可视化的 Java 多线程程序错误定位工具

Visual Fault Localization Tool for Java Multi-threaded Program

黄亚铭 指导老师: 孙玉霞

暨南大学计算机系

暨南大学本科生毕业设计(论文)答辩

2013.5.17



并行程序的难点

● 编写并行程序比编写串行程序更加困难 (死锁、数据争用...)



并行程序的难点

- 编写并行程序比编写串行程序更加困难(死锁、数据争用...)
- ② 测试并行程序比测试串行程序更加困难
 - 难以重现错误
 - 不易定位



Falcon

- Fault Localization in Concurrent Programs
- Sangmin Park, et al.



错误类型的划分

数据访问模式 (Data Access Pattern)

如果在并发程序运行的过程中,不同线程访问共享数据的方式符合这些数据访问模式,就有出现并发错误的可能性。



错误类型的划分

数据访问模式 (Data Access Pattern)

如果在并发程序运行的过程中,不同线程访问共享数据的方式符合这些数据访问模式,就有出现并发错误的可能性。

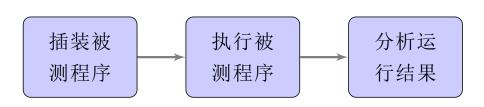
顺序破坏(Order Violation)

访问一个共享变量的两个线程(其中一个线程进行写操作)因为访问顺序交错导致读、写错误的数据。

原子性破坏(Atomicity Violation)

一个线程操作共享变量的过程中,另外一个线程操作被错误的插 入。

Falcon 工作流程





插装被测程序

插装的难点:

- 如何判断变量是否被多个线程访问?
- ② 如何判断变量是读还是写?
- ◎ 如何插入探针?



插装被测程序

插装的难点:

- 如何判断变量是否被多个线程访问?
- ② 如何判断变量是读还是写?
- ◎ 如何插入探针?

解决方法: Soot

- 使用 Soot 进行线程逃逸分析
- ② 使用 Jimple 分析语句中的变量
- 使用 Soot API 插入探针,并生成.class 文件



运行被测程序

多次运行经过插装的被测程序。



运行被测程序

多次运行经过插装的被测程序。

- 识别数据访问模式
- 记录数据访问模式和运行结果 (序列化为 XML 文件保存)



统计运行结果

Jaccard 系数

对于一个数据访问模式 p, 有通过的执行数 pass(p), 失败的执行数 failed(p) 和该被测程序所有测试失败的个数 totalfalied, 可以得到可疑度:

$$suspiciousness(p) = \frac{failed(p)}{totalfailed + passed(p)}$$



对 Falcon 的补充和改进





9 / 11

对 Falcon 的补充和改进

- 补充代码
- 2 完善算法



对 Falcon 的补充和改进

- 补充代码
- ② 完善算法
- 3 重构



对 Falcon 的补充和改进

- 补充代码
- ② 完善算法
- 3 重构
- 可视化



效果截图

顺序破坏





效果截图

原子性破坏

