



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCE  
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**UNDERGRADUATE THESIS**

## **Doop-Soot: Parallel Fact Generation**

**Mouris Dimitris**

**Supervisors:** **Smaragdakis Yannis**, Associate Professor NKUA  
**Antoniadis Anastasis**, M.Sc. Student NKUA

**ATHENS**

**JULY 2016**

# **UNDERGRADUATE THESIS**

Doop-Soot: Parallel Fact Generation

**Mouris Dimitris**

**R.N.: 1115201200114**

**SUPERVISORS:** **Smaragdakis Yannis**, Associate Professor NKUA  
**Antoniadis Anastasis**, M.Sc. Student NKUA



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Door-Soot: Παραλληλοποίηση της Δημιουργίας Γεγονότων**

**Μούρης Δημήτρης**

**Επιβλέποντες:** **Smaragdakis Yannis**, Associate Professor NKUA  
**Antoniadis Anastasis**, M.Sc. Student NKUA  
**Σμαραγδάκης Γιάννης**, Αναπληρωτής Καθηγητής ΕΚΠΑ  
**Αντωνιάδης Αναστάσης**, Μεταπτυχιακός Φοιτητής ΕΚΠΑ

**ΑΘΗΝΑ**

**ΙΟΥΛΙΟΣ 2016**

## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

Door-Soot: Παραλληλοποίηση της Δημιουργίας Γεγονότων

**Μούρης Δημήτρης**

**A.M.: 1115201200114**

**ΕΠΙΒΛΕΠΟΝΤΕΣ:** **Smaragdakis Yannis**, Associate Professor NKUA  
**Antoniadis Anastasis**, M.Sc. Student NKUA  
**Σμαραγδάκης Γιάννης**, Αναπληρωτής Καθηγητής ΕΚΠΑ  
**Αντωνιάδης Αναστάσης**, Μεταπτυχιακός Φοιτητής ΕΚΠΑ

## ΠΕΡΙΛΗΨΗ

Παραλληλοποίηση του Fact Generation του Doop. Το Doop χρησιμοποιείται για μπλαμπλαμπλ

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Τεκμηρίωση

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** static program analysis, doop: fact generation, soot, πτυχιακές εργασίες, τμήμα πληροφορικής και τηλεπικοινωνιών

Πανεπιστήμιο Αθηνών

## **ABSTRACT**

Παραλληλοποίηση του Fact Generation του Doop. Το Doop χρησιμοποιείται για μπλαμπλαμπλαμπλ

**SUBJECT AREA:** Documentation

**KEYWORDS:** static program analysis, doop: fact generation, soot, undergraduate thesis, dept. of informatics

University of Athens

*Αφιέρωση σε κάποιους.*

# ACKNOWLEDGEMENTS

blahblahblahblah

blahblahblahblah

blahblahblahblah



# CONTENTS

|                                      |           |
|--------------------------------------|-----------|
| <b>PROLOGUE</b>                      | <b>10</b> |
| <b>1. INTRODUCTION</b>               | <b>11</b> |
| <b>2. DOOP</b>                       | <b>12</b> |
| 2.1 Fact Generation                  | 12        |
| 2.2 Doop Time Examples               | 12        |
| 2.3 Fact Table Example               | 13        |
| <b>3. SOOT</b>                       | <b>14</b> |
| 3.1 Applying the latest Soot version | 14        |
| 3.2 Bytecode To Jimple               | 15        |
| 3.3 Compiling & Running Soot         | 16        |
| 3.4 Jimple Examples                  | 16        |
| 3.4.1 Hello World                    | 17        |
| 3.4.2 Inheritance Test               | 18        |
| <b>4. FOUR APPROACHES</b>            | <b>20</b> |
| 4.1 One Thread Per Method            | 21        |
| 4.2 One Thread Per Class             | 22        |
| 4.3 Fork-Join Framework              | 23        |
| 4.4 Multiple Classes Per Thread      | 25        |
| <b>5. LOCKING</b>                    | <b>29</b> |
| 5.1 Doop Side                        | 29        |
| 5.1.1 CSVDatabase                    | 29        |
| 5.1.2 Repesantation                  | 29        |
| 5.2 Soot Side                        | 29        |
| 5.2.1 Type Assigner                  | 30        |
| 5.2.2 Pack Manager                   | 30        |
| 5.2.3 Shimple -ssa                   | 30        |

|   |           |
|---|-----------|
| <b>6. TIME RESULTS . . . . .</b>            | <b>31</b> |
| <b>ACRONYMS AND ABBREVIATIONS . . . . .</b> | <b>33</b> |
| <b>REFERENCES . . . . .</b>                 | <b>34</b> |

## LIST OF FIGURES

|           |  |    |
|-----------|--|----|
| Σχήμα 1:  | Facts Example . . . . .  | 13 |
| Σχήμα 2:  | Fact Generation with different Soot versions . . . . .               | 15 |
| Σχήμα 3:  | Compiling Soot . . . . .   | 16 |
| Σχήμα 4:  | Generating Jimple from .class . . . . .                              | 16 |
| Σχήμα 5:  | Generating Jimple from .jar . . . . .                                | 16 |
| Σχήμα 6:  | HelloWorld.java . . . . .  | 17 |
| Σχήμα 7:  | HelloWorld.jimple . . . . .  | 17 |
| Σχήμα 8:  | inheritanceTest.java . . . . .                                       | 18 |
| Σχήμα 9:  | inheritanceTest.jimple . . . . .                                     | 19 |
| Σχήμα 10: | Sequential Fact Generation . . . . .                                 | 20 |
| Σχήμα 11: | One Thread Per Method . . . . .                                      | 21 |
| Σχήμα 12: | One Thread Per Class . . . . .                                       | 22 |
| Σχήμα 13: | Fork-Join Basic-Use . . . . .  | 23 |
| Σχήμα 14: | Fork-Join Framework . . . . .  | 24 |
| Σχήμα 15: | Multiple Classes Per Thread . . . . .                                | 26 |
| Σχήμα 16: | Increasing the number of threads with 3 classes per thread . . . . . | 28 |
| Σχήμα 17: | Increasing the number of threads with 4 classes per thread . . . . . | 28 |
| Σχήμα 18: | CSVDatabase.java . . . . .   | 29 |
| Σχήμα 19: | Repesantation.java . . . . .   | 29 |
| Σχήμα 20: | JimpleBodyPack.java . . . . .  | 30 |
| Σχήμα 21: | PackManager.java . . . . .   | 30 |
| Σχήμα 22: | Fact Generation Time Results . . . . .                               | 31 |

## LIST OF TABLES

|  |    |
|--|----|
| Πίνακας 1: Soot 2.5.0 times . . . . .  | 12 |
| Πίνακας 2: Sequential Fact Generation Time Examples . . . . .  | 21 |
| Πίνακας 3: One Thread Per Method Time Examples, with pool size: 4, 16, 32 . . . . .  | 22 |
| Πίνακας 4: One Thread Per Class Time Examples, with pool size: 4, 16, 32 . . . . .   | 23 |
| Πίνακας 5: Fork/Join Time Examples, with threshold 2, 3, 4 and pool size 16 . . . . .  | 25 |
| Πίνακας 6: Multiple Classes Per Thread Time Examples, with pool size: 4, 16,<br>32 and classes per thread: 2, 3, 4 . . . . . | 27 |
| Πίνακας 7: Summarizing best times of all approaches with pool size 16-32 . . . . .   | 32 |

## PROLOGUE

Το παρόν έγγραφο δημιουργήθηκε στην Αθήνα, το 2016

blahblahblahblah

blahblahblahblah

## 1. INTRODUCTION

eisagwgh gia doop kai soot

## 2. DOOP

From Doop Website:

*"Doop is a framework for pointer, or points-to, analysis of Java programs. Doop implements a range of algorithms, including context insensitive, call-site sensitive, and object-sensitive analyses, all specified modularly as variations on a common code base."*

*"Doop builds on the idea of specifying pointer analysis algorithms declaratively, using Datalog: a logic-based language for defining (recursive) relations. Doop carries the declarative approach further than past work by describing the full end-to-end analysis in Datalog and optimizing aggressively through exposition of the representation of relations (for example indexing) to the Datalog language level. Doop uses the Datalog dialect and engine of LogicBlox."*

Compared to alternative context-sensitive pointer analysis implementations (such as Paddle) Doop is much faster, and scales better. Also, with comparable context-sensitivity features, Doop is more precise in handling some Java features (for example exceptions) than alternatives. //kati prepei na alla3ei edw..

Doop is launched by specifying the type of analysis to run and the target directory that contains java bytecode (.jar file). First doop generates the facts and imports them into a database -or more precisely a knowlegdebase- and then it runs the analysis that was asked to. [1]

### 2.1 Fact Generation

Doop before running a pointer or points-to analysis, intergrates with Soot to generate either Jimple (**J**ava **s**imple) or Shimple (a **S**sa version of **J**imple) intermediate representations. Jimple is a typed 3-address IR suitable for performing optimizations; it only has 15 statements. Then from jimple the facts are generated and imported into a database with multiple tables. Shimple is a SSA-version of jimple; obviously first jimple is generated and then Soot applies a transformation pack to jimple body to create shimple.

The main problem is that fact generation takes more than 50% of total execution time (either jimple or shimple). Below are presented a few time examples of the sequential Doop Fact Generation and total execution time (both fact generation and analysis times).

### 2.2 Doop Time Examples

| Soot 2.5.0             | antlr.jar | hsqldb.jar | batik.jar |
|------------------------|-----------|------------|-----------|
| <b>Fact Generation</b> | 1.16 min. | 1.23 min.  | 2.26 min. |
| <b>Total time</b>      | 3.18 min. | 3.21 min.  | 4.34 min. |

Table 1: Soot 2.5.0 times

## 2.3 Fact Table Example

Here are presented some of the facts that created for the *helloWorld* example.

---

```

1 facts/ActualParam.facts:0 helloWorld.main/java.io.PrintStream.println/0
  helloWorld.main/$stringconstant0
2 facts/ApplicationClass.facts:helloWorld
3 facts/AssignHeapAllocation.facts:helloWorld.main/invoke/instruction3 3 helloWorld
  helloWorld.main/$stringconstant0 <helloWorld: void main(java.lang.String[])>
4 facts/AssignLocal.facts:helloWorld.<init>/definition / instruction1 1 helloWorld.<init>/@this
  helloWorld.<init>/r0 <helloWorld: void <init>()>
5 facts/AssignLocal.facts:helloWorld.main/definition / instruction1 1
  helloWorld.main/@param0 helloWorld.main/r0 <helloWorld: void main(java.lang.String[])>
6 facts/ClassObject.facts:<class helloWorld> java.lang.Class helloWorld
7 facts/ClassType.facts:helloWorld
8 /* ... */

```

---

Figure 1: Facts Example



### 3. SOOT

Originally, Soot started off as a Java optimization framework. By now, researchers and practitioners from around the world use Soot to analyze, instrument, optimize and visualize Java and Android applications. Soot provides four intermediate representations for analyzing and transforming Java bytecode:

1. Baf: a streamlined representation of bytecode which is simple to manipulate.
2. Jimple: a typed 3-address intermediate representation suitable for optimization.
3. Shimple: an SSA variation of Jimple.
4. Grimp: an aggregated version of Jimple suitable for decompilation and code inspection.
5. Jimple is Soot's primary IR and most analyses are implemented on the Jimple level. Custom IRs may be added when desired.

In our case, we use Soot to generate Jimple (or Shimple) facts from Java bytecode to run a pointer or points-to analysis. [2]

#### 3.1 Applying the latest Soot version

To minimize the time consumed in fact generation phase, before we tried to parallelize it, we applied the latest (Sept. 2015) Soot version. We faced a lot of compatibility problems in order to use it properly but it gave a speed up of 40%.

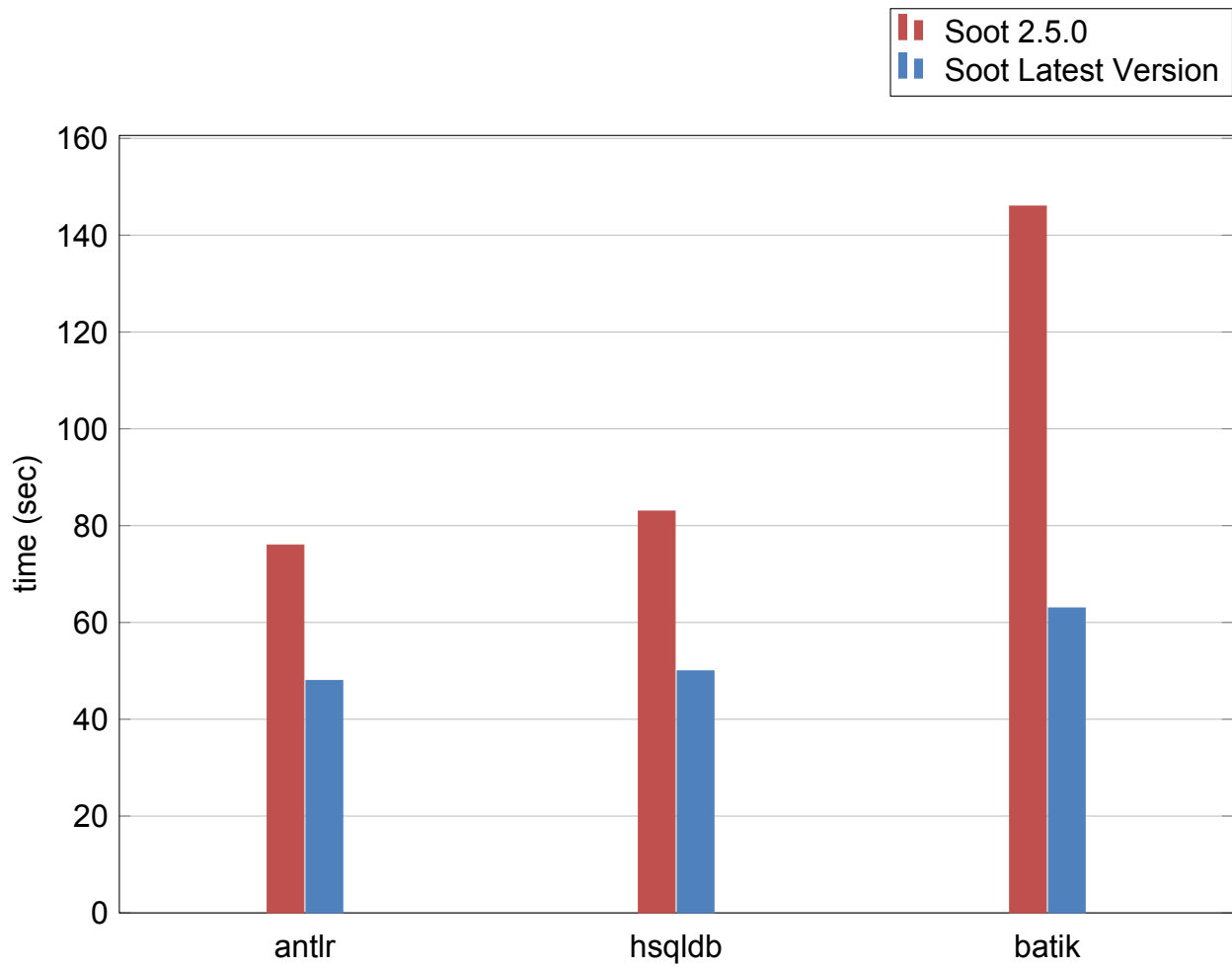


Figure 2: Fact Generation with different Soot versions

Then, to gain more speedup we tried to parallelize the fact generation part. In order to do that, we had to understand the way bytecode is translated to jimple. Below we explain in more detail this procedure.

### 3.2 Bytecode To Jimple

As we mentioned before, Soot is able to translate Java bytecode to a typed 3-address IR, Jimple. Jimple (**J**ava **s**imple) is a very convenient IR for performing optimizations, it only has 15 statements.

Soot has various phases and a lot of different options for transformations given. The one that is responsible for bytecode to jimple translation is jb phase. In this phase, first Soot translates bytecode to untyped Jimple and introduces new local variables; Jimple is stackless, Soot is using variables for stack locations. Then it infers Types to the untyped jimple. The next step is to linearize all the expressions to statements that only reference at most 3 local variables or constants.

Getting a little deeper, in a general case the way Soot handles Java bytecode classes is the following:

Soot is launched by specifying a directory with the Application code as a parameter. In

this example only Java bytecode, either a class file or a jar. First, the *main()* method of the Main class is executed and calls *Scene.loadNecessaryClasses()* (In our case Doop calls it). This method loads basic Java classes and then loads specific Application classes by calling *loadClass()*. Then, *SootResolver.resolveClass()* is called. The resolver calls *SourceLocator.getClassSource()* to fetch a reference to a *ClassSource*, an interface between the file containing the Java bytecode and Soot. In our case the class source is a *CoffiClassSource* because it is the coffi module which handles the conversion from Java bytecode to Jimple. Then, the resolver having a reference to a class source, calls *resolve()* on it. This methods in turn calls *soot.coffi.Util.resolveFromClassFile()* which creates a *SootClass* from the corresponding Java bytecode class. All source fields of Soot class methods are set to refer to a *CoffiMethodSource* object. This object is used later to get the Jimple representation of the method. For example, if during an analysis with Soot the analysis code calls *SootMethod.getActiveBody()* and the Jimple code of the method was not already generated, *getActiveBody()* will call *CoffiMethodSource.getBody()* to compute Jimple code from the Java bytecode. The Jimple code representation of the method can then be analyzed and/or transformed.

### 3.3 Compiling & Running Soot

---

```

1 ant          /* To compile */
2 ant classesjar /* To generate the sootclasses jar file */
3 ant fulljar   /* To generate the complete soot jar file */

```

---

Figure 3: Compiling Soot

---

```

1 (create a test.java)
2 javac test.java
3 java -cp ./lib/soot-trunk.jar soot.Main -f J -cp
   ./usr/lib/jvm/java-7-openjdk-amd64/jre/lib/rt.jar test

```

---

Figure 4: Generating Jimple from .class

---

```

1 java -cp soot-trunk.jar soot.Main -f J -cp ./usr/lib/jvm/java-7-openjdk-amd64/jre/lib/rt.jar
   -process-dir pathtotest.jar

```

---

Figure 5: Generating Jimple from .jar

### 3.4 Jimple Examples

Below are two simple java programs along with their jimple translation. The first one is the classic HelloWorld, and the second is a simple inheritance test that depends on the user's input. The local variables which start with a \$ sign represent stack positions and not local variables in the original program whereas those without \$ represent real local variables.

### 3.4.1 Hello World

---

```

1  public class helloWorld {
2      public static void main(String[] args) {
3          System.out.println("Hello, World");
4      }
5  }

```

---

Figure 6: HelloWorld.java

---

```

1  public class helloWorld extends java.lang.Object {
2
3      public void <init>() {
4          helloWorld r0;
5          r0 := @this: helloWorld;
6          specialinvoke r0.<java.lang.Object: void <init>()>();
7          return;
8      }
9
10     public static void main(java.lang.String []) {
11         java.lang.String [] r0;
12         java.io.PrintStream $r1;
13         r0 := @parameter0: java.lang.String[];
14         $r1 = <java.lang.System: java.io.PrintStream out>;
15         virtualinvoke $r1.<java.io.PrintStream: void println (java.lang.String)>("Hello,
16             World");
17         return;
18     }
19 }

```

---

Figure 7: HelloWorld.jimple

### 3.4.2 Inheritance Test

---

```
1  public class inheritanceTest {
2      public static void main(String[] args) {
3          testA a;
4          if (args.length < 1) {
5              a = new testA(5);
6          } else {
7              a = new testB(5);
8          }
9          int result = a.getA();
10         System.out.println("the value of a is " + result);
11     }
12
13     public static class testA {
14         int a;
15
16         public testA(int a) {
17             this.a = a;
18         }
19
20         public int getA() {
21             return this.a;
22         }
23     }
24
25     public static class testB extends testA {
26         public testB(int a) {
27             super(a+100);
28         }
29     }
30 }
```

---

Figure 8: inheritanceTest.java

---

```

1  public class inheritanceTest extends java.lang.Object {
2      public void <init>() {
3          inheritanceTest r0;
4          r0 := @this: inheritanceTest;
5          specialinvoke r0.<java.lang.Object: void <init>()>();
6          return;
7      }
8
9      public static void main(java.lang.String []) {
10         java.lang.String [] r0;
11         int $i0, i1;
12         inheritanceTest$testA $r1, r2;
13         inheritanceTest$testB $r3;
14         java.io.PrintStream $r4;
15         java.lang.StringBuilder $r5, $r6, $r7;
16         java.lang.String $r8;
17         r0 := @parameter0: java.lang.String[];
18         $i0 = lengthof r0;
19         if $i0 >= 1 goto label1;
20         $r1 = new inheritanceTest$testA;
21         specialinvoke $r1.<inheritanceTest$testA: void <init>(int)>(5);
22         r2 = $r1;
23         goto label2;
24     label1:
25         $r3 = new inheritanceTest$testB;
26         specialinvoke $r3.<inheritanceTest$testB: void <init>(int)>(5);
27         r2 = $r3;
28     label2:
29         i1 = virtualinvoke r2.<inheritanceTest$testA: int getA()>();
30         $r4 = <java.lang.System: java.io.PrintStream out>;
31         $r5 = new java.lang.StringBuilder;
32         specialinvoke $r5.<java.lang.StringBuilder: void <init>()>();
33         $r6 = virtualinvoke $r5.<java.lang.StringBuilder: java.lang.StringBuilder
           append(java.lang.String)>("the value of a is ");
34         $r7 = virtualinvoke $r6.<java.lang.StringBuilder: java.lang.StringBuilder
           append(int)>(i1);
35         $r8 = virtualinvoke $r7.<java.lang.StringBuilder: java.lang.String toString()>();
36         virtualinvoke $r4.<java.io.PrintStream: void println (java.lang.String)>($r8);
37         return;
38     }
39 }

```

---

Figure 9: inheritanceTest.jimple

## 4. FOUR APPROACHES

We now describe the basic idea of Fact Generation from Doop side. Given all the classes (sootClasses) to generate, Doop iterates each one of them; writes all the superClasses, if exist, and then generate all fields (sootFields) and methods (sootMethods). Below is presented the *FactGenerator.java* which implements the work described above and then calls Soot.

---

```

1  public class FactGenerator {
2      /* ... */
3
4      public void generate(sootClass) {
5          if (c.hasSuperclass() && !c.isInterface())
6              _writer.writeDirectSuperclass(c, c.getSuperclass());
7          for(SootField f : c.getFields())
8              generate(f);
9          for(SootMethod m : c.getMethods()) {
10              Session session = new Session();
11              generate(m, session);
12          }
13      }
14
15      public void generate(SootMethod m, Session session) {
16          /* ... */
17
18          /* This instruction spends more than 80% of FG time */
19          m.retrieveActiveBody()
20
21          /* ... */
22      }
23
24      /* ... */
25  }

```

---

**Figure 10: Sequential Fact Generation**

Having the previous basic structure in mind, and considering that *m.retrieveActiveBody()* spends more than 80% of total fact generation time, we tried to parallelize the method which calls *m.retrieveActiveBody()*. In order to do that, we approached the problem in four ways. The three of them are pretty much similar while the other one is based on a recursive Java Framework (Fork/Join Framework). Below are presented some Fact Generation time examples with the sequential FG.

| Jars    | Time (sec.) |
|---------|-------------|
| antlr   | 48          |
| eclipse | 27          |
| jython  | 32          |
| hsqldb  | 50          |
| batik   | 63          |

Table 2: Sequential Fact Generation Time Examples

#### 4.1 One Thread Per Method

Our first approach to parallelize Fact Generation is similar as the sequential one, but instead of having a loop over all Soot Methods and call *generate(m, session)*, we assign the task to a thread for each one of them. We created a new Java class, MethodGenerator, which is identical to FactGenerator and in addition has a *run()* method to generate sootMethods.

---

```

1  public class FactGenerator {
2      private ExecutorService MgExecutor = new ThreadPoolExecutor(8, 16, 0L,
        TimeUnit.MILLISECONDS, new LinkedBlockingQueue<Runnable>());
3      /* ... */
4
5      public void generate(sootClass) {
6          if (c.hasSuperclass() && !c.isInterface())
7              _writer.writeDirectSuperclass(c, c.getSuperclass());
8          for(SootField f : c.getFields())
9              generate(f);
10         for(SootMethod m : c.getMethods()) {
11             Session session = new Session();
12             Runnable mg = new MethodGenerator();
13             MgExecutor.execute(mg);
14         }
15     }
16 }
17
18 public class MethodGenerator {
19     public void run() {
20         generate(this.m, this.s)
21     }
22
23     /* ... */
24 }

```

---

Figure 11: One Thread Per Method

Below are presented some Fact Generation time examples with the *One Thread Per Method FG* approach for various thread-pool sizes (such as 4, 16 and 32).



| <b>Jars</b>      | <b>Time (sec.)</b> |           |           |
|------------------|--------------------|-----------|-----------|
| <b>Pool Size</b> | <b>4</b>           | <b>16</b> | <b>32</b> |
| antlr            | 21                 | 14        | 13        |
| eclipse          | 13                 | 7         | 8         |
| jython           | 14                 | 9         | 9         |
| hsqldb           | 23                 | 15        | 16        |
| batik            | 26                 | 23        | 18        |

Table 3: One Thread Per Method Time Examples, with pool size: 4, 16, 32

## 4.2 One Thread Per Class

In our second approach, we tried to find out ways to gain more speedup. So, we observed that some threads did not have much work to do and finishing their task instantly. Allocating a new object and assigning to it a task just to finish instantly was an overhead. As a result, we tried to feed the threads more than just a method, so we created a new thread for each class not for each method.

```

1  public class FactGenerator {
2      private ExecutorService CgExecutor = new ThreadPoolExecutor(8, 16, 0L,
3          TimeUnit.MILLISECONDS, new LinkedBlockingQueue<Runnable>());
4          /* ... */
5      public void generate(sootClass) {
6          Runnable cg = new ClassGenerator();
7          CgExecutor.execute(cg);
8      }
9  }
10
11 public class ClassGenerator {
12     public void run() {
13         if (c.hasSuperclass() && !c.isInterface())
14             _writer.writeDirectSuperclass(c, c.getSuperclass());
15         for(SootField f : c.getFields())
16             generate(f);
17         for(SootMethod m : c.getMethods()) {
18             Session session = new Session();
19             Runnable mg = new MethodGenerator();
20             MgExecutor.execute(mg);
21             generate(m, session);
22         }
23     }
24     /* ... */
25 }

```

Figure 12: One Thread Per Class

The results were slightly better than the previous but without achieving a remarkable speedup.

Below are presented some Fact Generation time examples with the *One Thread Per Class FG* approach for various thread-pool sizes (such as 4, 16 and 32).

| <b>Jars</b>      | <b>Time (sec.)</b> |           |           |
|------------------|--------------------|-----------|-----------|
| <b>Pool Size</b> | <b>4</b>           | <b>16</b> | <b>32</b> |
| antlr            | 21                 | 14        | 13        |
| eclipse          | 13                 | 7         | 8         |
| jython           | 14                 | 9         | 9         |
| hsqldb           | 23                 | 15        | 16        |
| batik            | 26                 | 23        | 18        |

**Table 4: One Thread Per Class Time Examples, with pool size: 4, 16, 32**

### 4.3 Fork-Join Framework

So in order to achieve more speedup we tried a completely different approach than the two previous, a recursive Java Framework (Fork/Join framework). As Oracle describes in Java Documentation, Fork/Join Framework is *"an implementation of the ExecutorService interface that helps you take advantage of multiple processors. It is designed for work that can be broken into smaller pieces recursively. The goal is to use all the available processing power to enhance the performance of your application."*

*The center of the fork/join framework is the ForkJoinPool class, an extension of the AbstractExecutorService class. ForkJoinPool implements the core work-stealing algorithm and can execute ForkJoinTask processes.*

*The idea of using the fork/join framework is to write code that performs a segment of the work. The basic structure should be like the following pseudocode."* [3]

---

```

1  if (my portion of the work is small enough) {
2      do the work directly
3  } else {
4      split my work into two pieces
5      invoke the two pieces and wait for the results
6  }
```

---

**Figure 13: Fork-Join Basic-Use**

---

```

1  public class FactGenerator {
2      private ForkJoinPool classGeneratorPool = new ForkJoinPool();
3      /* ... */
4      public void generate(sootClass) {
5          if (c.hasSuperclass() && !c.isInterface())
6              _writer.writeDirectSuperclass(c, c.getSuperclass());
7          for(SootClass i : c.getInterfaces())
8              _writer.writeDirectSuperinterface(c, i);
9          for(SootField f : c.getFields())
10             generate(f);
11         if (c.getMethods().size() > 0) {
12             ClassGenerator classGenerator = new ClassGenerator(_writer, _ssa, c, 0,
13                 c.getMethods().size());
14             classGeneratorPool.invoke(classGenerator);
15         }
16     }
17 }
18 public class ClassGenerator {
19     /* ... */
20     public void compute() {
21         List<SootMethod> sootMethods = _sootClass.getMethods();
22         /* if (my portion of the work is small enough) */
23         if (_to - _from < threshold) { /* How many classes can I process? */
24             for (int i = _from ; i < _to ; i++) {
25                 SootMethod m = sootMethods.get(i);
26                 Session session = new Session();
27                 generate(m, session);
28             }
29         } else { /* split work*/
30             int half = (_to - _from)/2;
31             ClassGenerator c1 = new ClassGenerator(_writer, _ssa, _sootClass, _from, _from
32                 + half);
33             ClassGenerator c2 = new ClassGenerator(_writer, _ssa, _sootClass, _from + half,
34                 _to);
35             invokeAll(c1, c2);
36         }
37     }
38     /* ... */
39 }

```

---

Figure 14: Fork-Join Framework

The results were worse than the two previous approaches (still better than then sequential approach). Below are presented some Fact Generation time examples with the *Fork/Join Framework FG* approach for threshold values (such as 2, 3 and 4). Threshold actually is the number of classes for a thread to process.

| <b>Jars</b>                            | <b>Time (sec.)</b> |          |          |
|--|--------------------|----------|----------|
| <b>Threshold (classes to generate)</b> | <b>2</b>           | <b>3</b> | <b>4</b> |
| antlr                                  | 23                 | 25       | 25       |
| eclipse                                | 13                 | 15       | 18       |
| jython                                 | 16                 | 17       | 19       |
| hsqldb                                 | 25                 | 28       | 31       |
| batik                                  | 34                 | 37       | 38       |

Table 5: Fork/Join Time Examples, with threshold 2, 3, 4 and pool size 16

#### 4.4 Multiple Classes Per Thread

Our last and final approach is similar as the second one, but instead of having one thread per class, we now have one thread per multiple classes. Even in the second approach some threads did not have much work to do. Below is presented in an abstract way the final stage of the code implementing the *Multiple Classes Per Thread* approach.

---

```

1  public class Driver {
2      public Driver(ThreadFactory factory, boolean ssa, int totalClasses) {
3          _factory = factory;
4          _ssa = ssa;
5          _classCounter = 0;
6          _sootClasses = new ArrayList<>();
7          _totalClasses = totalClasses;
8          _cores = Runtime.getRuntime().availableProcessors();
9          _executor = new ThreadPoolExecutor(_cores/2, _cores, 0L, TimeUnit.MILLISECONDS,
              new LinkedBlockingQueue<Runnable>());
10     }
11
12     public void doInParallel( List<SootClass> sootClasses) {
13         for(SootClass c : sootClasses)
14             generate(c);
15         _executor.shutdown();
16         _executor.awaitTermination(Long.MAX_VALUE, TimeUnit.NANOSECONDS);
17     }
18
19     void generate(SootClass _sootClass) {
20         _classCounter++;
21         _sootClasses.add(_sootClass);
22         if (( _classCounter % _classSplit == 0) || ( _classCounter + _classSplit-1 >=
23             _totalClasses)) {
24             Runnable runnable = _factory.newRunnable(_sootClasses);
25             _executor.execute(runnable);
26             _sootClasses = new ArrayList<>();
27         }
28     }
29
30     public class ThreadFactory {
31         /* ... */
32         public Runnable newRunnable(List<SootClass> sootClasses) {
33             if (_makeClassGenerator)
34                 return new FactGenerator(_factWriter, _ssa, sootClasses);
35             else
36                 return new FactPrinter(_ssa, _toStdout, _outputDir, _printWriter, sootClasses);
37         }
38     }

```

---

Figure 15: Multiple Classes Per Thread

Getting away from the overhead produced by assignments and allocations of the first and second approach, this one had so far the best time results. Below are presented some Fact Generation time examples with the *Multiple Classes Per Thread FG* approach for various thread-pool sizes (such as 4, 16 and 32) and various number of classes per thread.

| Jars    | Classes Per Thread<br>Pool Size | Time (sec.) |    |    |
|---------|---------------------------------|-------------|----|----|
|         |                                 | 2           | 3  | 4  |
| antlr   | 4                               | 22          | 18 | 19 |
|         | 16                              | 13          | 12 | 14 |
|         | 32                              | 14          | 13 | 13 |
| eclipse | 4                               | 12          | 10 | 11 |
|         | 16                              | 7           | 8  | 6  |
|         | 32                              | 8           | 8  | 8  |
| jython  | 4                               | 13          | 14 | 13 |
|         | 16                              | 11          | 7  | 9  |
|         | 32                              | 9           | 8  | 8  |
| hsqldb  | 4                               | 25          | 22 | 20 |
|         | 16                              | 17          | 14 | 16 |
|         | 32                              | 16          | 18 | 14 |
| batik   | 4                               | 23          | 24 | 25 |
|         | 16                              | 22          | 20 | 17 |
|         | 32                              | 21          | 17 | 17 |

**Table 6: Multiple Classes Per Thread Time Examples, with pool size: 4, 16, 32 and classes per thread: 2, 3, 4**

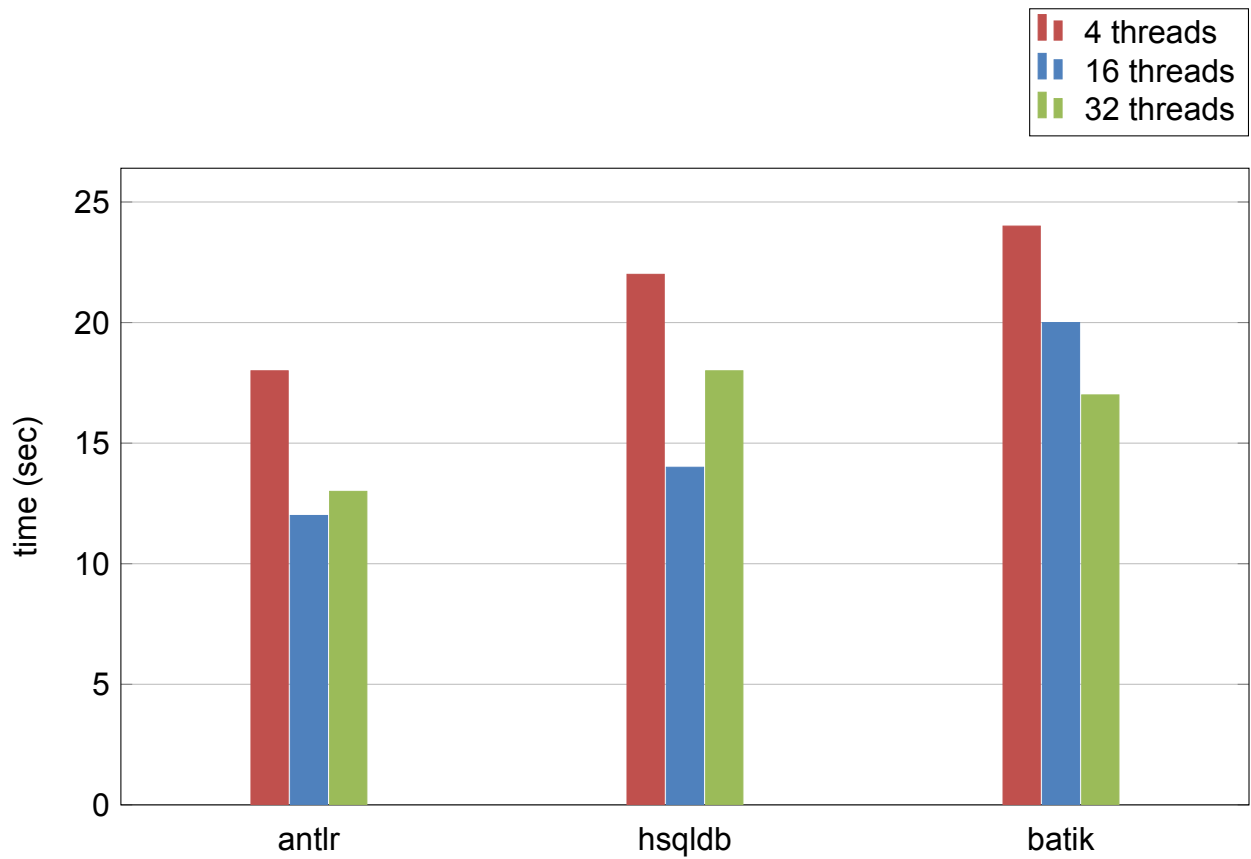


Figure 16: Increasing the number of threads with 3 classes per thread

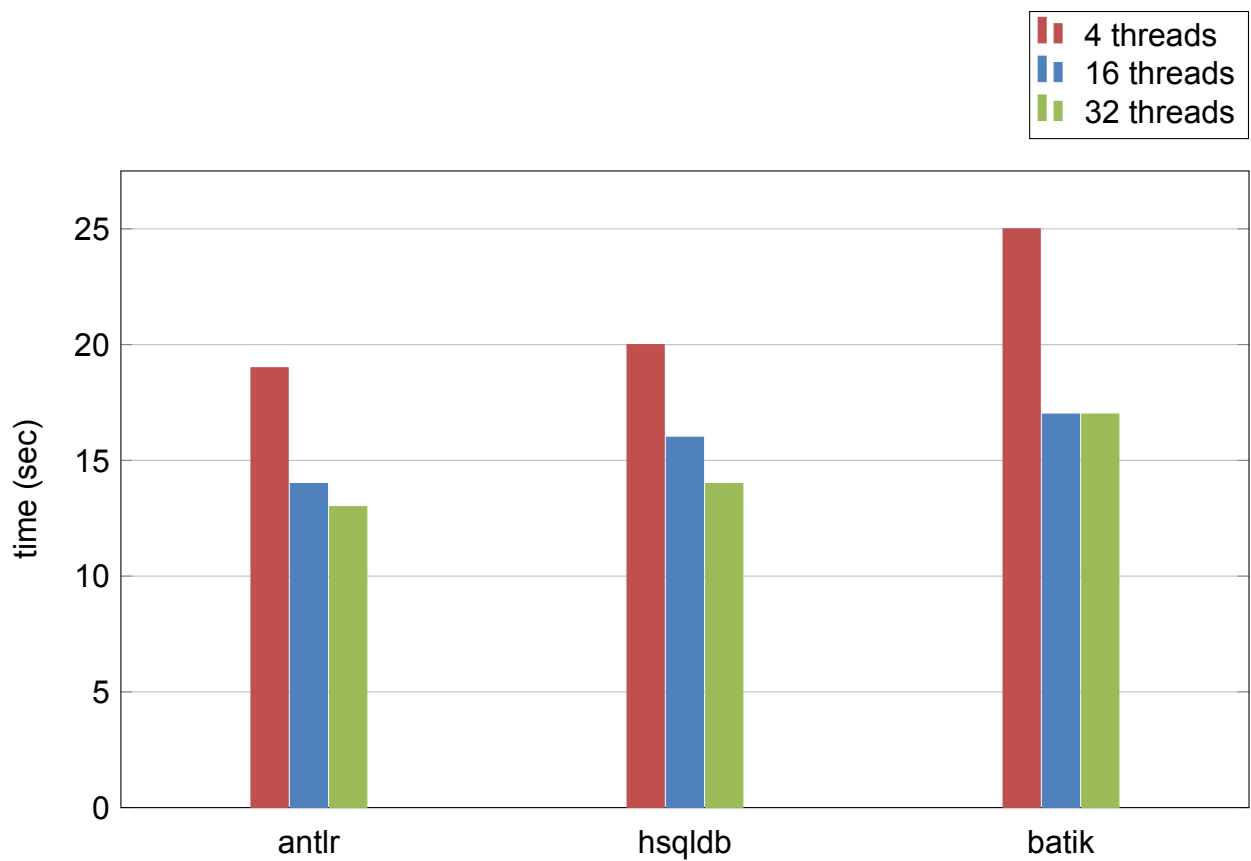


Figure 17: Increasing the number of threads with 4 classes per thread

## 5. LOCKING

Along with threads come locks. An incorrect use of locking could have very negative(??) effects on the results. Having more locks than we actually needed would have led to a time result as slow as the sequential one. Having less, would have led to races and deadlocks. So, locks should be used very consciously.

Luckily, in our case we used very few locks in both Doop and Soot. Below we explain in more detail.

### 5.1 Doop Side

The only two things we had to lock to prevent races and deadlocks were an access to the output file and three methods that accessing a sootMethod.

#### 5.1.1 CSVDatabase

Lock file to prevent more than one thread to write at the same time.

---

```

1 synchronized(predicateFile) {
2     Writer writer = getWriter(predicateFile);
3     addColumn(writer, arg, shouldTruncate);
4     for (Column col : args)
5         addColumn(writer.append(SEP), col, shouldTruncate);
6     writer.write(EOL);
7 }

```

---

Figure 18: CSVDatabase.java

#### 5.1.2 Representation

The other synchronization we had to provide was in three methods that were accessing SootMethods while threads were active.

---

```

1 public synchronized String signature(SootMethod) { /*...*/ }
2 public synchronized String handler(SootMethod, Trap, Session) { /*...*/ }
3 public synchronized String compactMethod(SootMethod) { /*...*/ }

```

---

Figure 19: Representation.java

### 5.2 Soot Side

The way Soot is implemented, it has a class that encloses all global objects (*G.java*). As we mentioned before, Soot does much more than just translating bytecode to jimple, so it has various phases and a lot of different options for transformations given. The phase that translates bytecode to jimple is jb phase.



### 5.2.1 Type Assigner

The class *JimpleBodyPack* applies the transformations corresponding to the given options. In our case, it applies a *"jb.tr"* which means *"jimple body transformation"*. From bytecode to jimple translation this is the only pack needed, so we lock before applying Type Assigner and unlock afterwards.

---

```

1 lock.lock();
2 PackManager.v().getTransform("jb.tr").apply(b);
3 lock.unlock();

```

---

Figure 20: *JimpleBodyPack.java*

### 5.2.2 Pack Manager

The class that manages the Packs containing the various phases and their options is *PackManager*. Lock before retrieving Class Hierarchy Analysis, unlock afterwards.

---

```

1 lock.lock();
2 p.add(new Transform("cg.cha", CHATransformer.v()));
3 p.add(new Transform("cg.spark", SparkTransformer.v()));
4 p.add(new Transform("cg.paddle", PaddleHook.v()));
5 lock.unlock();

```

---

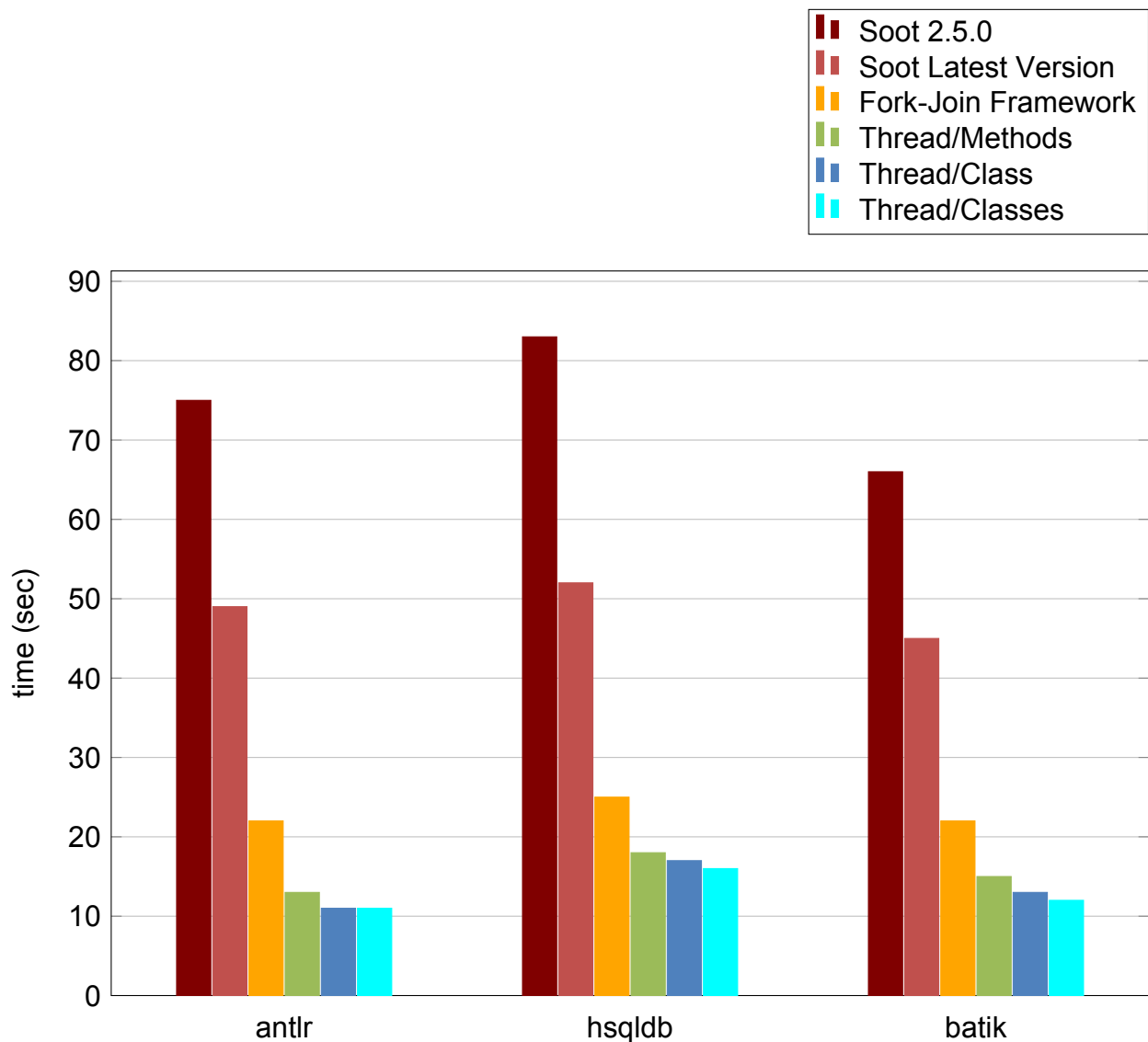
Figure 21: *PackManager.java*

### 5.2.3 Shimple -ssa

We also use Soot to provide **Shimple** which is an **SSA** variation of Jimple. Shimple generation is very similar to jimple but with a few differences. blahblabh prepei na to dw kiallo auto The class that handles the translation from jimple to shimple is *Shimple.java*. So far we have synchronized all Shimple-Body-creation methods.

## 6. TIME RESULTS

Below are presented fact generation times for Soot 2.5.0, latest Soot version and all our approaches.



**Figure 22: Fact Generation Time Results**

In conclusion, just changing the soot version and applying the latest one gave a speedup up to 30-40%. With the latest soot-version and our best approach we achieved a speedup of 60-80%. Our third attempt (Fork/Join Framework) was not a successful one, it was better than the sequential fact generation (obviously) but not as good as the other three.

| <b>Jars</b>     | <b>Time (sec.)</b> |                  |                      |                     |                       |
|-----------------|--------------------|------------------|----------------------|---------------------|-----------------------|
| <b>Approach</b> | <b>Sequential</b>  | <b>Fork/Join</b> | <b>Thread/Method</b> | <b>Thread/Class</b> | <b>Thread/Classes</b> |
| antlr           | 48                 | 23               | 13                   | 13                  | 12                    |
| eclipse         | 27                 | 13               | 7                    | 7                   | 6                     |
| jython          | 32                 | 16               | 8                    | 8                   | 7                     |
| hsqldb          | 50                 | 25               | 15                   | 14                  | 14                    |
| batik           | 63                 | 34               | 18                   | 17                  | 17                    |

**Table 7: Summarizing best times of all approaches with pool size 16-32**

## ACRONYMS AND ABBREVIATIONS

|         |                             |
|---------|-----------------------------|
| IR      | Intermediate Representation |
| SSA     | Static Single Assignment    |
| Jimple  | Soot typed 3-address IR     |
| Shimple | An SSA-version of Jimple    |
| FG      | Fact Generation             |
| jb      | Jimple Body                 |

## REFERENCES

- [1] "Doop: Framework for Java Pointer Analysis" [Online] Available: <http://doop.program-analysis.org/>
- [2] "Sable: Soot" [Online] Available: <https://sable.github.io/soot/>
- [3] "Oracle Java Fork/Join Framework" [Online] Available: <https://docs.oracle.com/javase/tutorial/essential/concurrency/forkjoin.html>