

Regular Paper

Constructing Object Groups Corresponding to Concepts for Recovery of a Summarized Sequence Diagram

KUNIHIRO NODA^{1,a)} TAKASHI KOBAYASHI^{1,b)} KIYOSHI AGUSA^{2,c)}

Received: March 10, 2020, Accepted: January 12, 2021

Abstract: Comprehending the behavior of an object-oriented system solely from its source code is troublesome owing to its dynamism. To aid comprehension, visualizing program behavior through reverse-engineered sequence diagrams from execution traces is a promising approach. However, because of the massiveness of traces, recovered diagrams tend to become very large causing scalability issues. To address these issues, we propose an object grouping technique that horizontally summarizes a reverse-engineered sequence diagram. Our technique constructs object groups based on Pree's meta patterns in which each group corresponds to a concept in the domain of a subject system. By visualizing interactions only among important groups, we generate a summarized sequence diagram depicting a behavioral overview of the system. Our experiment showed that our technique outperformed the state-of-the-art trace summarization technique in terms of reducing the horizontal size of reverse-engineered sequence diagrams. Regarding the quality of object grouping, our technique achieved an F -score of 0.670 and a Recall of 0.793 on average under the condition of #lifelines (i.e., the horizontal size of a sequence diagram) < 30 , whereas those of the state-of-the-art technique were 0.421 and 0.670, respectively. The runtime overhead imposed by our technique was 129.2% on average, which is relatively smaller than other figures found in the reference literature.

Keywords: reverse-engineered sequence diagram, trace summarization, object grouping, meta patterns, program comprehension

1. Introduction

Software documentation is a vital source of information for program comprehension; however, such documentation tends to become outdated because numerous modifications can occur after an application is released. Generating reverse-engineered sequence diagrams from execution traces is a promising approach that can help developers comprehend the behavior of a system. Because an execution trace contains a vast amount of information, the trace information must be summarized or abstracted in order to generate reverse-engineered sequence diagrams of a reasonable size [1], [2].

The vertical size of a reverse-engineered sequence diagram increases according to the execution time, while the horizontal size increases in proportion to the number of generated objects. Most existing approaches focus on reducing the vertical size of reverse-engineered sequence diagrams, using strategies such as compacting repetitive behavior [3], [4], [5] and dividing an entire trace into several phases [6], [7], [8], [9]. To improve practicality, horizontal reduction is also important. Some approaches perform horizontal reduction using techniques such as grouping objects [10], [11] and visualizing only interactions related to core (important) objects [12].

In this paper, we propose a technique for reducing the horizon-

tal size of reverse-engineered sequence diagrams, with the goal of helping developers comprehend a behavioral overview of a system. Our technique constructs object groups based on Pree's meta patterns [13], [14] that are the most primitive design patterns. In our technique, each group corresponds to a concept in the domain of a subject system. In conjunction with the existing core object identification technique [12], we identify important object groups and visualize only intergroup interactions. Performing these tasks generates a summarized version of a reverse-engineered sequence diagram. The summarized diagram depicts a behavioral overview of the subject system and is expected to be a valuable tool for developers in an early stage of program comprehension.

To improve maintainability in object-oriented programming, a concept is often divided into several classes by using design patterns [14], [15]. For instance, in a game application, the concept of *player* is divided into several state classes such as *normal state* and *death state* by using the GoF (Gang of Four) state pattern. From the perspective of maintainability, implementing a concept with several classes is beneficial; however, from the perspective of program comprehension, it produces numerous design elements to understand, which can confuse developers and increase the amount of effort required for program comprehension tasks. Our technique constructs object groups based on Pree's meta patterns, in which each group corresponds to a concept. By reunifying divided concepts as object groups, we generate a summarized sequence diagram that depicts a behavioral overview of a system at the concept level.

We evaluated the feasibility and effectiveness of our technique with traces generated from various types of open source software.

¹ Tokyo Institute of Technology, Meguro, Tokyo 152-8552, Japan

² Advanced Science, Technology & Management Research Institute of KYOTO, Kyoto 600-8813, Japan

a) knhr@sa.cs.titech.ac.jp

b) tkobaya@c.titech.ac.jp

c) agusa@astem.or.jp

The results showed that our technique outperformed the state-of-the-art trace summarization technique in terms of horizontal summarization of reverse-engineered sequence diagrams. With regard to the quality of object grouping, under the condition of #lifelines (i.e., the horizontal size of a sequence diagram) being less than 30, the *F*-score and *Recall* of our technique were 0.670 and 0.793 on average, respectively, whereas those of the state-of-the-art technique were 0.421 and 0.670, respectively. Our technique imposed a runtime overhead of 129.2% on average which is relatively small compared with recent scalable dynamic analysis techniques.

The main contributions of this paper are as follows:

- We propose an algorithm for constructing object groups based on Pree’s meta patterns, where each object group corresponds to a concept. This is the first object-grouping algorithm that leverages Pree’s meta patterns in the literature, leading to many opportunities to summarize object behavior.
- We present a new and effective algorithm for drawing a summarized sequence diagram that depicts a behavioral overview as intergroup interactions among important object groups.
- The feasibility and effectiveness of our technique are evaluated using traces generated from various types of open source software.

This paper is an extended version of our previous work [16].

The main differences from our previous work are as follows:

- Clarification and an extension of the algorithms for meta pattern detection and object grouping.
- A quantitative evaluation using traces generated from various open source applications.

The remainder of this paper is organized as follows. In Section 2, we briefly provide the background knowledge required for describing our technique. Section 3 elaborates our trace summarization technique. We evaluate our technique in Section 4 and discuss the threats to validity in Section 5. Section 6 describes key related work and Section 7 concludes this paper.

2. Background

2.1 Pree’s Meta Patterns

Pree’s meta patterns [13], [14] provide a classification of template-hook structures that are commonly used in object-oriented programming. **Figure 1** shows the classification consisting of seven types of template-hook structural patterns. Pree’s meta patterns are the most primitive design patterns. More concrete design patterns (e.g., the GoF design patterns [15]) are realized by using some of the meta patterns. For example, the GoF decorator pattern is an instance of the 11-RCon pattern.

In our technique, by identifying template-hook objects involved in the same meta pattern, we construct object groups that each correspond to a concept.

2.2 Core Object Identification

We previously presented a core object identification technique (COIT) [12]. By analyzing reference relations and dynamic properties, the COIT identifies core objects that are important in comprehending the design overview of a system. The COIT visu-

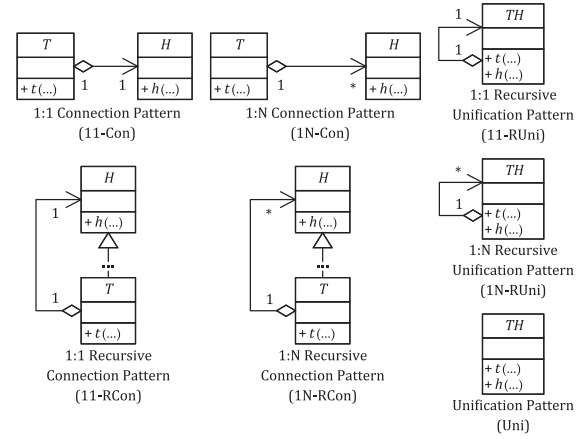


Fig. 1 Pree’s meta patterns. *T*, *H*, *t()*, and *h()* denote a template class, a hook class, a template method, and a hook method, respectively. *TH* denotes that a template class and a hook class are unified into one class. The name in the parentheses below each pattern is the abbreviated name used in this paper.

alizes only interactions related to core objects and thereby generates a summarized sequence diagram that depicts a behavioral overview.

The core identification process consists of the following two steps: (1) eliminating temporary objects and (2) estimating the importance of each object.

In step (1), the COIT analyzes reference relations and lifetimes for each object to identify temporary objects. By analyzing reference relations among objects in a manner similar to compilers’ escape analysis, the COIT identifies the dynamic scope for each object and assigns an escape state to each object. The escape states are categorized into three types: GlobalEscape, ReferenceEscape, and Captured. GlobalEscape (resp. ReferenceEscape) denotes an object that is referenced from another static (resp. non-static) object. An object is marked as Captured if the object is not referenced from any other objects.

Based on the escape state and lifetime for each object, the COIT identifies an object o_i as a temporary object if it satisfies the following condition.

$$\begin{aligned}
 (\text{EscapeState}(o_i) = \text{“Captured”}) \\
 \wedge \text{Lifetime}(o_i) < \text{Lifetime}_{\max}(O) \cdot L_{t\text{-long}} \\
 \vee (\text{EscapeState}(o_i) = \text{“ReferenceEscape”}) \\
 \wedge \text{Lifetime}(o_i) < \text{Lifetime}_{\max}(O) \cdot L_{t\text{-short}}
 \end{aligned}$$

Here, O is a set of all the objects. Lifetime_{\max} returns the maximum lifetime over all the objects. $L_{t\text{-long}}$ and $L_{t\text{-short}}$ are threshold factors for deciding whether an object is long-lived, short-lived, or neither. The above condition means an object o_i is identified as a temporary object if o_i satisfies one of the following conditions: o_i is referenced from no other objects and is not long-lived; o_i is referenced only from other non-static objects and is short-lived.

In step (2), the COIT estimates the importance of each non-temporary object based on the access frequency. Important objects are expected to be heavily accessed from other objects. The COIT calculates the importance value of each object as the harmonic mean of the write, read, and method-invocation frequencies. By building an importance-based object ranking

$R = \langle o_1, o_2, \dots, o_n \rangle$, the COIT identifies an object o_i as a core object if the importance of o_i is greater than the threshold I_t .

3. Recovering a Summarized Sequence Diagram by Constructing Object Groups

In this section, we describe the details of our technique for recovering a summarized version of a reverse-engineered sequence diagram. Our technique consists of the following components:

- (1) Meta pattern detection (Section 3.1).
- (2) Construction of object groups based on meta patterns (Section 3.2).
- (3) Visualization of intergroup interactions among important object groups (Section 3.3).

Figure 2 shows an overview of our technique. Given a subject program, our technique first statically analyzes the source code to detect all meta pattern instances. By utilizing the detected patterns and an execution trace, it constructs object groups corresponding to concepts in the domain of the subject program. Then, leveraging a core identification technique, it identifies important object groups. Finally, intergroup interactions only among important objects are visualized, resulting in a summarized version of a reverse-engineered sequence diagram that depicts a behavioral overview of the subject program.

3.1 Meta Pattern Detection

Because Pree’s meta patterns employ a structural classification, we can identify the meta patterns via static analysis. Algorithm 1 shows our meta pattern detection algorithm. First, we detect hook methods (1.1). Then, we identify meta patterns by detecting template methods that invoke the hook methods (1.2).

In hook method detection, by visiting all the method declaration nodes in an abstract syntax tree (AST), we check whether each method overrides/is-overridden-by another method. If a method overrides/is-overridden-by another method, the method is identified as a hook method (11.3–15). Note that we do not treat all constructors and all methods declared in the topmost type (e.g., `Object#equals(...)` and `Object#toString(...)`) as hook methods.

In meta pattern detection, by traversing all the method invocation nodes in an AST, we test whether each method invokes the detected hook methods. If a method invokes the hook methods, the caller method is treated as a template method, and a meta

pattern consisting of the template method and the invoked hook methods is detected (11.16–26).

We determine the type of each meta pattern by analyzing the relationships between the template and hook classes (11.27–37). Note that the multiplicity of the relationship between the template and hook classes is determined on the basis of the fields of the template class. If a template class refers to a hook class via a field whose type is not a subtype of `java.util.(Iterable|Map)`, we determine the multiplicity as 1 (i.e., 1:1 meta patterns); otherwise, the multiplicity is treated as N (i.e., 1:N meta patterns). Although our algorithm for determining multiplicity generates some false

Algorithm 1 meta pattern detection.

Input: a set of source code $S = \{s_1, s_2, \dots, s_n\}$.

Output: a set of meta-patterns detected $P = \{p_1, p_2, \dots, p_m\}$.

```

1:  $H \leftarrow \text{DETECTHOOKS}(S)$ 
2: return  $\text{DETECTMETAPATTERNS}(S, H)$ 

3: function  $\text{DETECTHOOKS}(S)$ 
  > In: a set of source code  $S$ .
  > Out: a set of sets  $HS$  s.t.  $H \in HS$  is a set of hook methods;
       $m_h \in H$  overrides or is-overridden-by other methods in  $H$ .
4:  $HS \leftarrow \emptyset$ 
5: for each  $s \in S$  do
6:    $a \leftarrow$  the AST of  $s$ 
7:   for each method declaration node  $n$  in  $a$  do
8:      $m \leftarrow$  the method declared in  $n$ 
9:     if  $(\exists H \in HS)[m \in H]$  then continue
10:     $M_{\text{super}} \leftarrow$  all the methods overridden by  $m$ 
11:     $M_{\text{sub}} \leftarrow$  all the methods that override  $m$ 
12:    if  $M_{\text{super}} \neq \emptyset \vee M_{\text{sub}} \neq \emptyset$  then
13:       $H \leftarrow \{m\} \cup M_{\text{super}} \cup M_{\text{sub}}$ 
14:       $HS \leftarrow HS \cup \{H\}$ 
15:  return  $HS$ 

16: function  $\text{DETECTMETAPATTERNS}(S, HS)$ 
  > In: a set of source code  $S$ ; a set of sets of hook methods  $HS$ .
  > Out: a set of meta patterns  $P$ , where  $p \in P$  is a triple  $\langle m_t, H, t \rangle$ ,
       $m_t$  is a template method,  $H$  is a set of hook methods,
      and  $t$  is a meta-pattern type.
17:  $P \leftarrow \emptyset$ 
18: for each  $s \in S$  do
19:    $a \leftarrow$  the AST of  $s$ 
20:   for each method invocation node  $n$  in  $a$  do
21:      $m_i \leftarrow$  the method invoked in  $n$ 
22:      $m_e \leftarrow$  the method enclosing the expression of  $n$ 
23:     if  $(\exists H \in HS)[m_i \in H]$  then
24:        $t \leftarrow \text{DETECTPATTERN}(m_e, m_i)$ 
25:        $P \leftarrow P \cup \{ \langle m_e, H, t \rangle \}$ 
26:  return  $P$ 

27: function  $\text{DETECTPATTERN}(m_t, m_h)$ 
  > In: a template method  $m_t$ ; a hook method  $m_h$ .
  > Out: a meta-pattern type  $t = \langle c, m \rangle$ , where  $c$  is a category name,
      and  $m$  is multiplicity.
28:  $t_t \leftarrow$  the type declaring  $m_t$  (i.e., the template type)
29:  $t_h \leftarrow$  the type declaring  $m_h$  (i.e., the hook type)
30:  $mul \leftarrow$  the multiplicity of the reference from  $t_t$  to  $t_h$ 
31: if  $m_h$  is invoked with the keyword ‘this.’ or ‘super.’ then
32:   return (“Unification”,  $mul$ )
33: if  $t_t$  is equal to  $t_h$  then
34:   return (“Recursive Unification”,  $mul$ )
35: if  $t_t$  is one of the subtypes of  $t_h$  then
36:   return (“Recursive Connection”,  $mul$ )
37: return (“Connection”,  $mul$ )

```

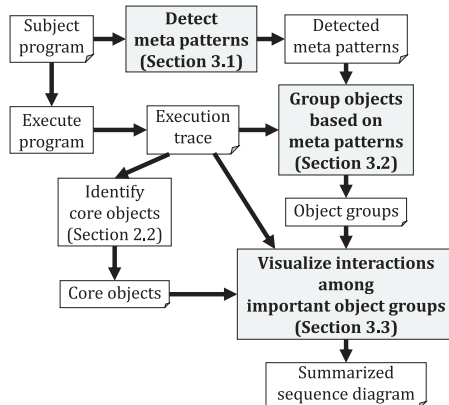


Fig. 2 Overview of our technique.

positives, the final result of object grouping is not affected because our grouping algorithm does not require multiplicity information (described in Section 3.2).

3.2 Object Grouping based on Meta Patterns

3.2.1 Overview

We construct object groups by using the meta pattern information detected in Section 3.1.

As mentioned in Section 1, in object-oriented programming, a concept is often divided into several classes by using design patterns to improve maintainability. Thus, we consider a set of template and hook objects involved in the same meta pattern as corresponding to a concept. Our grouping algorithm identifies template/hook objects for each detected meta pattern and groups those objects.

In a template-hook structure, the template method defines the outline of a process and the hook methods correspond to the details of the process. Our technique groups the template/hook objects such that the behavior of the template method is retained and the behavior of the hook methods is omitted after grouping. The resulting summarized sequence diagram thereby depicts a behavioral outline of a system.

We categorize the seven types of Pree’s meta patterns into the following three pattern types in terms of behavioral aspects.

- Recursive patterns (consisting of 11-RUni, 1N-RUni, 11-RCon, and 1N-RCon).
- Connection patterns (consisting of 11-Con and 1N-Con).
- Unification pattern (consisting of Uni).

For each pattern type, we describe our grouping approach in the following sections.

3.2.2 Object Grouping for the Recursive Patterns

In a meta pattern of the recursive patterns, a template object recursively refers to template/hook objects; that is, a reference chain consisting of template/hook objects is constructed. Typically, the template and hook methods have the same method signature. Once the template method is invoked, template/hook method calls are propagated in the reference chain; as a result, a chain of template/hook method calls is constructed.

For example, assume a concept of *file system* that consists of two elements: *directory* and *file*. Typically, a file system is realized using the GoF composite pattern (i.e., 1N-RCon pattern). There is a FileBase class that declares some abstract methods for file operations. A Dir (resp. File) class that extends the FileBase class as a template (resp. hook) class corresponds to *directory* (resp. *file*).

Assume a method FileBase#getDiskUsage() that returns disk usage information. The getDiskUsage() method is overridden in the Directory and File classes. Once the Dir#getDiskUsage() (a template method) is invoked, template/hook methods (i.e., Dir#getDiskUsage and File#getDiskUsage()) are recursively invoked; as a result, a chain of template/hook method calls is constructed as shown in Fig. 3.

If the template and hook methods have the same method signature, we group all template/hook objects that can be reached by traversing a chain of template/hook method calls. If the signature of the template method is different from that of the hook meth-

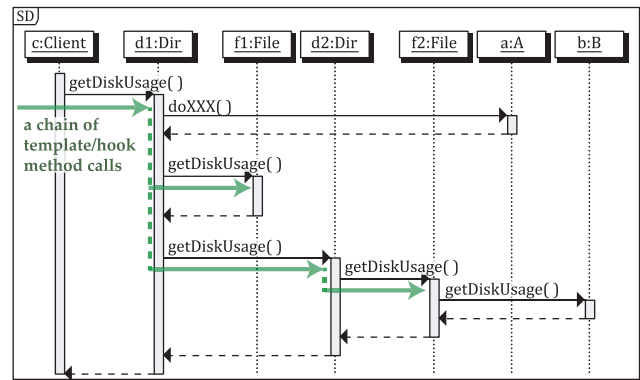


Fig. 3 A chain of template and hook method calls. The Dir (resp. File) class is a template (resp. hook) class. Classes A and B are neither template classes nor hook classes.

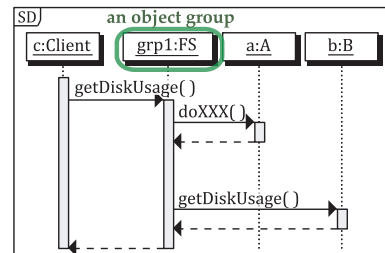


Fig. 4 Resulting summarized sequence diagram that depicts an outline of the behavior shown in Fig. 3. The lifeline named grp1:FS corresponds to the concept of *file system*. The lifeline named b:B is not gathered into grp1:FS because class B is neither a template class nor a hook class.

ods, we group template/hook objects (except for the template object that receives the first message in a chain of template/hook method calls) such that the first template and hook method calls are retained after grouping (Recall that intra-group interactions are omitted in a summarized sequence diagram).

In the case of *file system* shown in Fig. 3, all the template/hook objects (i.e., Dir and File instances) are gathered into one group. Figure 4 shows the resulting summarized sequence diagram that depicts a behavioral outline of Fig. 3. In the summarized diagram, only the first call of getDiskUsage() is retained and successive template/hook method calls are omitted. Note that although B#getDiskUsage() has the same signature as the template/hook methods, the object of class B is not added into the group because the class B is neither a template class nor a hook class.

3.2.3 Object Grouping for the Connection Patterns

In a meta pattern of the connection patterns, a template object refers to hook objects. Typically, the method signature of the template method is different from those of the hook methods. Once the template method is invoked, the template object calls the hook method for each hook object.

For example, assume a GUI application built using Model-View-Controller (MVC) architecture. Once a property of a model object is changed, view objects receive notifications of the property change and update their presentations. This notification and update mechanisms are typically realized using the GoF observer pattern (i.e., 1N-Con pattern). A property change in a model is notified by a template method such as Model#notifyPropertyChanged(...); subsequently, for each hook object, a hook method such as View#onPropertyChanged(...) is

invoked in the template method, which updates the presentations.

The hook objects involved in a connection pattern, whose hook methods are invoked in the template method of the same template object, are gathered into one group. For example, in a GUI application using MVC architecture mentioned above, view objects (i.e., hook objects) that are notified from a model object (i.e., template object) are grouped. The resulting summarized sequence diagram depicts interactions between the model object and the group of the view objects.

3.2.4 Object Grouping for the Unification Pattern

In the unification pattern, the template and hook methods belong to the same class. In other words, a template object also behaves as a hook object. We do not (cannot) perform any grouping for the unification pattern. Note that each non-grouped object will be regarded as a group consisting of a single object during visualization (described in Section 3.3).

3.2.5 Overall Algorithm of Object Grouping

We show the overall algorithm of our object grouping in Algorithm 2. By Algorithm 2, object groups are constructed as mentioned in the previous sections.

We assume that an execution trace is represented in the form of an event sequence based on the behavior model (B-model) proposed by Noda et al. [17]. The B-model represents the behavior of an object-oriented system. The B-model consists of event elements such as *EntryEvent* / *ExitEvent* events, which represent “entry into a constructor/method” / “exit from a constructor/method,” respectively, and *VariableDefinition* / *VariableReference* events, which denote that “a value is assigned to a variable” / “a value is read from a variable,” respectively. An execution trace can be represented in the form $ES = \langle e_1, e_2, \dots, e_n \rangle$, where e_i is an event element in the B-model. ES is composed of all events from all threads. To support thread-sensitive analyses, each event e_i has as its attribute the id of the thread where the event occurs.

In Algorithm 2, once a template method is invoked, either `DETECTOBJGRPFORRECURSIVEPATTERN` or `DETECTOBJGRPFORCONNECTIONPATTERN` is called according to the pattern type, which constructs an object group (ll.1–9).

For the recursive patterns, if the template and hook methods have the same name, the template object is inserted into the resulting group (ll.23–24). The function `CALLEEObj(e)` returns the callee object of the specified entry event e . Then, callee objects in a chain of template/hook method calls are added into the resulting group; that is, if all the invoked methods in the call stack cs are the template/hook methods, all the callee objects are added into the resulting group (ll.25–37). The function `THREADID(e)` returns the id of the thread where e occurs. The function `TRACE` returns a partial trace in which each event occurs during the specified period and on the specified thread. Note that some code fragments in the template/hook methods might be extracted as private-methods whose names are different from those of the template/hook methods. To handle such a case, by ignoring self-calls and focusing on inter-object interactions, we detect a chain of template/hook method calls (ll.19, 20, and 33).

For the connection patterns, we group hook objects whose hook methods are invoked in the template methods of the same template object (ll.40–43).

Algorithm 2 Object grouping based on meta patterns.

Input: an execution trace $ET = \langle e_1, e_2, \dots, e_l \rangle$;
meta patterns detected $P = \{p_1, p_2, \dots, p_m\}$.

Output: a set of object groups $OGS = \{OG_1, OG_2, \dots, OG_n\}$.

```

1:  $T \leftarrow \emptyset$ 
2: for each EntryEvent  $e_i \in ET$  do
3:   for each  $p \in P$  do
4:     if the template method of  $p$  is invoked at  $e_i$  then
5:       if  $p$  belongs to recursive patterns then
6:          $t \leftarrow \text{DETECTOBJGRPFORRECURSIVEPATTERN}(e_i, p)$ 
7:       if  $p$  belongs to connection patterns then
8:          $t \leftarrow \text{DETECTOBJGRPFORCONNECTIONPATTERN}(e_i, p)$ 
9:        $T \leftarrow T \cup t$ 
10:  $OGS \leftarrow \emptyset$ 
11:  $O_{\text{all}} \leftarrow$  all the objects generated during the execution
12: for each  $o_i \in O_{\text{all}}$  do
13:   for each  $p \in P$  do
14:      $OGS \leftarrow OGS \cup \{o \mid o \in OG \wedge \langle o_i, p, OG \rangle \in T\}$ 
15:  $OGS \leftarrow \{OG \mid OG \in OGS \wedge (\nexists OG_i \in OGS)[OG \subset OG_i$ 
16:    $\wedge \exists o_j \exists o_k \exists p[\langle o_j, p, OG \rangle \in T \wedge \langle o_k, p, OG_i \rangle \in T]]\}$ 
17: return  $OGS$ 

18: function DETECTOBJGRPFORRECURSIVEPATTERN( $e_{\text{entry}}, p$ )
  ▷ In: an EntryEvent  $e_{\text{entry}}$ ; a meta pattern  $p = \langle m_t, H, t \rangle$ .
  ▷ Out: a triple  $\langle o_t, p, OG \rangle$ , where  $o_t$  is a template object,
   $p$  is a meta pattern, and  $OG$  is an object group.
19: if  $e_{\text{entry}}$  is a self-call then
20:    $e_{\text{entry}} \leftarrow$  the latest entry event  $e_l$ 
   s.t.  $e_l$  is not a self-call and occurs before  $e_{\text{entry}}$ 
21:  $OG \leftarrow \emptyset$ 
22:  $m_{\text{at}} \leftarrow$  the method invoked at  $e_{\text{entry}}$ 
23: if  $m_{\text{at}}$  and  $m_h \in H$  have the same name then
24:    $OG \leftarrow OG \cup \{\text{CALLEEObj}(e_{\text{entry}})\}$ 
25:  $e_{\text{exit}} \leftarrow$  the ExitEvent that corresponds to  $e_{\text{entry}}$ 
26:  $i_{\text{entry}}, i_{\text{exit}} \leftarrow$  the indices of  $e_{\text{entry}}, e_{\text{exit}}$  in  $ET$ , resp.
27:  $th_{\text{entry}} \leftarrow \text{THREADID}(e_{\text{entry}})$ 
28:  $p_{\text{trace}} \leftarrow \text{TRACE}(i_{\text{entry}} + 1, i_{\text{exit}}, th_{\text{entry}})$ 
29:  $cs \leftarrow$  a new call stack
30: for  $e_i$  in  $p_{\text{trace}}$  do
31:   if  $e_i$  is an EntryEvent then
32:      $cs.\text{push}(e_i)$ 
33:      $M \leftarrow \{m \mid m \text{ is the method invoked at } e_k \in cs$ 
    $\wedge e_k \text{ is not a self-call}\}$ 
34:     if  $(\forall m \in M) [(\exists m_h \in H)$ 
    $[m \text{ has the same name and declaring class as } m_h]]$  then
35:        $OG \leftarrow OG \cup \{\text{CALLEEObj}(e_i)\}$ 
36:     if  $e_i$  is an ExitEvent then
37:        $cs.\text{pop}(e_i)$ 
38:   return  $\langle \text{CALLEEObj}(e_{\text{entry}}), p, OG \rangle$ 

39: function DETECTOBJGRPFORCONNECTIONPATTERN( $e_{\text{entry}}, p$ )
  ▷ The types of the inputs/output are the same
  as DETECTOBJGRPFORRECURSIVEPATTERN.
40:  $e_{\text{exit}} \leftarrow$  the ExitEvent that corresponds to  $e_{\text{entry}}$ 
41:  $i_{\text{entry}}, i_{\text{exit}} \leftarrow$  the indices of  $e_{\text{entry}}, e_{\text{exit}}$  in  $ET$ , resp.
42:  $th_{\text{entry}} \leftarrow \text{THREADID}(e_{\text{entry}})$ 
43:  $O_{\text{callee}} \leftarrow \{o \mid o \text{ is an object that receives a message } m$ 
   $\wedge m \text{ has the same name and declaring class as } m_h \in H$ 
   $\wedge m \text{ comes from } \text{CALLEEObj}(e_{\text{entry}}) \text{ in } \text{TRACE}(i_{\text{entry}}, i_{\text{exit}}, th_{\text{entry}})\}$ 
44: return  $\langle \text{CALLEEObj}(e_{\text{entry}}), p, O_{\text{callee}} \rangle$ 

45: function TRACE( $i_{\text{start}}, i_{\text{end}}, threadId$ )
46: return a partial trace  $p = \langle e_{k_1}, e_{k_2}, \dots \rangle$  extracted from  $ES$ 
   s.t.  $\forall e_{k_i} \in p [i_{\text{start}} \leq i \leq i_{\text{end}} \wedge \text{THREADID}(e_{k_i}) = threadId]$ .

```

Our algorithm ignores the types of the return values and the parameters when testing the equality of methods (ll.23, 34, and

43) or in other words we equate overloaded methods.

After constructing object groups for each meta pattern, we unify object groups having the same template object and meta pattern (ll.12–14); e.g., in the GoF state pattern (i.e., 11-Con pattern), several state objects (i.e., the hook objects) that receive messages from the same template object are gathered into one group. Finally, if an object group is a subset of another group associated with the same meta pattern, we eliminate the subgroup (l.15).

It is worth noting that the template/hook objects involved in a meta pattern p might also be involved in another meta pattern p' . Thus, Algorithm 2 constructs soft clusters of objects; an object is allowed to belong to multiple object groups.

If a template (or hook) method is invoked via reflection, a chain of template-hook method calls is interrupted by some special method calls (e.g., `java.lang.Method#invoke(...)`). For instance, if a template method is invoked via reflection, we would obtain a method call sequence like “template(...) → template(...) → invoke(...) → ... (library-method calls) ... → template(...) → hook(...)”. To deal with this issue, we need to skip such a chain of special library-method calls during object-grouping. In actual circumstances, library method calls usually are not recorded, meaning that entry/exit-events of library methods do not appear in a B-model event sequence. This is because (1) developers usually are not interested in detailed interactions within libraries when program comprehension and (2) tracing library behavior incurs a heavy logging overhead, as described in Section 4.2.4. In case developers would like to know detailed interactions within libraries, reflection would be a limitation of our technique. However, the main objective of this research is to help developers comprehend behavior specific to their subject system rather than libraries; thus, reflection does not greatly impair the usefulness of our technique.

3.2.6 Object Grouping for Delegation Patterns

In object-oriented programming, delegation can be used as an alternative to inheritance; as a result, some delegate methods might appear in a chain of template/hook method calls. Grouping delegate objects that are the callee of those delegate methods together with template/hook objects might improve the quality of object grouping.

For example, in the case of the *file system* mentioned in Section 3.2.2, the File class might delegate the disk usage calculation to another class under certain conditions. In Fig. 3, suppose that class B is a delegate of the File class and `B#getDiskUsage()` is called as a delegate method in `File#getDiskUsage()` under certain conditions. Because the object of the class B is also an element in the concept of *file system*, the object should be added to the object group named `grp1`.

Modifying Algorithm 2 as follows, we can group delegate objects together with template/hook objects.

- Relaxing the condition of line 34 as follows: $(\forall m \in M) [(\exists m_h \in H)[m \text{ has the same name as } m_h]]$ (i.e., the equality check regarding the declaring class is deleted; if a method is invoked by another method having the same name, we consider the invoked method as a delegate method).
- Grouping objects involved in a connection pattern in the

same manner used for grouping objects involved in a recursive pattern. The only exception is that the template object is never added to the resulting group (i.e., ll.23–24 are skipped).

The grouping method that allows (resp. disallows) delegate methods to appear in a method chain of template/hook method calls is referred to as “MP+D” (resp. “MP”). We investigate how grouping delegate objects affects performance in Section 4 (RQ₂). Our technique groups delegate objects by default because “MP+D” outperforms “MP” as described in Section 4.3.2.

3.3 Visualizing intergroup Interactions among Important Object Groups

Combined with the core identification technique (COIT) [12] mentioned in Section 2.2, we identify important object groups. By visualizing intergroup interactions only among the important object groups, we obtain a summarized version of a reverse-engineered sequence diagram that depicts a behavioral overview of a system.

Algorithm 3 shows our algorithm for drawing a summarized sequence diagram. Given an importance-based object ranking created by the COIT, if an object group contains an important object o_{imp} whose importance is greater than the threshold I_t , we treat the object group as an important one that should appear as a lifeline in the resulting summarized sequence diagram (ll.1–8). If no groups contain an important object o_{imp} , we create a new object group whose only member is the object o_{imp} (ll.6–7).

There are two types of methods for drawing a sequence diagram: class-level and instance-level. In a class-level sequence diagram, lifelines having the same types are unified into one lifeline and thereby the horizontal size of the diagram is reduced. Meanwhile, an instance-level sequence diagram provides a de-

Algorithm 3 Drawing a summarized sequence diagram.

Input: an object ranking $R = \langle o_1, o_2, \dots, o_m \rangle$,
 where the importance of o_i is greater than that of o_{i+1} ;
 a threshold of the importance I_t ;
 all the object groups $OGS = \{OG_1, OG_2, \dots, OG_n\}$.

```

1:  $OGS_{imp} \leftarrow \emptyset$ 
2: for  $i \leftarrow 1$  to  $m$  do
3:   if  $IMPORTANCE(o_i) < I_t$  then
4:     break
5:    $S \leftarrow \{OG \mid OG \in OGS \wedge o_i \in OG\}$ 
6:   if  $S = \emptyset$  then
7:      $S \leftarrow \{o_i\}$ 
8:    $OGS_{imp} \leftarrow OGS_{imp} \cup S$ 
9:  $TGS \leftarrow \emptyset$ 
10: for each object group  $OG \in OGS_{imp}$  do
11:    $TGS \leftarrow TGS \cup \{TYPE\_NAME\_SET(OG)\}$ 
12:  $OGS_{class} \leftarrow \emptyset$ 
13: for each  $TG \in TGS$  do
14:    $OG_{class} \leftarrow \{o \mid o \in OG \wedge OG \in OGS_{imp}$ 
      $\wedge TYPE\_NAME\_SET(OG) = TG\}$ 
15:    $OGS_{class} \leftarrow OGS_{class} \cup \{OG_{class}\}$ 
16:  $DRAW\_INTERGROUP\_INTERACTIONS(OGS_{class})$ 

17: function  $TYPE\_NAME\_SET(OG)$ 
   > In: an object group  $OG$ .
   > Out: a set of type names.
18: return  $\{t \mid t \text{ is the type name of } o \in G\}$ 

```

tailed behavioral view for each object, which increases the horizontal size of the diagram. A class-level diagram is useful in an early stage of program comprehension. As the understanding of a subject system deepens, an instance-level sequence diagram would become more suitable.

In Algorithm 3, object groups are converted from instance-level into class-level at lines 9–15. If an instance-level diagram is needed, one only has to skip lines 9 through 15.

The function `DRAWINTERGROUPINTERACTIONS` draws intergroup interactions among the given groups. Each group is visualized as a lifeline in the resulting diagram. For each object group, a *group-id* and a *group-type-name* are displayed in the box on the top of the lifeline where the *group-id* is a unique identifier for the group and the *group-type-name* is the type name of an arbitrary object in the object group.

Because our object grouping constructs soft clusters, an object can belong to multiple groups. Suppose an object o belongs to multiple groups G_1, G_2, \dots, G_n . If an entry event e whose callee object is o occurs, we must determine which of the groups should receive the message corresponding to e . If the message is either the template or hook method of a meta pattern p , and G_i is constructed in regard to p , we send the message to G_i ; otherwise, we send the message to an arbitrary group that contains the object o .

A previous study observed that in software maintenance tasks, successful developers (maintainers) first comprehended high-level structures of a system, then prepared a detailed plan of changes [18]. Another study stated that the most important concepts of a system are usually implemented by very few key classes, and those key classes would be good starting points to comprehend an unfamiliar system [19].

Our summarization technique would help developers build initial knowledge on high-level behavioral overviews of an unfamiliar system. They can in this way obtain a compact view describing what kinds of interactions (events) occur among key concepts, which are important for successful maintenance activities, without investigating tens or hundreds of classes involved in execution scenarios of interest. Thus, we consider that it is suitable to use our technique at the beginning of maintenance tasks so that maintainers can grasp a big picture of a system and the responsibilities of important concepts, which are required for planning proper changes to be made.

After the initial understanding of high-level overviews is established, developers need to dive deeper into the detailed interactions within each important concept (object group). At such a stage, they would need other functionalities for facilitating detailed investigation, such as unfolding grouped lifelines or visualizing (extracting) the behavior only of specific classes of interest. Our technique would not be a valuable aid for those situations. Also, if a maintenance task requires much finer-grained investigation such as checking the values of specific variables (e.g., debugging faulty behavior caused by incorrect variable values), our technique would not be beneficial to developers. These situations are outside the scope of our technique and should be supported by different approaches and tools.

4. Experiment

4.1 Research Questions and Evaluation Approaches

We address the following research questions through the experiment.

RQ₁: How effective is our technique in terms of reducing the horizontal size of reverse-engineered sequence diagrams?

Motivation: We aim to investigate the performance of our technique with respect to reducing the horizontal size of reverse-engineered sequence diagrams, which is the primary objective of this work. If a sequence diagram is small and contains object groups that are important in comprehending a design overview, the diagram is useful for program comprehension.

Evaluation Approach: For each subject system, we extract the ground truth of important concepts from execution scenarios, documents, and tutorials (Section 4.2.2). Utilizing the ground truth, we evaluate the quality of the resulting object groups by using the evaluation measures described in Section 4.2.3. The better the quality measures are, the more useful the resulting summarized diagram will be for program comprehension. We evaluate the effectiveness of our technique by investigating the trade-off between the values of the quality measures and the horizontal size (i.e., #lifelines) of the resulting diagram.

Baseline Selection: Our primary focus is horizontal summarization of reverse-engineered sequence diagram, while most existing studies focus on vertical summarization. Possible candidates of baseline techniques in our experiment are the existing techniques [10], [11], [12], [20], [29] that perform horizontal summarization (Details of the techniques are described in Section 6.1.2).

Because the summarization by OGAN [29] is based on user queries that are not required in our technique, we cannot fairly compare OGAN with our technique. The rest of the candidate techniques [10], [11], [12], [20] are fully automated. Of those, we have to exclude the technique by Dugerdil and Repond [10] because their tool implementation is unavailable. The technique by Toda et al. [11] constructs object groups based on GoF's design pattern usage. Their work is still at an early stage and incomplete. For example, grouping algorithms have been defined only for 5 GoF's design patterns; grouping behavior remains undefined for the rest of the 18 patterns. Also, it is undefined how to group objects that participate in multiple patterns. Thus, we consider it is better to select as our baseline the studies [12], [20] rather than the work by Toda et al. [11]. Because the technique [20] was outperformed in the subsequent study [12], we finally selected the technique in the paper [12] as our baseline.

The baseline technique [12] identifies core objects as described in Section 2.2 and then constructs object groups based on lifetimes and reference relations. With an importance-based object ranking R and a threshold of importance I_t , the baseline technique constructs object groups by hard clustering as follows. For each object o_{imp} whose importance is greater than I_t , the baseline technique gathers objects $o \in \mathcal{O}$ where each object $o \in \mathcal{O}$ is (in)directly referenced from o_{imp} and the lifetime of o is smaller than that of o_{imp} . Thus, for each important object o_{imp} , the base-

Table 1 Subject systems and execution scenarios.

Project	Ver.	KLOC ⁵	Execution scenario
jpacman ¹	SVN r53	6.0	launch the application; start a new game; move the Pac-Man to the right; have the Pac-Man obtain a power cookie and change its state to the power state; quit the application.
JModeller ²	SVN r1015	45.5	launch the application; add two class figures side by side; add attributes and methods; edit the names of the class, attributes, and methods; add connectors indicating association, dependency, and inheritance between the classes; quit the application.
wro4j ³	1.7.7	34.0	execute <i>wro4j-runner</i> while specifying test resources (*.js and *.css files) as target files, 'cssMin' as a pre-processor, and 'jsMin' as a post-processor.
JMeter ⁴	3.1	187.7	execute the application from the command line (non-GUI mode) with the following settings: sending HTTP requests to a web page of a university; #threads=5, #ramp-up=2, and #loop=2; saving the results into report files.

¹ <http://code.google.com/p/jpacman/> ² <https://sourceforge.net/p/jhotdraw/svn/HEAD/tree/trunk/> ³ <http://wro4j.github.io/wro4j/>

⁴ <http://jmeter.apache.org/> ⁵ Test code is excluded.

line technique constructs an object group such that o_{imp} and the gathered objects constitute a composition relation.

The core identification algorithm used by our technique proposed in this paper is the same as the baseline. The difference between our technique and the baseline lies in the algorithms of object-grouping. Namely our technique constructs important object groups based on Pree's meta patterns usage (Section 3.2), while the baseline groups core-related objects based on lifetimes and reference relations as described above.

The baseline technique generates an instance-level sequence diagram, whereas our technique generates a class-level sequence diagram. For comparison, we convert the object groups constructed by the baseline technique into class-level groups by applying 11.9–15 in Algorithm 3.

RQ₂: How much is performance improved by grouping delegate objects together with template/hook objects?

Motivation: As mentioned in Section 3.2.6, delegate methods might appear in a chain of template/hook method calls. In this research question, we investigate how grouping delegate objects together with template/hook objects affects performance.

Evaluation Approach: We compare the performance of "MP+D" and "MP" (see Section 3.2.6). As RQ₁, we investigate the trade-off between the values of the quality measures and the #lifelines.

RQ₃: How much is the runtime overhead imposed by our technique?

Motivation: Our technique weaves logging codes into a subject system for analyzing runtime information, which causes a runtime overhead. In terms of practicality, it is highly important to ensure that the runtime overhead will be small. In this research question, we investigate the overhead imposed by our technique.

Evaluation Approach: For each subject system, we measure the execution times of the woven/original programs to calculate the runtime overhead. We evaluate the overhead through a comparison against the overhead incurred by recent scalable dynamic analysis techniques.

4.2 Experimental Setup

4.2.1 Subject Systems and Execution Scenarios

There are no standard datasets to evaluate trace summarization techniques in the literature; thus, it needs to seek subject systems

that are suitable for the evaluation of our technique. The followings are the criteria we set for subject system selection.

- Each subject system should have documentation especially on its core aspects, or have well-known concepts, so that we can construct ground truth of core-related object grouping.
- The domains of subject systems should be different from one another to improve the external validity of the experiment.
- Each subject system should not be a toy example but a realistic application in order to conduct practical evaluation.
- Subject systems should be publicly available so that future studies can do experiments with the same dataset and compare their performance with ours.

We found the four systems listed in **Table 1** met the above criteria; thus, we select those as our subject systems. Note that it is practically quite difficult to conduct our experiment on dozens or hundreds of subject systems because we have to manually compose a set of ground truths via careful inspection of each system. All the four subjects we select are completely different kinds of realistic applications in different domains: a game application (jpacman), an event-driven GUI modeling application (JModeller), a command line application with file I/O (wro4j), and a multi-threaded application involving network communication (JMeter). Hence, we consider the set of the four systems in Table 1 as a good subject collection for a practical evaluation of our technique. For each subject, we select a representative execution scenario as shown in Table 1.

4.2.2 Ground Truths

For each subject system, we define the ground truth for object grouping. We extract the important concepts in the domain of each subject system from scenario descriptions and documents. Then, we examine the source code and identify the types (classes) that implement the extracted concepts. **Table 2** shows the ground truth for each subject system.

jpacman is a Pac-Man game application written in Java. We define the main entities in the Pac-Man game as the important concepts. These entities include *player* (Pac-Man), *ghost*, and *map*.

JModeller is a modeling application for drawing class diagrams; it is built on top of JHotDraw which is known as a well-designed GUI framework. We extract the important concepts from an introductory article^{*1} that describes the design details of

^{*1} <https://www.javaworld.com/article/2074997/swing-gui-programming/become-a-programming-picasso-with-jhotdraw.html>

Table 2 Important concepts and types corresponding to the concepts (i.e., ground truths).

Project	Concept	Types corresponding to the concept ¹
jpacman	Player	<i>Player, StateManager, Player\$NormalState, Player\$PowerState</i>
	Ghost	<i>Ghost, StateManager, Ghost\$NormalState, Ghost\$EatableState, RunningGhostBrain, RedGhostBrain</i>
	Gem	<i>Gem, EatGem</i>
	Map	<i>Map</i>
JModeller	Class	<i>ClassFigure, SeparatorFigure, GraphicalCompositeFigure, jmodeller.ClassFigure\$(1 4 5) [TextFigure], RectangleFigure, JModellerClass</i>
	Layout	<i>StandardLayouter</i>
	Connection	<i>AssociationLineConnection, DependencyLineConnection, InheritanceLineConnection, JModellerClass</i>
	Tool	<i>DelegationSelectionTool, TextTool, JModellerApplication\$1 [CreationTool], ConnectionTool, UndoableTool</i>
	Tool palette	<i>ToolButton</i>
wro4j	Model	<i>WroModel, Group, Resource</i>
	WroManager	<i>WroManager</i>
	PreProcessor	<i>PreProcessorExecutor\$2 [DefaultProcessorDecorator], BenchmarkProcessorDecorator, ExceptionHandlingProcessorDecorator, SupportAwareProcessorDecorator, MinimizeAwareProcessorDecorator, ImportAwareProcessorDecorator, CssMinProcessor, CSSMin</i>
	PostProcessor	<i>GroupsProcessor\$1 [DefaultProcessorDecorator], BenchmarkProcessorDecorator, ExceptionHandlingProcessorDecorator, SupportAwareProcessorDecorator, MinimizeAwareProcessorDecorator, ImportAwareProcessorDecorator, JSMinProcessor, JSMin</i>
	Locator	<i>StandaloneServletContextUriLocator</i>
JMeter	Test configuration	<i>TestPlan, ThreadGroup, HTTPSamplerProxy, ResultCollector, Arguments, LoopController</i>
	Sampler	<i>HTTPSamplerProxy, HTTPHC4Impl</i>
	Listener	<i>ResultCollector, Summariser, Summariser\$Totals, SummariserRunningSample</i>
	Test executor	<i>StandardJMeterEngine, PreCompiler, JMeterThread</i>

¹ The binary names of the types are shown in the column. The package names of the types are omitted for simplicity. If a type is used as an anonymous class in the code, the name of the base class to extend is shown in the following brackets.

the JModeller application; e.g., *class*, which is a primary element in a class diagram, and *connection*, which defines a relationship between two classes.

wro4j is an application used for improving the loading time of a web application. For example, wro4j provides the functionalities of merging and minifying web resources (e.g., *.js and *.css files). Documentation ^{*2} written by developers provides an overview of the design of wro4j. We consider components appearing in the architecture diagram in the document as the important concepts; e.g., *model*, which represents web resources to process, and *pre/post-processors*, which process web resources.

JMeter is an application that provides functionality for measuring the performance of a web application. The user manual ^{*3} describes the components of JMeter. We extract major concepts relating to our execution scenario from the user manual. In addition, an overview of a test-execution flow is provided in the wiki ^{*4}. We also extract important concepts from the description; e.g., *test configuration*, which specifies the method for measuring the performance, and *sampler*, which sends requests to target servers.

4.2.3 Evaluation Measures

One commonly used external quality measure for evaluating cluster quality is the F-measure. We also use the F-measure to evaluate the quality of object grouping. We calculate the F-measure in the same manner as Steinbach et al. [21], [22].

Let $RS = \{R_1, R_2, \dots, R_n\}$ be the ground truth (the reference set) of grouping, where R_i ($1 \leq i \leq n$) is a set of type names defined in Table 2. Let $OGS_{class} = \{OG_1, OG_2, \dots, OG_m\}$ be the resulting groups of our technique or the baseline technique, where OG_j ($1 \leq j \leq m$) is a set of objects (OGS_{class} is obtained at 1.15 in Algorithm 3). Let $TS = \{T_1, T_2, \dots, T_m\}$ be the result-

ing groups of type names, where $T_i = \{t \mid t \text{ is the type name of } o \in OG_j \wedge OG_j \in OGS_{class}\}$.

First, for the reference group R_i and the resulting group T_j , the *Recall* and *Precision* are defined as follows:

$$Recall(i, j) = |R_i \cap T_j| / |R_i|,$$

$$Precision(i, j) = |R_i \cap T_j| / |T_j|.$$

Then, the *F* score for the reference group R_i and the resulting group T_j is defined as follows:

$$F(i, j) = \frac{2 \cdot Recall(i, j) \cdot Precision(i, j)}{Recall(i, j) + Precision(i, j)}.$$

Finally, the *F* score for the reference set RS and the grouping result TS is calculated as follows:

$$n = \sum_{R_i \in RS} |R_i|,$$

$$F = \sum_{R_i \in RS} \frac{|R_i|}{n} \max_{T_j \in TS} F(i, j).$$

The $\max_{T_j \in TS} F(i, j)$ is the maximum value of $F(i, j)$ over all the elements in TS and R_i . The *F* is the weighted average of $\max_{T_j \in TS} F(i, j)$ over all the elements in RS .

We investigate the trade-off between the *F* score and the horizontal size of the resulting sequence diagram (i.e., #lifelines). Note that, by definition, the *F* score easily reaches 1 if we create the power set of the set of all the objects. Therefore, using only the *F* score is inappropriate for evaluating performance, and we must investigate the trade-off between the *F* score and #lifelines.

The more the objects of the important types shown in Table 2 are contained in the resulting sequence diagram, the more useful the resulting diagram will be for program comprehension. Thus, we also evaluate the rate of the important types contained in the resulting diagram (i.e., the recall). The *Recall* for the reference set RS and the grouping result TS is calculated in the same manner

^{*2} <https://wro4j.readthedocs.io/en/stable/DesignOverview/>

^{*3} <https://jmeter.apache.org/usermanual/index.html>

^{*4} <https://wiki.apache.org/jmeter/JMeterTestExecution>

Table 3 Recorded runtime information.

Project	Total events	Messages	Loaded classes	Objects
jpacman	19,024,287	8,387,304	57	1,273
JModeller	1,074,279	568,940	153	6,993
wro4j	311,192	147,858	295	1,504
JMeter	630,969	333,190	372	5,562

as F :

$$Recall = \sum_{R_i \in RS} \frac{|R_i|}{n} \max_{T_j \in TS} Recall(i, j).$$

4.2.4 Weaving Extent

Our technique requires results of meta pattern detection (Section 3.1) and an execution trace. The execution trace must contain information regarding method-entries/exits, objects, and field accesses. Note that the field access information is used for obtaining an importance-based object ranking R (see Algorithm 3). We use *SELogger*, which is a part of REMViewer [23], for recording execution traces.

We aim to help developers comprehend the behavior specific to the domain of a subject system. For this reason, we weave logging codes only into the classes defined in the subject system. In other words, libraries are not instrumented. As the only exception, collection libraries are instrumented in order to analyze reference relations more precisely (details are provided in the paper describing the baseline technique [12]). This weaving condition is realistic in terms of avoiding a heavy logging overhead.

4.3 Results

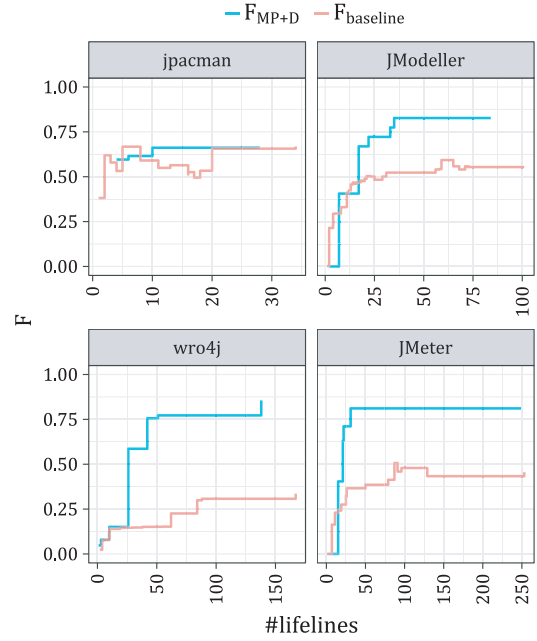
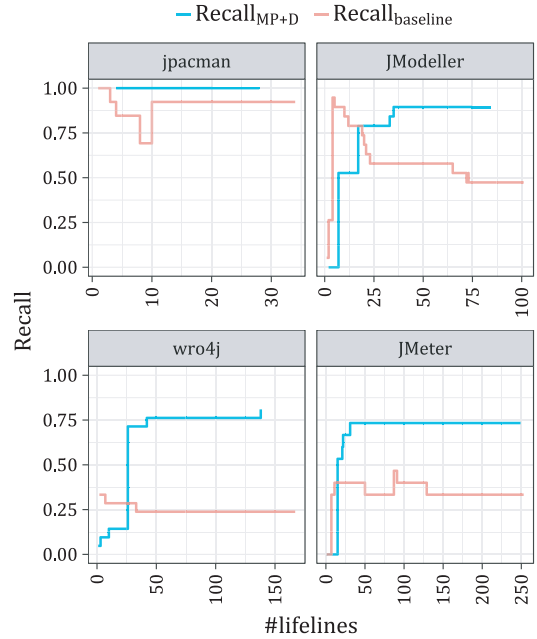
The recorded runtime information is shown in **Table 3**. The numbers of messages and objects affect the size of a reverse-engineered sequence diagram. The number of lifelines (i.e., the horizontal size) of a class-level (resp. instance-level) reverse-engineered sequence diagram is equal to the number of loaded classes (resp. objects).

4.3.1 Answer to RQ₁

Varying the threshold I_t (an input of Algorithm 3) causes the resulting object groups to change, which affects the values of the quality measures (i.e., the F score and $Recall$) and #lifelines. **Fig. 5** and **Fig. 6** show the trade-off relationship between the quality measures and #lifelines.

In our technique, the entire set of resulting object groups (i.e., the output OGS of Algorithm 2) is independent of the value of threshold I_t . As the value of I_t decreases, the number of groups displayed in the resulting diagram increases. Thus, as #lifelines is increased, the values of the quality measures monotonically increase.

On the other hand, in the baseline technique, the entire set of resulting object groups depends on the value of I_t . After the value of I_t is determined, the baseline technique constructs a set of object groups from scratch. Thus, the values of the quality measures do not monotonically increase along with the #lifelines. For instance, if an object o , which (in)directly refers to numerous objects (i.e., o is close to the root object in reference relations), ranks as one of the most important objects, many objects are gathered into the same group, which causes the $Recall$ value to increase. Because the baseline technique constructs hard clusters of objects, reducing the value of I_t increases the number of


Fig. 5 Performance of our technique and the baseline technique (F score vs. #lifelines).

Fig. 6 Performance of our technique and the baseline technique ($Recall$ vs. #lifelines).

groups (i.e., causes the large object group containing o to be decomposed); this might cause the $Recall$ value to decrease, which also affects the F score (e.g., the JModeller case).

As shown in Fig. 5 and Fig. 6, in most of the subject systems, the values of our technique's quality measures increase significantly at the portion of the graph with the fewest #lifelines, and our technique outperforms the baseline technique at #lifelines > 25 .

A sequence diagram with many #lifelines (e.g., > 100) is not suitable for practical use because it requires significant effort for developers to comprehend the content of the diagram. Thus, we focus on the area where #lifelines is less than 30 (i.e., small enough for manual investigation) in Fig. 5 and Fig. 6. The maxi-

Table 4 Maximum performance under the condition of #lifelines < 30. Each shaded cell indicates the maximum performance value for each project.

Project	MP+D		MP		baseline	
	<i>F</i>	<i>Recall</i>	<i>F</i>	<i>Recall</i>	<i>F</i>	<i>Recall</i>
jpacman	0.662	1.000	0.636	0.538	0.667	1.000
JModeller	0.722	0.789	0.655	0.632	0.503	0.947
wro4j	0.586	0.714	0.586	0.714	0.146	0.333
JMeter	0.711	0.667	0.767	0.667	0.366	0.400
Average	0.670	0.793	0.661	0.638	0.421	0.670

imum performance under the condition of #lifelines < 30 is shown as **Table 4**. Table 4 shows that our technique (the “MP+D” column) outperforms the baseline technique for almost all cases. Our technique achieved an *F* score (resp. a *Recall*) of 0.670 (resp. 0.793) on average, which is 0.249 (resp. 0.123) higher than that of the baseline technique. Our technique is therefore more effective compared with the baseline in terms of reducing the horizontal size of reverse-engineered sequence diagrams.

Our technique achieved an *F* score (resp. a *Recall*) of 0.670 (resp. 0.793) on average, under the condition of #lifelines < 30 (which is acceptable for manual investigation); this is 0.249 (resp. 0.123) higher than that of the baseline technique. Thus, our technique is more effective than the baseline technique in terms of reducing horizontal size of reverse-engineered sequence diagrams.

4.3.2 Answer to RQ₂

We show the performance of “MP+D” and “MP” (see Section 3.2.6) in **Fig. 7** and **Fig. 8**. Fig. 7 (resp. Fig. 8) shows the trade-off relationship between the *F* score (resp. the *Recall*) and #lifelines. As shown in Fig. 7 and Fig. 8, allowing delegate methods in a chain of template/hook method calls (i.e., “MP+D”) tends to improve the performance of our technique.

Table 4 shows the maximum performance when there are fewer than 30 #lifelines, which is a size acceptable for manual investigation. As shown in Table 4, the *F* score (resp. the *Recall*) of “MP+D” is 0.009 (resp. 0.155) higher than that of “MP” on average. For all the subject systems, allowing delegate methods in a chain of template/hook method calls increased the number of opportunities for object grouping. Moreover, grouping delegate objects together with template/hook objects improved performance in all cases (except for the *F* score in the JMeter case); i.e., the grouping of delegate objects rarely impaired the performance.

Grouping delegate objects together with template/hook objects improved performance in almost all cases. Under the condition of #lifelines < 30, the *F* score (resp. the *Recall*) of “MP+D” is 0.009 (resp. 0.155) higher than that of “MP” on average.

Regarding RQ₁ and RQ₂, the prevalence of meta patterns and template-hook objects deserves detailed investigation. Our technique constructs object groups based on meta patterns; therefore, the number of opportunities for object grouping depends on the number of template/hook methods and objects in subject systems. If meta patterns are seldom used in subject systems, the number of object groups constructed could be small; this might lessen the effectiveness of our technique. In the following, we show the

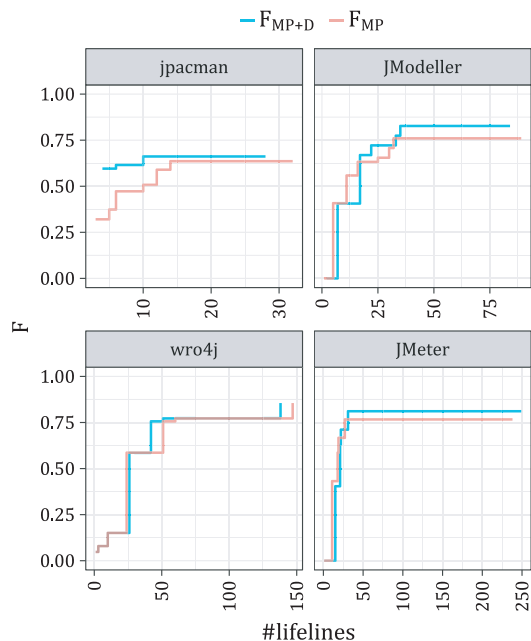


Fig. 7 Effect of allowing delegate methods in a chain of template/hook method calls (*F* score vs. #lifelines).

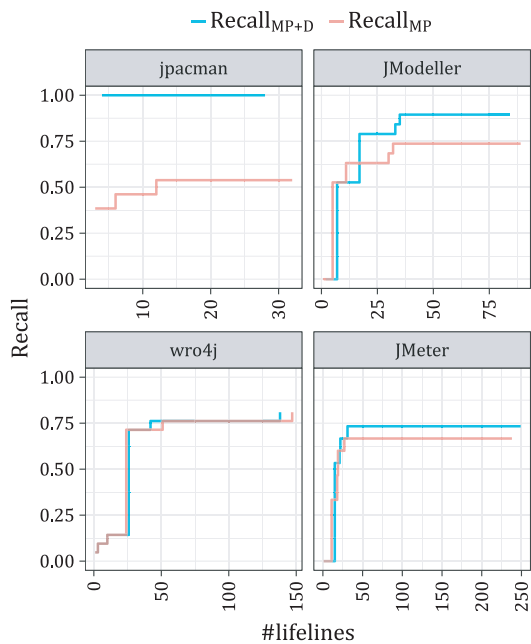


Fig. 8 Effect of allowing delegate methods in a chain of template/hook method calls (*Recall* vs. #lifelines).

results of a detailed investigation into the prevalence of meta patterns and template-hook objects.

Table 5 and **Table 6** show the number of meta patterns and delegations for each subject system. Table 5 shows the number of method invocations (i.e., #pairs of caller and callee methods) that are statically detected (e.g., in the jpacman case, there are 62 pairs of template and hook methods). Table 6 shows the number of pairs of template and hook methods for each pattern type.

As shown in Table 5, 33.2% of all method invocations are hook/delegate method invocations on average. There are a large number of meta patterns and delegations in the subject systems. Moreover, as shown in Table 6, the unification pattern and connection patterns tend to be used more frequently than the other

Table 5 Numbers of method invocations, meta-patterns, and delegations. The “All” column shows the total number of method invocations in each subject system. The “MP” column shows the number of meta-patterns (i.e., #pairs of template and invoked hook methods). The “Delegation” column shows the number of delegations (i.e., #pairs of delegating and delegated methods) in which the delegate method is not a hook method. Each number is statically counted. Static methods, methods declared in `java.lang.Object`, and library methods are excluded. The “Rate of MP+D” is calculated by $(\text{“MP”} + \text{“Delegation”}) / \text{“All”} \cdot 100$.

Project	All	MP	Delegation	Rate of MP+D [%]
jpacman	425	62	8	16.5
JModeller	7,627	3,649	100	49.2
wro4j	5,035	1,837	97	38.4
JMeter	24,477	6,560	480	28.8
Average	9,391	3,027	171	33.2

Table 6 Number of meta-patterns detected.

Project	Uni	11-RUni	1N-RUni	11-Con	1N-Con	11-RCon	1N-RCon
jpacman	10	0	0	4	42	0	6
JModeller	1,596	0	9	336	1,393	143	172
wro4j	309	0	1	727	664	48	88
JMeter	2,214	3	13	687	2,887	335	421

Table 7 Numbers of grouped and non-grouped objects. The “Grouped” column shows the number of objects grouped by “MP+D”. The “Non-grouped” column shows the number of non-grouped objects. The “Rate of grouped” column is the rate of the grouped objects to all the objects. The number outside (resp. inside) the parentheses shows the number of non-temporary objects (resp. all the objects).

Project	Grouped	Non-grouped	Rate of grouped [%]
jpacman	147 (278)	409 (995)	26.4 (21.8)
JModeller	149 (4,625)	119 (2,368)	55.6 (66.1)
wro4j	107 (240)	277 (1,264)	27.9 (16.0)
JMeter	3,962 (4,543)	392 (1,019)	91.0 (81.7)
Average	1,091 (2,422)	299 (1,399)	50.2 (46.4)

patterns.

We show the numbers of grouped and non-grouped objects in **Table 7**. In **Table 7**, the number outside (resp. inside) the parentheses shows the number of non-temporary objects (resp. all the objects). During a program execution, numerous temporary objects are generated and those are not important for program comprehension. We identified temporary objects using the baseline technique [12]. Note that those temporary objects are not included in an importance-based object ranking R , an input of Algorithm 3.

As shown in **Table 7**, on average, 50.2% of the non-temporary objects are grouped by our technique (i.e., 50.2% of the objects are either template, hook, or delegate objects).

As shown in **Table 5** and **Table 7**, our technique uncovered many opportunities for object grouping. In addition, our technique achieved high F scores and $Recalls$ for all the subject systems, regardless of the differences in the rates of grouped objects (see **Table 4**, **Fig. 5**, and **Fig. 6**). In other words, objects corresponding to the important concepts tend to utilize template/hook structures or delegations.

Numerous meta patterns and delegations are used in the subject systems. On average, 33.2% of all the method invocations were hook/delegate method invocations. 50.2% of all the non-temporary objects were grouped by our technique. There were

Table 8 Runtime overhead. The “Base” (resp. “With logging”) column shows the execution time without (resp. with) logging codes. The “Overhead” is calculated by $(\text{“With logging”} - \text{“Base”}) / \text{“Base”} \cdot 100$.

Project	Base [s]	With logging [s]	Overhead [%]
jpacman	9.84	13.76	39.9
JModeller	5.15	10.24	98.9
wro4j	4.59	6.16	34.0
JMeter	4.97	22.06	344.1
Average	6.14	16.54	129.2

many grouping opportunities found by our technique. Therefore, we expect our technique to have a wide range of application. Moreover, our technique achieved high quality object grouping, regardless of the rates of meta patterns and delegations used; objects of important concepts tended to utilize meta patterns or delegations.

4.3.3 Answer to RQ₃

Table 8 shows the runtime overhead. For each execution scenario, we measured the execution time five times, both with and without the logging codes for recording an execution trace. We then calculated the average overhead. We used an Intel Xeon E5-2620 v4 2.10 GHz machine and assigned 16 GB of RAM to the heap of the Java VM. We set the options for *SELLogger* as follows: four background threads were used for writing trace data onto a disk; the trace data was recorded in an uncompressed format.

Our technique imposed a runtime overhead of 129.2% on average. This overhead is relatively small compared with recent scalable dynamic analysis techniques [24], [25], which incur runtime overhead of approximately 100% to 800%.

Developers only need to execute an instrumented application once to produce the behavioral visualization. Thus, in many cases, the overhead from our technique is expected to prove acceptable in a development phase rather than in a production phase.

Our technique imposed a runtime overhead of 129.2% on average, which is relatively small compared with recent dynamic scalable analysis techniques. In many cases, this overhead is expected to be acceptable in program comprehension tasks.

Finally, to facilitate reader understanding, we show an example covering a portion of the resulting (summarized) sequence diagram from the *JModeller* case in **Fig. 9**. **Figure 9** provides the knowledge for a behavioral overview of the subject system as follows.

When a user clicks on the canvas to add a new class figure, the tool objects (the lifeline named “10:DelegationSelectionTool”) receive a mouse-down event (`mouseDown(...)`), and a message communicating the addition of a new figure (`add(...)`) is sent to the view objects (the lifeline named “17:StandardDrawingView”). Then, the view objects are set as containers of the newly added figure object (`addToContainer(...)`).

A paint-message (`paintComponent(...)`) is sent from library code. Then, the view objects draw the figures in the view (`draw(...)`).

When a user clicks on the canvas to add a connector between two classes, the tool objects receive a mouse-up event

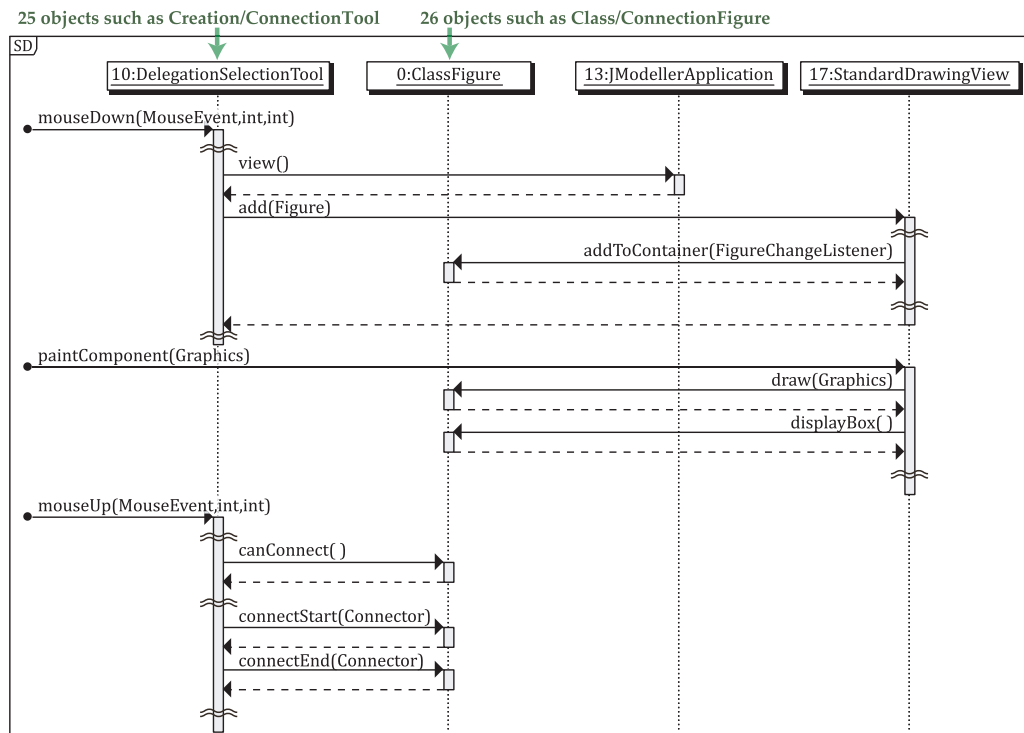


Fig. 9 Portion of the resulting diagram from the *JModeller* case.

(`mouseUp(...)`). After testing whether it is possible to connect the two classes (`canConnect(...)`), the tool objects add a new connection (`connectStart/End(...)`).

In the resulting diagram, objects are abstracted at a concept level (for example, *figure* (“0:ClassFigure”) or *tool* (“10:DelegationSelectionTool”)), which is useful for comprehending the behavioral overview of the subject system. The resulting diagram is expected to become a valuable tool for developers in an early stage of program comprehension.

5. Threats to Validity

To improve the external validity, we used various types of applications in our experiment: a game application that periodically updates/renders the display according to the frame rate (*jpacman*); a modeling application written as an event-driven GUI program (*JModeller*); a command line application that includes file processing (*wro4j*); a multi-threaded command line application that includes network accesses (*JMeter*). However, because the number of subject systems is limited, it is unclear whether the results of our experiment can be generalized further.

Our algorithm groups objects involved in template-hook (and delegation) structures. We cannot theoretically validate whether focusing only on those structures is sufficient for obtaining useful summarization. Instead, we showed the following through our experiment.

- Template-hooks and delegations are widely used in object-oriented systems: over 33% of method invocations are hooks or delegations, and over 50% of non-temporary objects are involved in hooks or delegations (Section 4.3.2). This indicates our algorithm would have many opportunities to summarize object behavior.
- The resulting summarization has not deviated significantly

from those described in execution scenarios and documents: F score and Recall are 0.670 and 0.793, resp (RQ_1). This means the level (degree) of our abstraction is neither too excessive nor too insufficient.

Focusing template-hook (and delegation) structures gives us numerous opportunities to group objects and a good abstraction level. We therefore consider that our algorithm based only on template-hooks (and delegations) works effectively for obtaining behavioral overviews of object-oriented systems.

We defined the ground truths by ourselves. To improve objectivity, we extracted the ground truths from the documents written by developers (except in the case of *jpacman*); however, the definitions of the ground truths might be incorrect. To mitigate the threat, we listed all the ground truths in this paper (Table 2) so that subsequent research can use and validate them.

We did not determine how much software maintenance task time is saved by our summarized sequence diagram; because this study focuses on the compactness of the diagram and the quality of the object grouping, an evaluation of time savings is outside the scope of this paper. Further studies are needed in order to evaluate the usefulness of summarized sequence diagrams in actual maintenance tasks.

6. Related Work

6.1 Coping with Large Execution Traces

To improve program comprehension, testing, and formal verification, many studies have focused on recovering sequence diagrams [1], [2], [26]. The existing research has stated that because execution traces contain vast amounts of information, reverse-engineered sequence diagrams are often afflicted by scalability issues.

6.1.1 Vertical Summarization of a Reverse-engineered Sequence Diagram

The primary approach for coping with the scalability issue is summarizing/abstracting the repetitive behavior in an execution trace [3], [4], [5]. Traces contain substantial amounts of repetitive behavior owing to iterative statements or recursive method calls in a program; thus, summarizing the repetitive behavior results in a significant reduction in the vertical size of a reverse-engineered sequence diagram. To detect iterative behavior in a trace, for example, previous studies utilized debug information [4] or proposed a regular expressions-based approach [5].

An execution trace is often composed of several phases (tasks); thus, phase detection approaches [6], [8], [9], [27], which divide an entire trace into phases, are effective in reducing the vertical size of reverse-engineered sequence diagrams. To detect phases, for example, previous studies examined the creation time of objects [6] or utilized text mining techniques [9].

Hamou-Lhadj and Lethbridge proposed a technique for removing unimportant methods (i.e., utilities) from a trace on the basis of the fan-in/out of each method [28]. By removing unimportant methods (implementation details), they achieved a vertical and horizontal summarization of a reverse-engineered sequence diagram.

6.1.2 Horizontal Summarization of a Reverse-engineered Sequence Diagram

Along with the vertical summarizations, horizontal summarizations of reverse-engineered sequence diagrams are an important factor in improving overall practicality. Some existing trace summarization approaches perform horizontal summarization. The technique proposed in this paper is also categorized as this type.

OGAN is a tool that visualizes interactions between objects related to two classes specified by a user [29]. For each object, OGAN computes the dynamic interaction context, which is a set of classes that use/are-used-by the object. Then, for each class, objects are clustered according to the equality of their dynamic interaction contexts. OGAN selects an object from a cluster for each specified class, and visualizes interactions between those and related objects. If developers already have a certain degree of knowledge about a subject system and know the class names of interest then OGAN is useful for obtaining a representative behavior of the classes. Meanwhile, our technique is more suitable when developers do not know the class names that are important in comprehending the design overview of a subject system (e.g., in an early stage of program comprehension).

Dugerdil and Repond proposed a class clustering technique in which each cluster corresponds to a functional entity in a subject system [10]. Their technique divides an entire trace into segments; then, for each loaded class, it creates a feature vector that represents the binary occurrence (presence (1) or absence (0)) of the class in each segment. If the feature vectors of two classes are similar, those classes are gathered into the same cluster.

A trace, which is generated by exercising several features, consists of several functional phases. Because the technique proposed by Dugerdil and Repond clusters classes from a functional perspective by segmenting a trace and calculating the similari-

ties of the binary occurrence vectors, each of the constructed clusters tends to correspond to a functional phase. As a consequence, the resulting summarized sequence diagram provides a highly abstracted behavioral view that depicts relationships (interactions/flows) among functional entities. Their technique is therefore suitable for comprehending a long/complex execution scenario that exercises several features. Compared with their technique, our technique aims at providing a finer-grained behavioral view. As its input, our technique assumes a simpler trace that exercises a few features, and visualizes only important behavior (interactions among important objects). Thus, our technique is suitable for comprehending how key objects behave in a simpler execution scenario that exercises a few features of interest.

Toda et al. proposed an object-grouping technique based on the GoF design patterns [11]. The key idea of using design patterns for clustering is similar to our approach. The technique by Toda et al. focused on the GoF design patterns, which are more specific than Pree's meta patterns. Therefore, although the risk of grouping irrelevant objects into the same cluster is low, the number of opportunities for object grouping is quite small. Our technique, which clusters objects based on more primitive design patterns, finds more opportunities to summarize object behavior.

We also proposed similar approaches in our previous papers [12], [20]. In the study [20], we first identified core objects based on reference relations and access frequencies. Then, grouping core-related objects based on lifetimes and reference relations, we obtained horizontally-summarized reverse-engineered sequence diagrams that depicted inter-group interactions among the core object groups. The subsequent study [12] presented a refined version of the core identification algorithm (described in Section 2.2) and produced better results.

We selected the study [12] as the baseline in our experiment. As described in Section 4.1, the technique proposed in this paper leverages the core identification algorithm proposed in the baseline paper [12]. The difference between the proposed technique and the baseline lies in the object-grouping algorithms. The proposed technique constructs important object groups based on Pree's meta patterns as described in Section 3.2. Meanwhile, for each important (core) object o_{imp} , the baseline technique identifies a set of objects that are in a composition relation together with o_{imp} , and gathers those objects into the same group. In Section 4, we investigated the performance of the baseline technique, and showed that the technique proposed in this paper outperformed the baseline technique in terms of horizontal summarization of reverse-engineered sequence diagrams.

6.1.3 Other Approaches

There are many other approaches for handling the massive amounts of information in a trace.

Reticella extracted a partial behavioral view from an entire interaction log through a specialized dynamic slicing [17]. The slice (partial behavioral view) of a reverse-engineered diagram is calculated on a B-model (behavior model) event sequence. As with traditional dynamic program slicing, the specialized slicing technique does not summarize (abstract) entire object interactions but pinpoints and extracts partial interactions of interest based on slicing criteria. Alimadadi et al. identified recurring struc-

tures (motifs) in a trace and represented the trace as a hierarchical view [30]. Busany and Maoz proposed a technique that suggested a stopping criteria for a trace analysis (e.g., the sample size of a trace), such that the analysis results could be statistically trusted at a specified level of confidence [31].

Some existing works proposed effective visualization tools: a dedicated view for comprehending a large trace [32], an interactive tool that effectively explores a sequence diagram [33], and a generic toolkit that provides a set of functionalities for smoothly exploring a massive-scale sequence diagram [34].

6.2 Detecting and Leveraging Design Patterns

A number of design pattern detection techniques have been proposed [35], [36]. While most existing techniques focus on GoF design patterns, some research detected and leveraged the Pree's meta patterns, which are the most primitive design patterns [37], [38]. Because some GoF design patterns have similar structures, detecting these patterns produces some false positives. On the other hand, because each Pree's meta pattern is structurally distinguishable, it is relatively easy to detect these patterns without false positives.

Detecting the meta patterns, Hayashi et al. reduced the computation and maintenance costs of the GoF design pattern detection logic [37]. Posnett et al. proposed a meta pattern detector named Thex that worked on Java bytecode and is scalable to larger codebases [38]. Our technique also detects meta patterns for constructing object groups corresponding to concepts.

7. Conclusion

The behavior of an object-oriented program is visualizable as a reverse-engineered sequence diagram, which is a valuable tool for program comprehension; however, owing to the massive size of execution traces, reverse-engineered sequence diagrams often suffer from scalability issues.

To address the issues, in this paper we propose a trace summarization technique that reduces the horizontal size of reverse-engineered sequence diagrams. Our technique constructs object groups in which each object corresponds to a concept in the domain of a subject system. Then, given an importance-based object ranking, important object groups are identified. By visualizing intergroup interactions only among the important groups, our technique generates a summarized version of a reverse-engineered diagram that depicts a behavioral overview of the subject system.

We evaluated our technique using traces generated from various types of open source software. The results showed that our technique outperformed the state-of-the-art trace summarization technique in terms of horizontal summarization of reverse-engineered sequence diagrams. Regarding the quality of object grouping, our technique achieved an F score (resp. a $Recall$) of 0.670 (resp. 0.793) on average, under the condition of $\#lifelines < 30$ (which is acceptable for manual investigation). This is 0.249 (resp. 0.123) higher than that of the baseline technique. Moreover, our technique imposed a runtime overhead of 129.2% on average, which is relatively small compared with recent scalable dynamic analysis techniques. In many cases, the runtime overhead of our technique is expected to be acceptable in maintenance

tasks.

Overall, our technique can recover a summarized sequence diagram that depicts a behavioral overview, while incurring a small runtime overhead. The resulting summarized diagram is expected to be a valuable tool for developers in an early stage of program comprehension.

Acknowledgments This work was partly supported by MEXT KAKENHI Grant Numbers JP15H02683, JP18H03221, and JP18J15087.

References

- [1] Bennett, C., Myers, D., Storey, M.A., German, D.M., Ouellet, D., Salois, M. and Charland, P.: A survey and evaluation of tool features for understanding reverse-engineered sequence diagrams, *Journal of Software Maintenance and Evolution: Research and Practice*, Vol.20, No.4, pp.291–315 (2008).
- [2] Ghaleb, T.A., Alturki, M.A. and Aljasser, K.: Program comprehension through reverse-engineered sequence diagrams: A systematic review, *Journal of Software: Evolution and Process*, Vol.30, No.11, pp.1–30 (2018).
- [3] Taniguchi, K., Ishio, T., Kamiya, T., Kusumoto, S. and Inoue, K.: Extracting sequence diagram from execution trace of Java program, *Proc. International Workshop on Principles of Software Evolution*, pp.148–154 (2005).
- [4] Myers, D., Storey, M.A. and Salois, M.: Utilizing debug information to compact loops in large program traces, *Proc. European Conference on Software Maintenance and Reengineering*, pp.41–50 (2010).
- [5] Jayaraman, S., Jayaraman, B. and Lessa, D.: Compact visualization of Java program execution, *Software: Practice and Experience*, Vol.47, No.2, pp.163–191 (2016).
- [6] Watanabe, Y., Ishio, T. and Inoue, K.: Feature-level phase detection for execution trace using object cache, *Proc. International Workshop on Dynamic Analysis*, pp.8–14 (2008).
- [7] Ishio, T., Watanabe, Y. and Inoue, K.: AMIDA: A sequence diagram extraction toolkit supporting automatic phase detection, *Proc. International Conference on Software Engineering*, pp.969–970 (2008).
- [8] Pirzadeh, H., Shanian, S., Hamou-Lhadj, A., Alawneh, L. and Shafiee, A.: Stratified sampling of execution traces: Execution phases serving as strata, *Science of Computer Programming*, Vol.78, No.8, pp.1099–1118 (2013).
- [9] Pirzadeh, H., Hamou-Lhadj, A. and Shah, M.: Exploiting text mining techniques in the analysis of execution traces, *Proc. International Conference on Software Maintenance*, pp.223–232 (2011).
- [10] Dugerdil, P. and Repond, J.: Automatic generation of abstract views for legacy software comprehension, *Proc. India Software Engineering Conference*, pp.23–32 (2010).
- [11] Toda, T., Kobayashi, T., Atsumi, N. and Agusa, K.: Grouping objects for execution trace analysis based on design patterns, *Proc. Asia-Pacific Software Engineering Conference*, pp.25–30 (2013).
- [12] Noda, K., Kobayashi, T. and Atsumi, N.: Identifying core objects for trace summarization by analyzing reference relations and dynamic properties, *IEICE Trans. Information and Systems*, Vol.E101.D, No.7, pp.1751–1765 (2018).
- [13] Pree, W.: Meta patterns — A means for capturing the essentials of reusable object-oriented design, *Proc. European Conference on Object-Oriented Programming*, pp.150–162 (1994).
- [14] Pree, W.: *Design Patterns for Object-Oriented Software Development*, ACM Press/Addison-Wesley Publishing Co. (1995).
- [15] Gamma, E., Helm, R., Johnson, R. and Vlissides, J.: *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley Longman Publishing Co., Inc. (1995).
- [16] Noda, K., Kobayashi, T. and Agusa, K.: Execution trace abstraction based on meta patterns usage, *Proc. Working Conference on Reverse Engineering*, pp.167–176 (2012).
- [17] Noda, K., Kobayashi, T., Yamamoto, S., Saeki, M. and Agusa, K.: Reticella: An execution trace slicing and visualization tool based on a behavior model, *IEICE Trans. Information and Systems*, Vol.E95.D, No.4, pp.959–969 (2012).
- [18] Robillard, M.P., Coelho, W. and Murphy, G.C.: How effective developers investigate source code: An exploratory study, *IEEE Trans. Software Engineering*, Vol.30, No.12, pp.889–903 (2004).
- [19] Zaidman, A. and Demeyer, S.: Automatic identification of key classes in a software system using webmining techniques, *Journal of Software Maintenance and Evolution: Research and Practice*, Vol.20, No.6, pp.387–417 (2008).
- [20] Noda, K., Kobayashi, T., Toda, T. and Atsumi, N.: Identifying core

objects for trace summarization using reference relations and access analysis, *Proc. Annual Computer Software and Applications Conference*, pp.13–22 (2017).

- [21] Steinbach, M., Karypis, G. and Kumar, V.: A comparison of document clustering techniques, *Proc. KDD Workshop on Text Mining* (2000).
- [22] Steinbach, M., Karypis, G. and Kumar, V.: A comparison of document clustering techniques, Technical Report #00-034, University of Minnesota (2000).
- [23] Matsumura, T., Ishio, T., Kashima, Y. and Inoue, K.: Repeatedly-executed-method viewer for efficient visualization of execution paths and states in Java, *Proc. International Conference on Program Comprehension*, pp.253–257 (2014).
- [24] Madsen, M., Tip, F., Andreasen, E., Sen, K. and Møller, A.: Feedback-directed instrumentation for deployed JavaScript applications, *Proc. International Conference on Software Engineering*, pp.899–910 (2016).
- [25] Huang, J.: Scalable thread sharing analysis, *Proc. International Conference on Software Engineering*, pp.1097–1108 (2016).
- [26] Lo, D. and Maoz, S.: Towards succinctness in mining scenario-based specifications, *Proc. International Conference on Engineering of Complex Computer Systems*, pp.231–240 (2011).
- [27] Pirzadeh, H. and Hamou-Lhadj, A.: A novel approach based on gestalt psychology for abstracting the content of large execution traces for program comprehension, *Proc. International Conference on Engineering of Complex Computer Systems*, pp.221–230 (2011).
- [28] Hamou-Lhadj, A. and Lethbridge, T.: Summarizing the content of large traces to facilitate the understanding of the behaviour of a software system, *Proc. International Conference on Program Comprehension*, pp.181–190 (2006).
- [29] Munakata, S., Ishio, T. and Inoue, K.: OGAN: Visualizing object interaction scenarios based on dynamic interaction context, *Proc. International Conference on Program Comprehension*, pp.283–284 (2009).
- [30] Alimadadi, S., Mesbah, A. and Pattabiraman, K.: Inferring hierarchical motifs from execution traces, *Proc. International Conference on Software Engineering*, pp.776–787 (2018).
- [31] Busany, N. and Maoz, S.: Behavioral log analysis with statistical guarantees, *Proc. International Conference on Software Engineering*, pp.877–887 (2016).
- [32] Cornelissen, B., Zaidman, A. and van Deursen, A.: A controlled experiment for program comprehension through trace visualization, *IEEE Trans. Software Engineering*, Vol.37, pp.341–355 (2011).
- [33] Sharp, R. and Rountev, A.: Interactive exploration of UML sequence diagrams, *Proc. International Workshop on Visualizing Software for Understanding and Analysis*, pp.1–6 (2005).
- [34] Lyu, K., Noda, K. and Kobayashi, T.: SDEplorer: A generic toolkit for smoothly exploring massive-scale sequence diagram, *Proc. International Conference on Program Comprehension*, pp.380–384 (2018).
- [35] Dong, J., Zhao, Y. and Peng, T.: A review of design pattern mining techniques, *International Journal of Software Engineering and Knowledge Engineering*, Vol.19, No.6, pp.823–855 (2009).
- [36] Ampatzoglou, A., Charalampidou, S. and Stamelos, I.: Research state of the art on GoF design patterns: A mapping study, *Journal of Systems and Software*, Vol.86, No.7, pp.1945–1964 (2013).
- [37] Hayashi, S., Katada, J., Sakamoto, R., Kobayashi, T. and Saeki, M.: Design pattern detection by using meta patterns, *IEICE Trans. Information and Systems*, Vol.E91.D, No.4, pp.933–944 (2008).
- [38] Posnett, D., Bird, C. and Devanbu, P.: Thex: Mining metapatterns from Java, *Proc. Working Conference on Mining Software Repositories*, pp.122–125 (2010).



Takashi Kobayashi is an associate professor in the Department of Computer Science, School of Computing, Tokyo Institute of Technology. His research interests include software reuse, software design, software maintenance, program analysis, software configuration management, Web-services compositions, workflow, multimedia information retrieval, and data mining. He received B.Eng., M.Eng. and Dr.Eng. degrees in computer science from Tokyo Institute of Technology in 1997, 1999, and 2004, respectively. He is a member of the JSSST, IPSJ, DBSJ, IEEE-CS, and ACM.



Kiyoshi Agusa received his B.Sc. and M.Sc. degrees in electrical engineering, and Ph.D. degree in computer science from Kyoto University in 1970, 1972 and 1980, respectively. He is a president, vice-chairman of Advanced Science, Technology & Management Research Institute of KYOTO (ASTEM). His research interests lie in the field of software engineering, especially programming environments, and requirements.



Kunihiro Noda received his B.Sc. degree in engineering and M.Sc. degree in information science from Nagoya University, and his Ph.D. degree in computer science from Tokyo Institute of Technology. He had worked as a software engineer in an automotive company for around five years, and is currently working as a software engineering researcher in industry. His research interests include program analysis, program comprehension, specification mining, automated program repair, and debugging.

His research interests include program analysis, program comprehension, specification mining, automated program repair, and debugging.