

## Group Meeting

Members: Yong Li, Depeng Liu, Weizhi Feng, Xie Li, Shizhen Yu, Yutian Zhu,  
Zongxin Liu

2021 年 8 月 11 日

## 差分隐私——刘德鹏

这周:

- 连续机制建模: 阅读强化学习 5/6/7 章节, 讨论了转移到差分隐私中建模及遇到的难点: 连续空间 event 转化为 action 的处理, reward 分配等;
- Pufferfish 文章: 工具比较需更新, 验证工具 (CheckDP, 20 'CCS) 不能直接适用于离散噪声场景, 适用于黑盒场景的基于机器学习工具 (DP-Sniper, 21 'SP) 应该会适用, 在读文章 (DP-Sniper: Black-Box Discovery of Differential Privacy Violations using Classifiers, 21 'SP), 其根据 DP 机制的输入、输出训练分类器, 拟合给定输出在输入下的后验分布, 来寻找最可能违反 DP 的反例。
- 参数选取: 调研了最早的一篇 How much is enough? choosing  $\epsilon$  for differential privacy(11' ISC) 文章, 相同大小的  $\epsilon$  对于不同机制、数据域、查询的影响不同, 表现为推断真实数据的分布的影响, 是一个相对意义的参数, 其选取比设计隐私机制甚至更有挑战。存在局限性为没有正确认清背景知识对差分隐私的保护削弱影响, 只能在小案例上分析等。

计划:

- 强化学习教材阅读, 考虑最简单连续机制的建模;
- 精读 21 'SP 的文章, 调研工具使用;
- 参数选取更多文章调研。

上周计划（进行中）：

- 根据 SV-COMP 加入库函数语义的支持。
- 代码的重构。
- 继续跑例子进行测试和 Debug。

SV-COMP:

选取用例：memsafety 文件夹共 38 个例子

在 38 个例子中：

- Successfully verified:  $2 \rightarrow 10$
- Unmatched result:  $2 \rightarrow 3$
- Exceptions raised:  $34 \rightarrow 25$

其中 Unmatched result 中，目前有 1 个误报，2 个漏报。

TODOs:

- ① 寻找误报漏报原因，特别是误报，可能是因为语义的 bug.
- ② 继续添加对 SV-COMP 的支持和 Debug.
- ③ **优化问题**：部分例子运行时间过长，50s 左右
- ④ 给出更具体的内存安全问题的划分，目前只是给出 UNSAFE 结果。

- 期刊文章：

- 上周计划是将 SDBA 取补算法正确性证明写完，然后增加一个 determinization (确定化，给定 SDBA，构建一个 DRA) 的章节，目前完成进度：SDBA 取补算法正确性写完了；determinization 构造写完了。

- 阅读

- 看之前 CAV 的一篇文章 decoupled search 做 composed Büchi automata 的 liveness verification，还没看完，计划下下周三组会的时候报告。

- 计划

- 期刊文章根据实验情况写实验部分。
- 读 CAV 文章。
- 读程序分析教材 3.1 3.2 节，根据硬件验证小组要求读相关内容：chisel 验证；或者 handbook of modelchecking 相关章节。

- 调研了 VeriSolid 的建模模块, 即 FSolidM。
  - 函数体的语句不能自动生成, 需要手动输入。
  - SolidM 对于 constructor 和 fallback 不能很好地支持。
  - 不支持多合约交互建模。
- VerX: Safety Verification of Smart Contracts (读了一半左右)
- 下周计划:
  - 继续调研 VeriSolid, 考察 FSolidM 的建模功能和阅读对应源码。
  - 准备讨论班。
  - 有时间就熟悉一下 smack 项目。

进展：

- 上周读了《Learning Deterministic Automata on Infinite Words》，跟李老师周末大概讲了讲主要的思路，文章中的一些证明细节也基本上弄清楚了。
- 这周在读《Boolean Satisfiability Solvers and Their Applications in Model Checking》，现在大概通读完一遍，再整理整理，明天做一下汇报。