-table's elements. Moreover, `HashMap` allows for null values and it does not contain any of `Hashtable`'s methods that produce `NullPointerExceptions.` The `Iterator` interface replaces `Enumeration` and features shorter method names and supports an additional `remove()` method for the safe removal of elements during iteration.

Figure 2 shows a migration specification, as would be written by a user, for migrating the example program. This specification consists of a set of rewrite rules that express how allocation sites and method calls on legacy classes are replaced. For example, rule (3) states that migrating calls to method `Vector.size()` requires no modification, since `ArrayList` defines a syntactically and semantically identical method. If methods in the legacy class are not supported by the replacement class, rewriting method calls becomes more involved. For example, `Vector` supports a method `firstElement()` not defined in `ArrayList`. Rule (4) states that a call `receiver.firstElement()`, where `receiver` is an expression of type `Vector`, should be transformed into `receiver.get(0)`. In cases where it is not possible to express the effect of a method call on the legacy class in terms of calls to methods on the replacement class, calls may be mapped to methods in user-defined classes. For example, `Vector` has a method `copyInto()` that copies the contents of a `Vector` into an array. Since `ArrayList` does not provide this functionality, rule (6) transforms `receiver.copyInto(array)` into a call to method `Util.copyInto(receiver, array)` of the `Util` class shown in Figure 3.

Rules (1) and (2) are both concerned with rewriting allocation sites of the form `new Vector()`. The former applies in cases where thread-safety need not be preserved, and transforms the allocation site into an expression `new ArrayList()`. The latter applies in situations where thread-safety must be preserved, and transforms

3As noted in e.g., [6, 1, 27], synchronization operations in Java programs may incur significant performance overhead.

4In this discussion we use the term `class` to refer to Java classes as well as to interfaces.
the allocation site into `Collections.synchronizedList(new ArrayList())` using a standard synchronization wrapper in the class `java.util.Collections`. Our tool relies on an escape analysis to determine which of the two rules should be applied, and prefers (1) over (2) whenever possible.

It is important to realize that, while the specification of Figure 2 describes how program fragments are transformed, it does not state when the transformation is allowed. We will now examine the program of Figure 1 to illustrate the limitations on migration.

Line (38) of the example program contains a call to a constructor `javax.swing.JComboBox(Vector)` of `Vector` (as opposed to `abstract` container classes such as `Map`) in the signature of public methods is often an indication of poor design, because it unnecessarily exposes implementation details. Nonetheless, this practice is pervasive. For example, we counted 215 public methods in the JDK 1.4.2 standard libraries whose signature refers to `Hashtable` or `Vector`.

in turn, entails migrating `elements.hasMoreElements()` on line (28) and `elements.nextElement()` on line (29), and all callers of `printElements()` must be updated accordingly, which involves further changes on lines (3) and (32)–(35).

Figure 1(b) shows the refactored program, using underlining to highlight transformed program fragments. It is evident from the above discussion that migrating references of legacy classes is both labor-intensive (many declarations, calls, and allocation sites are affected) and error-prone (not all uses of legacy classes can be migrated, and dependencies between migrations exist). The limitations on class library migration can be categorized as follows:

- **interaction with external libraries.** If an application calls a method of an external library, opportunities for migration may be limited if the method signature refers to a concrete class such as `Vector`.
- **type correctness requirements due to assignments.** Program constructs such as assignments require that certain subtype-relationships hold between the types of variables and expressions. As a result, migrating one variable often requires migrating others.
- **dependencies between migrations.** If a method in one legacy class refers to another, one migration may depend on another.
- **thread safety.** A major difference between legacy collection classes and their replacements is the fact that the latter do not perform automatic synchronization on every method call. If thread-safety is required, synchronization wrappers must be inserted.

### 3. THE FRAMEWORK

We now present the formal framework for expressing migrations between classes. In what follows, we will use `P` and `P'` to refer to the original program, and to the refactored program, respectively. We will first make precise the concept of a migration between classes. Then, we present the type constraints model.

#### 3.1 Migrations

Fix two classes, `T` and `T'`. We say that a mapping `m` over fragments of Abstract Syntax Trees (AST) is a *migration* from `T` to `T'` under the following conditions:

- `m` maps allocation expressions of type `T`, i.e. expressions of the form `new T(...)`, into allocation expressions of type `T'`.
- `m` maps method call expressions on methods of `T` to expressions not written in terms of `T`.

Given a migration from `T` to `T'`, we refer to `T` and to `T'` as a *legacy class*, and a *replacement class*, respectively. An expression `E` is said to be a *migration candidate* if it is either an allocation expression of type `T`, or if it is a call to a method of `T`.

The input to the framework, in addition to the program being refactored, is a set of migrations, each of which is specified by a set of rules. The example in Figure 2 consists of three migrations: From `Vector` to `ArrayList`, from
Given a set of migrations \( M \) and a program \( P \), a system of type constraints \([25]\) is generated. Type constraints are a formalism for expressing relationships between the types of declarations and expressions that has traditionally been used for type-checking, type inference, and more recently for refactoring \([31, 30, 14]\). We extend the type constraints framework of \([31]\) to determine where declarations and allocations sites can be transformed, as per migrations in \( M \), without affecting program behavior. Type constraints are generated from the AST of a program in a syntax-directed manner, and encode relationships between the types of declarations and expressions that must be satisfied in order to preserve type correctness or program behavior. Figure 4 shows the notation used to formulate type constraints. Figure 5 shows the syntax of type constraints. These constraints are generated by rules (25)–(29) in Figure 6, and are discussed in Section 3.2.2.

The use of a partial specification language has the disadvantage that the framework may be able to migrate smaller portions of a program than with a more complete one. The grammar of the specification language, as well as an example of a complete migration from \([31]\) to determine where declarations and allocations sites can be transformed, as per migrations in \( M \), without affecting program behavior. Type constraints are generated from the AST of a program in a syntax-directed manner, and encode relationships between the types of declarations and expressions that must be satisfied in order to preserve type correctness or program behavior. Figure 4 shows the notation used to formulate type constraints. Figure 5 shows the syntax of type constraints. These constraints are generated by rules (25)–(29) in Figure 6, and are discussed in Section 3.2.2.

1. Constraints that express relationships between types of declarations and expressions that must be preserved in order to preserve type-correctness. These constraints are generated by rules (1)–(24) in Figure 6, and are largely the same as in \([30, 31]\), the main difference being special treatment for legacy classes. Section 3.2.1 presents some of these constraints in more detail.

2. Constraints that express relationships between the types of declarations and expressions that must be preserved in order to preserve a program’s run-time behavior. These constraints are generated by rules (25)–(29) in Figure 6, and are discussed in Section 3.2.2.

3. Constraints that preserve type-correctness of migration candidates. These are shown in Figure 7, and will be discussed in Section 3.2.4.

2.1 Preserving Type Correctness

Rules (1)–(24) in Figure 6 generate type constraints for preserving type-correctness. Each rule generates, for a given program construct, one or more type constraints that express the subtype-relationships that must exist between the declared types of the construct’s constituent expressions, in order for that program construct to be type-correct. By definition, a program is type-correct if the type constraints for all its constructs are satisfied. We will now study a few of the constraint generation rules in detail, referring the reader to \([30, 31]\) for further details.

For example, rule (1) states that an assignment \( E_1 = E_2 \) is type correct if the type of \( E_2 \) is the same as or a subtype of the type of \( E_1 \). For a field-access expression \( E.f \), rule (2) defines the type of this expression to be the same as the declared type of \( F \) and rule (3) requires that the type of expression \( E \) be a subtype of the type in which \( F \) is declared.

Rules (4)–(6) are concerned with a virtual method call \( E.m( E_1, \ldots, E_k) \) that refers to a method \( M \). Rule (4) defines the type of the call-expression to be the same as \( M ’s return type. Further, the type of each actual parameter \( E_i \) must be the same as or a subtype of the type of the corresponding formal parameter \( Param(M, i) \) (rule (5)). Rule (6) states that
### Figure 6: Type constraints for a set of core Java language features.

<table>
<thead>
<tr>
<th>program construct(s)/analysis fact(s)</th>
<th>implied type constraint(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>assignment $E_1 = E_2$</td>
<td>$E_2 \leq E_1$</td>
</tr>
<tr>
<td>access $E.f$ to field $F$</td>
<td>$E.f = F$</td>
</tr>
<tr>
<td>method call</td>
<td>$E.m(E_1, \ldots, E_n) = M$</td>
</tr>
<tr>
<td>to a virtual method $M$</td>
<td>$E \leq \text{Decl}(M)$</td>
</tr>
<tr>
<td>$[E]_p$ is not a migration candidate</td>
<td></td>
</tr>
<tr>
<td>return $E$ in method $M$</td>
<td>$[E] \leq [M]$</td>
</tr>
<tr>
<td>Constructor call $\text{new } C(E_1, \ldots, E_n)$ to constructor $M$</td>
<td>$[E_i] \leq \text{Param}(M, i)$</td>
</tr>
<tr>
<td>Class $C$ is not a legacy class</td>
<td></td>
</tr>
<tr>
<td>direct call</td>
<td>$E.m(E_1, \ldots, E_n) = M$</td>
</tr>
<tr>
<td>to method $M$</td>
<td>$E \leq \text{Decl}(M)$</td>
</tr>
<tr>
<td>down-cast expression $(C)E$</td>
<td>$[(C)E] \leq [E]$</td>
</tr>
<tr>
<td>for every type $T$</td>
<td>$T \leq \text{java.lang.object}$</td>
</tr>
<tr>
<td>implicit declaration of $\text{this}$ in method $M$</td>
<td>$\text{this} = \text{Decl}(M)$</td>
</tr>
<tr>
<td>declaration of method $M$ (declared in type $T$)</td>
<td>$\text{Decl}(M) = T$</td>
</tr>
<tr>
<td>declaration of field $F$ (declared in type $T$)</td>
<td>$\text{Decl}(F) = T$</td>
</tr>
<tr>
<td>explicit declaration of $v$ th parameter $v$ of method or constructor $M$</td>
<td>$\text{Param}(M, i) = [v]$</td>
</tr>
<tr>
<td>down-cast expression $(T)E$</td>
<td>$[(T)E] \leq T'$</td>
</tr>
<tr>
<td>$T$ is a legacy type</td>
<td>$T \leq [(T)E]$</td>
</tr>
<tr>
<td>down-cast expression $(T)E$</td>
<td>$[(T)E] = T$</td>
</tr>
<tr>
<td>$T$ is not legacy type</td>
<td></td>
</tr>
<tr>
<td>expression $\text{new } C(E_1, \ldots, E_n)$</td>
<td>$[\text{new } C(E_1, \ldots, E_n)] \leq C'$</td>
</tr>
<tr>
<td>$C$ is a legacy class with replacement $C'$</td>
<td>$C_1 \leq C'$</td>
</tr>
<tr>
<td>expression $\text{new } C(E_1, \ldots, E_n)$</td>
<td>$[\text{new } C(E_1, \ldots, E_n)] = C'_{\text{new}}$</td>
</tr>
<tr>
<td>$C$ is not a legacy class</td>
<td></td>
</tr>
<tr>
<td>$M'$ overrides $M$, $M' \neq M$</td>
<td>$[\text{Param}(M', i)] = [\text{Param}(M, i)]$</td>
</tr>
<tr>
<td>actual parameter $E$ in call to method in external library, $[E]_p = T$</td>
<td>$[E] = T$</td>
</tr>
<tr>
<td>for each down-cast expression $(C)E$, and each allocation expression $E' \in \text{PointsTo}(P, E)$ such that $[E']_p \leq [(C)E]_p$</td>
<td>$[E'] \leq [(C)E]$</td>
</tr>
<tr>
<td>for each down-cast expression $(C)E$, and each allocation expression $E' \in \text{PointsTo}(P, E)$ such that $[E']_p \not\leq [(C)E]_p$</td>
<td>$[E'] \not\leq [(C)E]$</td>
</tr>
</tbody>
</table>

### Figure 7: Type constraints for migration candidates.

<table>
<thead>
<tr>
<th>program construct(s)/analysis fact(s)</th>
<th>implied type constraint(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression $E$ of the form $\text{new } C(\cdots)$</td>
<td>$(E] = C) \rightarrow c$</td>
</tr>
<tr>
<td>$[E]_p = C$</td>
<td>$(E] = C') \rightarrow c'$</td>
</tr>
<tr>
<td>$C$ is a legacy class with replacement $C'$</td>
<td>for every $c \in \text{consts}(E), c' \in \text{consts}(E')$</td>
</tr>
<tr>
<td>$E'$ is a migration of $E$</td>
<td></td>
</tr>
<tr>
<td>Expression $E$ of the form $E_1.m(\cdots)$</td>
<td>$(E] = T) \rightarrow c$</td>
</tr>
<tr>
<td>$[E_1]_p = T$</td>
<td>$(E] = T') \rightarrow c'$</td>
</tr>
<tr>
<td>$T$ is a legacy type with replacement $T'$</td>
<td>for every $c \in \text{consts}(E), c' \in \text{consts}(E')$</td>
</tr>
<tr>
<td>$E'$ is a migration of $E$</td>
<td></td>
</tr>
</tbody>
</table>
a declaration of a method with the same signature as $M$ must occur in a supertype of the type of $E$. The complexity in this rule stems from the fact that $M$ may override one or more methods $M_1, \ldots, M_k$ that are declared in supertypes $T_1, \ldots, T_k$ of the type $\text{Decl}(M)$. Here, the type-correctness of the method call only requires that the type of receiver expression $E$ is a subtype of one of these $T_i$. This is expressed by way of a disjunction in rule (6) using auxiliary function $\text{RootDefs}$ of Figure 4.

Rules (22) and (23) are concerned with allocation expressions of legacy classes. The need for rules specific to legacy classes is illustrated as follows: In \cite{31}, the following conclusions of legacy classes. The rules for down-cast expressions have been solved; the refactored program will not refer to these auxiliary classes. First, rule (27) states that the types of actual parameters in calls to methods in external libraries cannot be changed. Second, since source code is unavailable, we must make conservative assumptions about casts that may be performed inside external libraries, and also for computing points-to information. Rule (27) satisfies these assumptions by effectively fixing the type of any allocation site that may flow to an external library.

### 3.2.3 Example

Figure 8 shows some of the basic constraints generated for the example of Figure 1 using the rules of Figure 6. This set of constraints, although not complete, is sufficient to demonstrate the migration possibilities and limitations of allocation sites $A_2$ and $A_3$. Note that $A_3$ is only constrained to be a subtype of $\text{Vector}^\top$, and a supertype of $\text{Vector}_\bot$, and can therefore be migrated. However, $A_2$ cannot be migrated due to a call to a constructor of the external class JComboBox. Rule (27) fixes the parameter of this constructor to be Vector, and the other constraints effectively propagate this fact to $A_2$, whose type accordingly must remain Vector.

### 3.2.4 Implication Constraints

Since a migration is a transformation over expressions, a migration candidate and its transformed counterpart often generate distinct sets of constraints. To illustrate this, consider the migration in Figure 2 and the expression $\text{v2.copyInto}([\text{dbData}])$ on line (13) of Figure 1. The transformed method call is $\text{Util.copyInto}([\text{v2}, \text{dbData}])$. Here, the original expression generates (by rules (5) and (6)) the constraints

$$[\text{v2}] \leq \text{Vector} \quad \text{[dbData]} \leq [\text{Param}(\text{Vector.copyInto}, 1)]$$

because $\text{v2}$ is a receiver of the method call, and $\text{dbData}$ is an actual parameter. However, in the transformed expression $\text{v2}$ and $\text{dbData}$ are both actual parameters. Hence the transformed expression generates (by rule (10)) the constraints

$$[\text{v2}] \leq [\text{Param}(\text{Util.copyInto}, 1)] \quad \text{[dbData]} \leq [\text{Param}(\text{Util.copyInto}, 2)]$$

Therefore, depending on the type assigned to certain constraint variables (in this case the variable $[\text{v2}]$), different sets of constraints need to be considered by a solving algorithm. To express this, we use implication constraints, rules for which are given in Figure 7.
4. SOLVING CONSTRAINTS

We will present the algorithm for solving systems of type constraints in two parts. First, an algorithm for solving basic constraints is presented. This is then embedded in a solver that accommodates implications. Our goal is to find a solution to the constraints that maximizes the number of migration candidates that are transformed to use replacement classes.

4.1 Solving Simple Constraints

An initial step in solving a system of simple constraints is to simplify all disjunction constraints into equivalent or stronger constraints over the relation $\leq$. Disjunctions are generated by rule (6) for dealing with virtual method calls. We first remove any disjunction that trivially follows from some other constraint in the system. Then, any remaining disjunction is strengthened to a $\leq$-constraint by replacing it with one of its disjuncts. If we assume that none of the disjuncts is derived from a migrated call site, it is guaranteed that a solution to the constraint system will still exist after strengthening the disjunctions, because the original program satisfies all disjuncts of all disjunctions. The manner in which a disjunct is chosen to replace the constraint is the same as in [30].

To solve a system of simple constraints over relations $\leq$ and $\not\leq$, we partition the set of constraint variables into equivalence classes according to equality constraints. That is, any solution to the system must assign an identical type to all members of a single equivalence class. The relations $\leq$ and $\not\leq$ are extended to equivalence classes in the obvious way. Then the set of possible types for each class is computed using an optimistic algorithm. The algorithm associates with each equivalence class $E$ a set $S_E$ of types, initialized as follows: (i) if $E$ contains an allocation expression new $C(\cdots)$, then $S_E = \{C\} \cup \{C' \mid C$ is a legacy class with replacement $C'\}$; otherwise (ii) $S_E$ contains all types except $T_1$ and $T_{\infty}$, for all $T$.

The algorithm SOLVE-SIMPLE in Figure 9 solves a system of simple constraints by iterating until a fixpoint is reached, in which no set $S_E$ contains more than one type. The algorithm consists of a pruning phase (lines 2-3), where invalid types are pruned from sets, and a heuristic phase (lines 9-10) where a non-singleton set is reduced to a single member. The heuristic phase is defined in terms of some heuristic choose that, given a set of classes, selects a most preferred class. In our implementation it greedily prefers replacement classes over legacy classes, though it does not consistently prefer any one replacement class over another. Space limitations prevent us from presenting a proof that SOLVE-SIMPLE always computes a solution, and that the heuristic phase does not eliminate all solutions from the search space. Termination follows from the fact that every loop body performs a strictly monotone step, reducing the size of at least one set.

4.2 Solving Implication Constraints

We now consider a general constraint system consisting of simple implications as well as simple constraints. Let $I$ be a set of implication constraints of the form $(\langle I \rangle = T) \rightarrow \psi$ and $C$ a set of simple constraints. The algorithm SOLVE-COMPLETE for solving $(I, C)$ is given in Figure 10. Its basic step is to select an implication constraint $(\langle I \rangle = T) \rightarrow \psi$, and to nondeterministically satisfy or violate its left side by removing types from the set $S_E$. Subsequently, the simple solver is applied with a possibly larger set of simple constraints. The requirement on the nondeterministic choice, expressed in line (8), is that it does not eliminate all solutions. If such a state is reached then the algorithm backtracks to the last nondeterministic choice point at which not all choices have been exhausted, and makes a different choice. Such a point is guaranteed to exist, since the original program is a valid solution to the constraint system. In practice the backtracking mechanism is coupled with a policy for directing nondeterministic choice. For example, given an implication $(\langle I \rangle = T) \rightarrow \psi$, the implemented policy is to attempt to satisfy $(\langle I \rangle = T)$ if $T$ is a replacement class.

---

Figure 9: Algorithm SOLVE-SIMPLE for a set $\mathcal{C}$ of simple constraints over $\leq$, $\not\leq$.

Figure 10: Algorithm SOLVE-COMPLETE for the sets $(I, C)$ of implication and simple constraints. The notation $V := A$ or $V := E$ denotes a nondeterministic assignment.
The running time of Solve-Complete is worst-case exponential in the number of call sites. However, such behavior is rare, since implication constraints are rarely independent of one another (i.e., multiple implications tend to have identical left hand sides). More importantly, due to the choice-directing policy, worst case behavior is only exhibited in programs where a majority of call sites cannot be migrated. In programs encountered in our evaluation this has not been the case.

4.3 Preserving Thread-Safety

Suppose that we migrate from a class T to a class T', where T is designed to be thread-safe by only allowing synchronized access to its internal state, but where T' is not. Unless countermeasures are taken, the refactored program may exhibit different behavior if it uses multiple threads whose executions are interleaved in ways that could not arise in the original program. The migration from Vector to ArrayList is an example of this case, because all of Vector’s public methods are synchronized, thus preventing multiple threads from concurrently accessing the same object, but none of the methods in class ArrayList are. The JDK library designers deliberately omitted synchronization from class ArrayList so as to avoid unnecessary overhead in clients that do not access a list concurrently. To handle cases where thread-safety is required, the Java collection classes provide a mechanism known as synchronization wrappers: an object of type ArrayList can be made thread-safe by “wrapping” its allocation site with a call to the static method Collections.synchronizedList(List).

Hence, instead of writing:

one would write:

Collections.synchronizedList(new ArrayList())

We support migrations from classes that are thread-safe to classes that are not by introducing synchronization wrappers. However, synchronization wrappers are only introduced for objects that may be accessed by multiple threads. To this end, we employ a simple escape analysis (see, e.g., [6, 1, 27]) to compute a conservative approximation of legacy allocation sites that may escape their allocating thread. If the analysis indicates that a legacy allocation site may escape, a synchronization wrapper is introduced. Otherwise, the allocation site is left unwrapped.

Figure 11 shows the algorithm used to determine allocation sites that may escape their allocating thread. It uses the notations SF(P) and fields(S) to denote, respectively, the set of all static field declarations in program P, and the set of all fields declared by types in the set S. It uses the notation PointsTo(P, E), discussed in Section 3.2.2, to denote the set

\[ 1 : S := \{ \text{class } C \mid C \leq \text{Thread} \} \cup \{ \text{class } C \mid C \leq \text{Runnable} \} \cup \bigcup_{F \in SF(P)} \{ \{A\} \mid A \in \text{PointsTo}(P, F) \} \]

\[ 2 : \text{repeat until fixpoint} \]

\[ 3 : \text{for each } F \in \text{fields}(S) \]

\[ 4 : \quad S := S \cup \{ \{A\} \mid A \in \text{PointsTo}(P, F) \} \]

\[ 5 : \text{for each allocation site } A \]

\[ 6 : \text{escaping}(A) := \text{Lib-Escaping}(A) \lor \exists F \in \text{fields}(S) . A \in \text{PointsTo}(P, F) \]

Figure 11: Escape Analysis Algorithm. Lib-Escaping(A) is informally defined as true if A \in PointsTo(P, E), where E is an actual parameter in a call to a method in an external library.

The rationale behind this algorithm is that static fields, subclasses of Thread, and implementors of the interface Runnable, which is often used to pass callback procedures to Thread objects, are the sole vehicles by which objects can escape their allocating thread. Therefore, any escaping object must be transitive referenced by one or more fields of these classes. Furthermore, since external libraries cannot partake in this analysis, the algorithm conservatively assumes that any allocation site flowing into a method of an external library class is thread-escaping. As we shall see in Section 5, this naïve algorithm has proved itself remarkably effective on the benchmark applications that we have analyzed.

5. EVALUATION

Our work is implemented in Eclipse, a widely-used open-source development environment for Java, using existing infrastructure for building refactorings [2] and type constraints [31, 14]. Class Hierarchy Analysis [7] is used to compute a call graph, followed by a variation on 0-CFA [29] to compute points-to information. Care is taken to ensure that correct points-to information is computed in the presence of reflection and external class libraries (similar to [32]).

We evaluated the refactoring tool on a number of Java applications of up to 53 KLOC that we migrated from Vector to ArrayList, from Hashtable to HashMap, and from Enumeration to Iterator. Table 1 states the essential characteristics for each benchmark program. From left to right, columns of the table indicate: (i) the benchmark name, (ii) the number of lines of source code (in thousands), (iii) the number of classes, (iv) the legacy classes used (here, V, HT, and E indicate Vector, Hashtable, and Enumeration, respectively), (v) the number of declarations of variables, parameters, and fields that refer to legacy classes, (vi) the number of allocation sites referring to legacy classes, and (vii) the number of call sites referring to methods in legacy classes.

The results are shown in Table 2. The columns of the table show, for each benchmark: (i) the number of legacy declarations that were migrated and those that could not be migrated for reasons discussed in Section 2 (these numbers are separated by a ‘/’ symbol), (ii) the number of legacy allocation sites that were migrated without synchronization.
that refer to these legacy classes, (vi) the number of allocation sites that refer to these legacy classes, and
that, on average, 90.4% of legacy declarations and 96.6% of
legacy call sites that were successful, our implementation does not yet make use of the new
optimization infrastructure in Eclipse.

Figure 12 shows a chart that visualizes the percentage of
optimized type constraint infrastructure in Eclipse.

remaining 36 call sites could not be migrated. Processing
692 of the 728 legacy call sites could be migrated, and the
2 legacy allocation sites could not be migrated at all, and (c)
out inserting synchronization wrappers, and the remaining
44 of the 46 legacy allocation sites could be migrated with-
out wrappers, the number of legacy allocation sites that were
migrated with synchronization wrappers, and the number of
legacy allocation sites that could not be migrated (using
'/' symbols to separate these numbers), and (iii) the number
of legacy call sites that were migrated, and the number of
legacy call sites that could not be migrated (again, using '/'
for separation), and (vii) the number of call sites that could
cannot be migrated, and (iii) the number of call sites that could be migrated and
and that could not be migrated. (iv) the time required to process the benchmark (minutes:seconds).

wrappers, the number of legacy allocation sites that were
migrated with synchronization wrappers, and the number of
legacy allocation sites that could not be migrated (using
'/' symbols to separate these numbers), and (iii) the number
of legacy call sites that were migrated, and the number of
legacy call sites that could not be migrated (again, using '/'
for separation), and (vii) the number of call sites that could
cannot be migrated, and (iii) the number of call sites that could be migrated and
and that could not be migrated. (iv) the time required to process the benchmark (minutes:seconds).

Table 1: Characteristics of the Java applications used to evaluate our techniques. From left to right, the columns of the table show: (i) the name of the benchmark, (ii) the number of lines of source code (in thousands), (iii) the number of classes in the benchmark, (iv) the legacy classes used in this benchmark (here, V denotes Vector, HT denotes Hashtable, and E denotes Enumeration), (v) the number of declarations that refer to these legacy classes, (vi) the number of allocation sites that refer to these legacy classes, and (vii) the number of call sites that refer to these legacy classes.

<table>
<thead>
<tr>
<th>benchmark</th>
<th>KLOC</th>
<th>#classes</th>
<th>legacy classes</th>
<th># legacy declarations</th>
<th># legacy allocation sites</th>
<th># legacy call sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanoi</td>
<td>3.0</td>
<td>21</td>
<td>V</td>
<td>3</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>JUnit</td>
<td>5.3</td>
<td>100</td>
<td>V, HT, E</td>
<td>62</td>
<td>24</td>
<td>118</td>
</tr>
<tr>
<td>JLex</td>
<td>7.9</td>
<td>26</td>
<td>V, HT, E</td>
<td>39</td>
<td>16</td>
<td>185</td>
</tr>
<tr>
<td>JavaCap</td>
<td>10.6</td>
<td>36</td>
<td>HT, E</td>
<td>56</td>
<td>14</td>
<td>153</td>
</tr>
<tr>
<td>Cassowary</td>
<td>12.2</td>
<td>68</td>
<td>V, HT, E</td>
<td>139</td>
<td>46</td>
<td>728</td>
</tr>
<tr>
<td>Azurues</td>
<td>13.9</td>
<td>100</td>
<td>V</td>
<td>14</td>
<td>12</td>
<td>51</td>
</tr>
<tr>
<td>HTMLParser</td>
<td>17.1</td>
<td>115</td>
<td>V, HT</td>
<td>134</td>
<td>25</td>
<td>467</td>
</tr>
<tr>
<td>JBidWatcher</td>
<td>22.9</td>
<td>154</td>
<td>V, HT</td>
<td>14</td>
<td>36</td>
<td>294</td>
</tr>
<tr>
<td>SpecJBB</td>
<td>31.3</td>
<td>110</td>
<td>V, HT, E</td>
<td>28</td>
<td>15</td>
<td>88</td>
</tr>
<tr>
<td>Jax</td>
<td>53.1</td>
<td>309</td>
<td>V, HT, E</td>
<td>254</td>
<td>94</td>
<td>706</td>
</tr>
</tbody>
</table>

Table 2: Results of applying our refactoring tool to the benchmarks. For each benchmark, the table shows: (i) the number of declarations that could be migrated and the number of declarations that had to be left unchanged, (ii) the number of allocation sites that could be migrated but not desynchronized, and the number of allocation sites that could not be migrated, and (iii) the number of call sites that could be migrated and that could not be migrated. (iv) the time required to process the benchmark (minutes:seconds).

<table>
<thead>
<tr>
<th>benchmark</th>
<th>declarations (migr./unchanged)</th>
<th>allocation sites (migr.-desync/migrated/unchanged)</th>
<th>call sites (migr./unchanged)</th>
<th>time (min:sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanoi</td>
<td>3/0</td>
<td>3/0/0</td>
<td>26/0</td>
<td>2:07</td>
</tr>
<tr>
<td>JUnit</td>
<td>55/7</td>
<td>23/1/0</td>
<td>111/7</td>
<td>8:29</td>
</tr>
<tr>
<td>JLex</td>
<td>29/10</td>
<td>12/0/4</td>
<td>101/18</td>
<td>9:43</td>
</tr>
<tr>
<td>JavaCap</td>
<td>56/0</td>
<td>14/0/0</td>
<td>155/0</td>
<td>16:56</td>
</tr>
<tr>
<td>Cassowary</td>
<td>121/18</td>
<td>14/0/2</td>
<td>692/36</td>
<td>63:45</td>
</tr>
<tr>
<td>Azurues</td>
<td>13/0</td>
<td>6/6/0</td>
<td>51/0</td>
<td>7:29</td>
</tr>
<tr>
<td>HTMLParser</td>
<td>141/3</td>
<td>21/0/2</td>
<td>461/6</td>
<td>27:35</td>
</tr>
<tr>
<td>JBidWatcher</td>
<td>67/4</td>
<td>32/1/3</td>
<td>291/3</td>
<td>32:17</td>
</tr>
<tr>
<td>SpecJBB</td>
<td>22/6</td>
<td>13/0/2</td>
<td>78/10</td>
<td>9:36</td>
</tr>
<tr>
<td>Jax</td>
<td>208/43</td>
<td>81/0/12</td>
<td>706/0</td>
<td>12:05</td>
</tr>
</tbody>
</table>

interaction with external libraries. In JUnit, one of the
Vectors is eventually passed to the constructor of the exter-

Table 2 shows a chart visualizing the effectiveness of the tool at migrating allocation sites. In this chart, the bottom part of each bar shows the percentage of allocation sites that were migrated without the insertion of synchronization wrappers, and the top part of each bar shows the percentage of allocation sites that could be migrated but that required the insertion of synchronization wrappers. As can be seen in Figure 13, an average of 92.0% of all allocation sites can be migrated (82.8% without the insertion of synchronization wrappers and 9.1% with the insertion of synchronization wrappers). We now discuss a few cases out of the experiments illustrating a number of nontrivial aspects of migration.

Nontrivial rewriting. JBidWatcher, JUnit, and SpecJBB contain calls to the previously discussed method Vector.copyInto() which requires nontrivial rewriting and introduction of an auxiliary class. In JBidWatcher, JUnit, SpecJBB and Jax, the percentages of migrated call sites for which the method’s name or signature was changed are 30%, 75%, 73%, and 47%, respectively. Clearly, manual migration of these applications would involve a significant amount of error-prone editing work.

Interaction with external libraries. In JUnit, one of the
Vectors is eventually passed to the constructor of the exter-
Figure 12: Percentage of declarations and call sites that can be migrated.

Figure 13: Percentage of allocation sites can be migrated. The bottom part of each bar indicates the percentage of allocation sites that can be migrated without inserting synchronization wrappers, and the top part depicts the additional allocation sites that can be migrated with the insertion of synchronization wrappers.

Synchronization preservation. The migration of JUnit includes a synchronization-wrapped allocation site. It is detected as escaping since it is assigned to a field whose declaring class declares a Runnable that references the field. The Runnable object is passed to Swing, which would cause any escape analysis without access to the Swing code to declare it as escaping. In Azureus, escape analysis reports that synchronization wrappers need to be introduced for certain ArrayLists, but the program already performs explicit synchronizations. Hence, in principle, a more precise escape analysis could enable migration without synchronization wrappers in Azureus.

6. RELATED WORK
Several categories of related work can be distinguished, as will be discussed below.

6.1 Refactoring
The field of refactoring and related semantics-preserving program transformations was pioneered in the early 90s by Opdyke and Johnson [23, 24] and by Griswold [16, 17]. Opdyke’s Ph.D. thesis [23, page 27-28] catalogs a number of refactorings, and informally specifies a number of invariants that any refactoring must respect in order to preserve program behavior. For example, the Type-Safe Assignments invariant states that “The type of each expression assigned to a variable must be an instance of the variable’s defined type, or an instance of one of its subtypes. This applies both to assignment statements and function calls”. The same constraints (encoded by rules (1), (5), (8), and (10)) are used in the present paper. The refactorings proposed by Opdyke et al. were implemented in the Smalltalk Refactoring Browser [26], which was the first tool to provide automated support for refactoring.

In recent years, refactoring has been popularized by the emergence of light-weight design methodologies such as “extreme programming” [3] that advocate continuous improve-
ment of a program’s design. For an overview of the field, the reader is referred to books by Fowler [13] and Kerievsky [20], to a recent survey article by Mens and Tourné [22], and to Martin Fowler’s www.refactoring.com site. Popular development environments such as eclipse (see www.eclipse.org) and IntelliJ IDEA (see www.jetbrains.com/idea/) currently offer automated support for a wide range of refactorings. However, we are not aware of any IDE that supports refactorings for migrating between functionally equivalent classes.

Dig and Johnson [8] conducted an empirical study on the role of refactorings in API migration. In their work, successive versions of several frameworks and libraries are examined, and the API-breaking changes are classified as refactorings or as other changes. In each of these frameworks, over 80% of the API-breaking changes are refactorings, thus providing evidence that automated tool support for API migration is highly desirable. Such tool support is provided by Henkel and Diwan [18], who developed a tool that records how refactorings are used to evolve a library. Then, applications that use the library under consideration can be refactored accordingly by applying the recorded refactorings. The work by Henkel and Diwan can be viewed as a tool for automatically deriving migration specifications from refactorings that have been applied to a library. These specifications capture a class of transformations that is more limited than the one we consider, because we consider transformations that cannot be expressed by refactorings.

6.2 Refactorings based on Type Constraints
Tip, Kiezun, and Bäumer [31] use a subset of the constraint generation rules presented in this paper to express the preconditions and to compute the allowable source code modifications for several refactorings related to generalization, including Extract Interface, Pull Up Members, Generalize Type, and Use Supertype Where Possible. Specifically, the refactorings in [31] restructure the type hierarchy and change declarations (but not allocation sites). This work was implemented in Eclipse, and our implementation reuses key parts of this infrastructure.

Donovan et al. [11] and Fuhrer et al. [14] present refactorings that convert Java applications to use Java 1.5 generics [4]. In this work, the assumption is that an application uses a class library (e.g., the Java Standard Collections Framework) for which a generic version has become available, and type inference is used to determine concrete classes that can be used as actual type parameters to instantiate generic library classes. An important benefit of this refactoring is that the use of generic container classes often makes it unnecessary to use down-casts when retrieving elements, and [11, 14] both report that a significant number of down-casts can be removed from the analyzed benchmark applications. Similar to the work in the present paper, Fuhrer et al. extend the basic type constraints of [31], but in a way that is completely different from the rules in the present paper. The work by Donovan et al. [11] uses type constraints that are similar to those used in the present paper, but augmented with type variables and type substitutions to support the inference of generic types. Interestingly, Donovan et al. also use a form of implication constraints (referred to as guarded constraints in [11]), in order to handle situations where the generation of a constraint depends on whether or not a declaration is left raw (i.e., not parameterized). To solve these, their solver also employs a form of backtracking.

Von Dincklage and Diwan [33] present a constraint-based approach for converting non-generic Java classes to use generics and updating non-generic usages of generic classes. Von Dincklage’s tool employs a suite of heuristics that resulted in the successful parameterization of several classes from the Java standard collections.

Most closely related to the work in the present paper is the work by De Sutter et al. [30], who present an optimization method in which Java applications are rewritten to use customized versions of library classes. These custom classes incorporate a number of high-level optimizations that are likely to result in reduced execution speed and heap consumption, and a combination of static analysis and profile information is used to select the optimizations to be applied. The program transformations of [30] are more limited than the ones we consider, because the methods supported by the generated custom classes are always a strict subset of those in the library classes that they replace, and a custom class always has the same superclass as the original class that it replaces. The type constraints used in [30] therefore do not need to allow for situations where auxiliary classes are needed, and hence there is no need for implication constraints (and a backtracking solver). There are several other significant differences between the work in the present paper and [30], most notably the fact that we permit non-trivial user-specified interface mappings between source and target classes, and that we allow the simultaneous migration of interdependent classes.

6.3 Support for Migration and Evolution
Although migration is an important theme in the maintenance of long-lived applications, limited tool support appears to exist, much of which is focused on data and inter-language migration. In [12], type inference is used to reverse-engineer date-related types in a given Cobol program, and migrate them to Year 2000-compliant ones. Unlike the present work, the interface mapping is trivial, since the source and target types are equivalent. A tool to migrate PL/IX programs to C++ programs appears in [21]. Curiously, that work focuses exclusively on language translation, and ignores API migration.

Dmitriev [9] presents a migration tool for converting persistent objects after certain kinds of changes (e.g., replacing a single field with two fields, and the renaming of classes). In this work, changes are specified using a simple specification language, and the programmer supplies converter classes that perform the data conversion. In later work, Dmitriev [10] presents a language-specific make facility for Java, in which various types of dependences between classes are distinguished in order to avoid unnecessary recompilation.

7. CONCLUSIONS AND FUTURE WORK
Even small applications may contain large numbers of references to legacy classes. Migrating these involves a significant amount of error-prone editing work, so automation is highly desirable. We have presented a framework for the automatic migration of (library) classes that has been implemented in
the Eclipse environment. We evaluated the approach on a number of moderate-sized Java applications and found that, in the benchmarks we studied, over 90% of declarations, allocation sites, and call sites were migrated successfully. Moreover, a careful analysis of the results revealed that nontrivial rewriting of method names and signatures was required in many cases. The simple escape analysis we use has proved sufficient to avoid insertion of synchronization wrappers in all but a few cases. Hence, there does not appear to be a need for a more precise escape analysis such as [6].

Plans for future work include supporting the migration of user-defined subclasses of legacy classes, and the introduction of the ADAPTER design pattern [15] to wrap allocation sites and method parameters that could otherwise not be migrated. We also plan to support migrations between methods that differ in terms of thrown exceptions. This situation is similar to the one in Section 2 where the return type of a method in a legacy class (e.g., Vector.elements()) is another legacy class (in this case, Enumeration), and can be handled similarly. An additional factor, however, is the fact that determining all the catch clauses that need to be migrated requires interprocedural analysis, but we do not foresee any major problems here. Lastly, our refactoring tool could be made more useful by providing explanations in cases a construct could not migrated.

Acknowledgments
We would like to thank Robert O’Callahan for suggesting the idea of refactoring Vectors into ArrayLists and Julian Dolby for suggesting an easily-implemented escape analysis algorithm. We are also grateful to Adam Kiezun and the anonymous reviewers for comments on drafts of this paper.

8. REFERENCES
The language for specifying migrations is a language of rewrite rules over templates of Java expressions. A partial grammar is given in Figure 14. The left side of a rewrite rule consists of a Java allocation or method call expression, with additional annotations (e.g., return types) and template variables. A template variable is used for matching the left side of a rule against fragments of the program AST. The right side of a rule consists of an arbitrary Java expression template—an expression with annotations and template variables. Here, the occurrence of a variable denotes a substitution with an appropriate fragment from the program AST, as matched with the left side.

A sample specification for the synchronization-less migration from class Vector to class ArrayList is given in Figure 15. Figure 16 specifies an auxiliary migration, from Utility to Class.

Figure 14: Migration Specification Language Grammar. The nonterminals prim-type and array-type denote any string representing a Java primitive or array type, respectively. Expression reduces to any construct that, after variable substitution, is a valid Java expression. denotes the empty string.

Enumeration to Iterator. Figure 17 presents the auxiliary class used in Figure 15 (a partial version of this class was presented earlier in Figure 3).

public class Util {
    public static void copyInto(List v, Object[] target) {
        if (target == null) throw new NullPointerException();
        for (int i = v.size() - 1; i >= 0; i--)
            target[i] = v.get(i);
    }
    public static void setSize(List v, int newSize) {
        while (v.size() < newSize) v.add(null);
        while (v.size() > newSize) v.remove(newSize);
    }
    public static int indexOf(List v, Object elem, int index) {
        if (elem == null) {
            for (int i = index; i < v.size(); i++)
                if (v.get(i) == null) return i;
        } else {
            for (int i = index; i < v.size(); i++)
                if (elem.equals(v.get(i))) return i;
        }
        return -1;
    }
    public static int lastIndexOf(List v, Object e, int index) {
        if (index >= v.size())
            throw new IndexOutOfBoundsException();
        if (e == null) {
            for (int i = index; i >= 0; i--)
                if (v.get(i) == null) return i;
        } else {
            for (int i = index; i >= 0; i--)
                if (e.equals(v.get(i))) return i;
        }
        return -1;
    }
}

Figure 17: Auxiliary class for migrating Vector to ArrayList

APPENDIX

A. SPECIFICATION LANGUAGE

The language for specifying migrations is a language of rewrite rules over templates of Java expressions. A partial grammar is given in Figure 14. The left side of a rewrite rule consists of a Java allocation or method call expression, with additional annotations (e.g., return types) and template variables. A template variable is used for matching the left side of a rule against fragments of the program AST. The right side of a rule consists of an arbitrary Java expression template—an expression with annotations and template variables. Here, the occurrence of a variable denotes a substitution with an appropriate fragment from the program AST, as matched with the left side.
new Vector(), unsynchronized → new ArrayList()
new Vector(Collection c), unsynchronized → new ArrayList(c)
new Vector(int c), unsynchronized → new ArrayList(c)
void Vector:receiver.copyInto(Object array) → void Util.copyInto(receiver, array)
boolean Vector:receiver.add(Object v) → boolean receiver.add(v)
boolean Vector:receiver.addAll(Collection c) → boolean receiver.addAll(c)
void Vector:receiver.addAll(int i, Collection c) → boolean receiver.addAll(i, c)
void Vector:receiver.addElement(Object v) → void receiver.add(v)
void Vector:receiver.clear() → Object receiver.clone()
boolean Vector:receiver.contains(Object o) → boolean receiver.contains(o)
boolean Vector:receiver.containsAll(Collection c) → boolean receiver.containsAll(c)
Object Vector:receiver.elementAt(int i) → Object receiver.get(i)
Enumeration Vector:receiver.elements() → Iterator receiver.iterator()
void Vector:receiver.ensureCapacity(int c) → void receiver.ensureCapacity(c)
boolean Vector:receiver.equals(Object o) → boolean receiver.equals(o)
Object Vector:receiver.firstElement() → Object receiver.get(0)
int Vector:receiver.get(int i) → int receiver.hashCode()
int Vector:receiver.indexOf(Object o, int i) → int Util.indexOf(receiver, o, i)
void Vector:receiver.insertElementAt(Object o, int i) → boolean receiver.add(i, o)
boolean Vector:receiver.isEmpty() → boolean receiver.isEmpty()
Iterator Vector:receiver.iterator() → Iterator receiver.iterator()
Object Vector:receiver.lastElement() → Object receiver.get(i)
int Vector:receiver.lastIndexOf(Object o, int i) → int Util.lastIndexOf(receiver, o, i)
ListIterator Vector:receiver.ListIterator() → ListIterator receiver.ListIterator()
ListIterator Vector:receiver.ListIterator(int i) → ListIterator receiver.ListIterator(i)
void Vector:receiver.removeAllElements() → void receiver.clear()
boolean Vector:receiver.removeAllElements(int i) → boolean receiver.remove(i)
boolean Vector:receiver.removeAll(Collection c) → boolean receiver.retainAll(c)
Object Vector:receiver.remove(int i) → Object receiver.remove(i)
boolean Vector:receiver.removeElement(Object o) → boolean receiver.remove(o)
void Vector:receiver.removeElementAt(int i) → boolean receiver.remove(i)
boolean Vector:receiver.retainAll(Collection c) → boolean receiver.retainAll(c)
Object Vector:receiver.set(int i, Object o) → Object receiver.set(i, o)
void Vector:receiver.setElementAt(Object o, int i) → Object receiver.set(i, o)
boolean Vector:receiver.size() → void Util.setSize(receiver, s)
int Vector:receiver.size() → int receiver.size()
List Vector:receiver.subList(int f, int t) → List receiver.subList(f, t)
Object[] Vector:receiver.toArray() → Object[] receiver.toArray()
Object[] Vector:receiver.toArray(Object[] a) → Object[] receiver.toArray(Object[] a)
void Vector:receiver.trimToSize() → void receiver.trimToSize()
String Vector:receiver.toString() → String receiver.toString

Figure 15: Specification for migrating from Vector to ArrayList

boolean Enumeration:receiver.hasMoreElements() → boolean receiver.hasNext()
Object Enumeration:receiver.nextElement() → Object receiver.next()

Figure 16: Auxiliary migration from Enumeration to Iterator