

数据库事务一致性高效验证技术研究

(软件学院聘期考核报告; 2023 年 01 月 ~ 2023 年 12 月)

魏恒峰

hfwei@nju.edu.cn

2023 年 11 月 22 日





学期	课程	学分	课时
2023 年春季	编译原理 (1 班)	3	54
2023 年暑期	大语言模型原理与应用	1	2
2023 年秋季	C 语言程序设计基础 (1 班)	2	36
2023 年秋季	C 语言程序设计基础 (2 班)	2	36



COMPILER

$$(\underbrace{200}_{\text{软件学院}} + \underbrace{88}_{\text{跨专业选修}}) + (\underbrace{98 + 95 + 87 + 90 + 86 + 90}_{\text{技术科学试验班}}) + \underbrace{32}_{\text{苏州校区重修班}} = 866$$



$$(\underbrace{8}_{\text{软件学院}} + (\underbrace{3 \times 6}_{\text{技术科学试验班}})) + \underbrace{1}_{\text{苏州校区重修班}} = 27 \text{ 名助教}$$

10 月 29 日, 已顺利完成第一次机考

评价指标

软件学院

技术科学试验班

本课程的分构成成为:

- **平时练习** (10%): 基本每周一次;
- **阶段机试** (15% + 20%): 学期中安排两次阶段性机试, 主要考察平时练习的掌握程度;
- **课程项目** (25%): 指选 + 自选题目, 学期期末项目 (很可能会作为 **寒假作业**);
- **期末机试** (30%): 和平时编程练习与阶段性机试的形式相同, 没有笔试。

定于 12 月 09 日, 第二次机考

每周安排 9 次答疑

！ 本学期, 如果你的代码风格很糟糕, 助教有权拒绝回答相关问题。

软件学院	技术科学试验班 1/2/3 班				技术科学试验班 4/5/6 班				答疑调查问卷统计数据
学院	软件学院 1 班	软件学院 2 班	软件学院 3 班	软件学院 4 班	软件学院 5 班	软件学院 6 班	软件学院 7 班	软件学院 8 班	
周一上午	4 (5.9%)	2 (2.6%)	2 (4.0%)	0 (0.0%)	2 (4.7%)	2 (4.4%)	5 (11.2%)	4 (9.1%)	
周一下午	7 (8.8%)	6 (6.7%)	9 (18.0%)	9 (17.6%)	6 (14.0%)	36 (40.0%)	2 (22.2%)	4 (9.1%)	
周二上午	5 (6.4%)	13 (14.4%)	10 (20.0%)	10 (18.2%)	10 (22.2%)	10 (11.1%)	10 (22.2%)	10 (22.2%)	
周二下午	6 (8.2%)	4 (4.4%)	5 (10.0%)	4 (7.6%)	4 (9.1%)	7 (7.8%)	5 (11.1%)	1 (2.2%)	
周三上午	9 (12.3%)	12 (13.3%)	2 (4.0%)	1 (2.0%)	4 (9.1%)	3 (6.7%)	3 (6.7%)	1 (2.2%)	
周三下午	6 (7.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	4 (9.1%)	
周四上午	6 (8.2%)	12 (13.3%)	0 (0.0%)	0 (0.0%)	4 (9.1%)	2 (4.4%)	7 (15.6%)	4 (9.1%)	
周四下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周一上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周一下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周二上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周二下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周三上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周三下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周四上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周四下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周一上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周一下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周二上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周二下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周三上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周三下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周四上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周四下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周五下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周六下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日上午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	
周日下午	5 (6.7%)	5 (5.6%)	5 (10.0%)	5 (9.1%)	5 (11.1%)	5 (5.6%)	5 (11.1%)	1 (2.2%)	

答疑收集表

答疑收集表									
2023CPL答疑收集表 Users									
ID	提问同学	类型	作业题编号	代码提交编号	知识点	问题	图片	文件	操作
47	0058 郭阿青 231250066	题目疑问-作业				812246 请问我按您给的提示将每一种情况都写了出来，但是时间超限了，遇到...			
48	0059 潘启华 221830040	题目疑问-作业				第七次作业，内存分配器，812960 使用链表，但是运行超时			
49	0060 廖云彪 231880346	题目疑问-作业	6-F			第六次作业三角形 请问为什么813652号代码的三角形本就没问题在oj上会偏移			
50	0061 刘敬东 231880324	题目疑问-作业	7-A			第七次作业A栈模拟题 总想30分答案错误 毫无思路 自己试了很多都没问题，提...			
51	0062 唐国川 231880521	题目疑问-作业				第七次作业括号序列判断题，按照书上给的那段代码，栈不久类似于数组吗，...			
52	0063 赵安睿 231200037	题目疑问-作业				请问第六次作业-最大区间中814295号作业作为什么运行错误啊（我也不知道通...			
53	0064 隋治平 231880255	题目疑问-作业				第七次作业，内存分配器，本地样例通过，显示运行错误，提交编号812563打...			
54	0065 李政吉 231830185	题目疑问-作业	6-E	814756	基本数据类型	老师好，第六次作业E题和分时间超限80分，如何优化呢？以及，为什么我的测...			
55	0066 刘敬宇 231870033	题目疑问-作业		815153		Submission 815153 第七次作业的brackets.c,请问有什么办法提高程序运行效率...			
56	0067 刘敬宇 231870033	题目疑问-机试				Submission 815197 请问第一次机试的三只小猪题目的代码为什么只能拿到20...			
57	0068 唐国川 231880521	题目疑问-作业			输入/输出	我的c语言代码输入的是 () ，也就是两个空格加一个括弧，为什么存入数组里...			
58	0069 丁宏业 231880397	题目疑问-作业	错题			变量mid是全局变量，但在自定义函数中加上double再定义一查可以满分，不...			
59	0070 赵训强 231880509	题目疑问-作业	6-E			助教好！孩子第6次作业积分答题一直20分改卷查看在于干题0分嘛！只能麻...			
60	0071 丁宏业 231880397	题目疑问-作业			递归	变量mid是全局变量，但在自定义函数中加上double再定义一查可以满分，不...			
61	0072 庄祺鑫 231880233	其他			代新调试	一般样例给的少的题目，而且样例不怎么好想题目怎么debug（抓狂）			
62	0074 黄智航 231880430	题目疑问-作业				我做题问三角形（triangle.c）题目中，如果二维数组没有初始化输出时为什么不...			
63	0075 丁慧阳 231880185	题目疑问-作业	6-C			7题 C语言序列 代码编号818498 代码在字符串数组开到10 48时运行报错，10 58时...			
64	0076 杨欣月 231880442	题目疑问-作业	其他		指针	关于指针的知识掌握的感觉很混乱。字符型指针的应用和二维指针不太理解，...			
65	0077 姚东申 231250187	题目疑问-作业				第七次作业的挂板刷题 我的代码试了很多样例都没问题，请问是哪个地方没考...			
66	0078 蔡宇斯 231250139	题目疑问-作业			函数	可以在自己写的函数里面直接调用其他自己写的函数吗？			

合集 | 18个视频 | 11-17更新

默认排序

升序排序

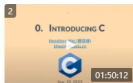
编辑



0-intro-Class2-20230915

▶ 2026

9-15



0-intro-20230915-Class1

▶ 1330

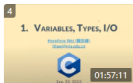
9-15



1-types-io-Class2-20230922

▶ 2017

9-22



1-types-io-Class1-20230922

▶ 1271

9-22



2-if-for-array-Class2-20231008

▶ 1096

10-8



2-if-for-array-Class1-20231008

▶ 595

10-8



3-for-a-while-Class2-20231013

▶ 1556

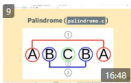
10-13



3-for-a-while-Class1-20231013

▶ 542

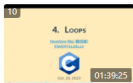
10-13



3-for-a-while-palindrome-20231016

▶ 945

10-16



4-loops-Class2-20231020

▶ 1188

10-20



4-loops-Class1-20231020 (剪辑版)

▶ 579

10-25



5-function-Class2-20231027

▶ 598

10-27



5-function-Class1-20231027

▶ 1128

10-27



第一次机试说明-Class2-20231027-李薛成

▶ 1015

10-27



第一次机试说明-Class1-20231027

▶ 681

10-27



6-recursion-Class2-20231110

▶ 904

11-10



6-recursion-Class1-20231110

▶ 597

11-12



7-data-types-Class2-20231117

我的合集和视频列表 > 合集CPL 视频教程

播放全部

合集 | 14个视频 | 11-14更新

默认排序

升序排序

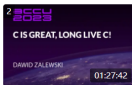
编辑



CLion 调试器使用方法

4485

2022-11-10



Programming in Modern C with a Sneak Peek into C23 (ACCU)

573

7-27



New Features in C (Dan Saks; 2019)

278

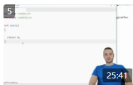
7-27



Modern C and What We Can Learn from it (emBO++ 21 Luca)

574

9-5



C Programming In One Video

563

9-8



20231008-C Code Style 同样是 C 语言, 你的代码怎么这么丑?

1224

10-8



20231009-for-CLion-Debug (今天你又 Bug 了吗?)

1377

10-9



20231014-scanf-indeterminate (你尽管写 bug, ChatGPT 会出)

1432

10-14



20231018-timing C 为您省时?

999

10-19



20231022-VSCode调试方法与 VSCode常用快捷键

2040

10-22



Let's Build a Computer in Conway's GAME OF LIFE (带中英)

403

10-25



20231111-static-local-variables (静态局部变量究竟是个什么东西?)

732

11-12



20231112-EAP 这样是不是就能合法地永久试用 JetBrains 产品了?

2829

11-12



20231114-C 语言学习资源-视频教程 (C 语言, 你入门了吗? 如入! 视)

903

11-14

2023 春季,《编译原理》由选修课改为专业必修课。



本学期: 作业 (0 分) + 实验 (75 分) + 期末测试 (25 分)

作业 (15 分) + 实验 (45 分) + 期末测试 (40 分)

实验分数高, 导致今年的高分段人数较多。

下学期考虑调整。

12 (11-llvm-ir)	2023-04-05 (周五)	LLVM IR 简介	LLVM IR, LLVM Java API
12 (12-ir-expr)	2023-05-06 (周六)	表达式的中间代码生成	LLVM IR, 表达式翻译
13 (13-ir-control (1))	2023-05-10 (周三)	控制流的中间代码生成 (方案一)	LLVM IR, 控制流翻译
14 (14-ir-control (2))	2023-05-17 (周三)	控制流的中间代码生成 (方案二)	LLVM IR, 控制流翻译
14 (15-ir-backpatch)	2023-05-19 (周五)	控制流的中间代码生成 (回填技术)	为什么需要回填技术?
15 (16-parser-lr0)	2023-05-24 (周三)	LR(0)、SLR	D4.5、D4.6
16 (17-parser-lr1)	2023-05-31 (周三)	LR(1)、LALR(1)	D4.7
16 (18-codegen-riscv)	2023-06-02 (周五)	RISC-V 程序设计	

计划编写《编译原理》课程讲义

逐步对外开放《C 语言程序设计基础》与《编译原理》课程资源
提升课程影响力

研究背景: 分布式系统

分布式系统应用广泛



(a) Google Docs



(b) Apache Wave



(c) Wikipedia



(d) L^AT_EX Editor



Figure 微信与分布式存储系统

Figure 协同文本编辑系统

分布式系统通常采用“数据副本”技术提高容错性与可用性

研究主题: 分布数据一致性

“数据副本”技术带来了数据一致性问题

研究问题丰富:

规约、实现、度量、验证、编程模型

博士论文工作偏重于“实现、度量”

研究工作: 分布数据一致性的形式化规约与验证

入职后, 研究重心有所调整:
近三年工作偏重“规约、验证”

工作特色: 使用形式化方法追求真实系统、重要协议的正确性

研究工作: 分布数据一致性的形式化规约与验证

这代表了学术界与工业界的一种共同趋势



Figure TLA⁺ 形式化规约语言 (由 Leslie Lamport 开发)

Engineers use TLA+ to prevent serious but subtle bugs from reaching production.

BY CHRIS NEWCOMBE, TIM RATH, FAN ZHANG, BOGDAN MUNTEANU, MARC BROOKER, AND MICHAEL DEARDEUFF

How Amazon Web Services Uses Formal Methods

Figure [Amazon: CACM2015]@CACM

*“At Amazon, formal methods are **routinely** applied to the design of **complex real-world software**, including public cloud services.”*

研究工作: 分布数据一致性的形式化规约与验证

这代表了学术界与工业界的一种共同趋势



Figure TLA⁺ 形式化规约语言 (由 Leslie Lamport 开发)

Engineers use TLA+ to prevent serious but subtle bugs from reaching production.

BY CHRIS NEWCOMBE, TIM RATH, FAN ZHANG, BOGDAN MUNTEANU, MARC BROOKER, AND MICHAEL DEARDEUFF

How Amazon Web Services Uses Formal Methods

Figure [Amazon: CACM2015]@CACM

*“Formal methods are **surprisingly feasible** for mainstream software development and **give good return on investment**.”*

研究工作: 分布数据一致性的形式化规约与验证

这代表了学术界与工业界的一种共同趋势



Figure TLA⁺ 形式化规约语言 (由 Leslie Lamport 开发)

Engineers use TLA+ to prevent serious but subtle bugs from reaching production.

BY CHRIS NEWCOMBE, TIM RATH, FAN ZHANG, BOGDAN MUNTEANU, MARC BROOKER, AND MICHAEL DEARDEUFF

How Amazon Web Services Uses Formal Methods

Figure [Amazon: CACM2015] @CACM

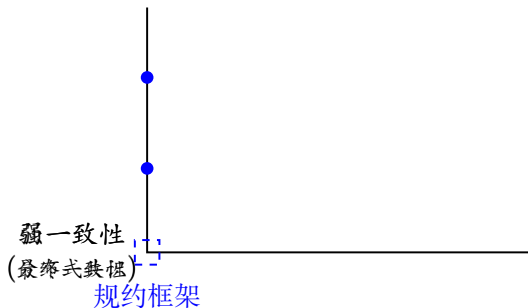
*“Formal methods find **bugs** in system designs that **cannot be found** through any other technique we know of.”*

研究工作: 分布数据一致性的形式化规约与验证



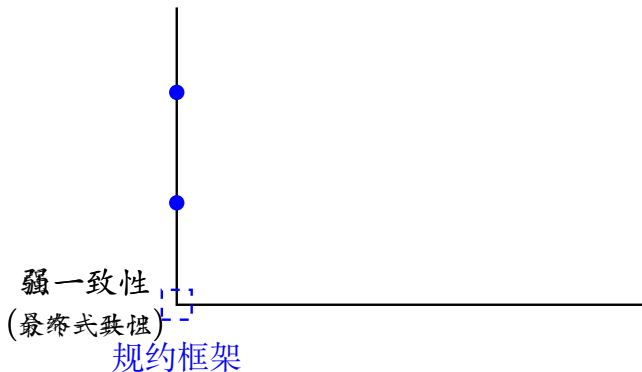
研究工作: 分布数据一致性的形式化规约与验证

不同应用、不同场景需要强弱不同的数据一致性规约



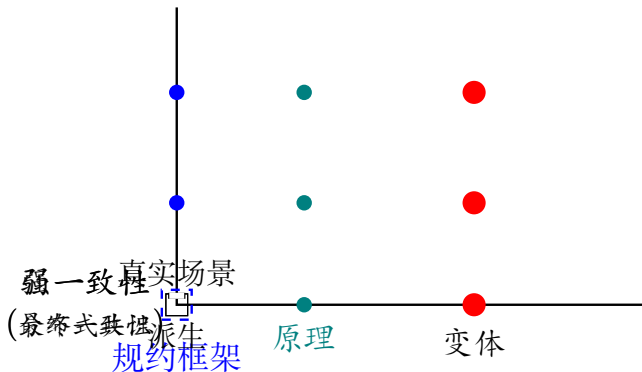
既关注典型的一致性规约、又研究统一的一致性规约框架

研究工作: 分布数据一致性的形式化规约与验证



研究工作: 分布数据一致性的形式化规约与验证

形式化验证方面的挑战: 不同应用、不同场景产生了不同的协议变体



使用精化技术研究众多变体的正确性以及它们之间的关系

(精化技术 (Refinement) [ERefinement:TCS1991] [Lamport:EATCS2018]:

数据精化 (Data Refinement) + 动作精化 (Action Refinement))

研究工作: 三份典型工作介绍

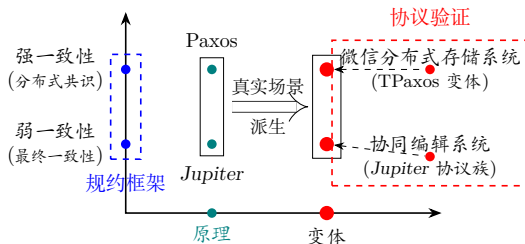


Figure 研究工作概述

- (1) *Jupiter* 协议族的验证
(已发表: [PODC-BA'2018](#), [OPODIS'2018](#); 在审: [TSE'2020](#))
- (2) TPaxos 协议的验证 (在审: [软件学报'2020](#))
- (3) 规约框架 (正在进行, 基本完成)

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化



(a) Google Docs



(b) Apache Wave



(c) Wikipedia



(d) L^AT_EX Editor

Figure 协同文本编辑系统

这是“协同工作”^a 与“人机接口”^b 领域的重要主题之一

[Ellis:SIGMOD89] [Nichols:UIST95] [Ressel:CSCW96] [Sun:TOCHI98] [Xu:

^a如 CSCW: Computer-Supported Cooperative Work and Social Computing

^b如 TOCHI: ACM Transactions on Computer-Human Interaction

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

这是“协同工作”与“人机接口”领域的重要主题之一
然而，这些工作所设计的协同协议大多缺少严格的规约与证明



Specification and Complexity of Collaborative Text Editing

Hagit Attiya
Technion

Sebastian Burckhardt
Microsoft Research

Alexey Gotsman
IMDEA Software Institute

Adam Morrison
Technion

Hongseok Yang
University of Oxford

Marek Zawirski*
Inria & Sorbonne Universités,
UPMC Univ Paris 06, LIP6

Figure[Attiya:PODC16]@PODC'2016

Hagit Attiya
(ACM Fellow)

(2011 年 Dijkstra 奖获得者)

提出两个重要规约: 弱列表规约与强列表规约
证明了 RGA [Roh:JPDC11] 满足强列表规约

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

这是“协同工作”与“人机接口”领域的重要主题之一
然而，这些工作所设计的协同协议大多缺少严格的规约与证明



Specification and Complexity of Collaborative Text Editing

Hagit Attiya
Technion

Sebastian Burckhardt
Microsoft Research

Alexey Gotsman
IMDEA Software Institute

Adam Morrison
Technion

Hongseok Yang
University of Oxford

Marek Zawirski*
Inria & Sorbonne Universités,
UPMC Univ Paris 06, LIP6

Figure[Attiya:PODC16]@PODC'2016

Hagit Attiya
(ACM Fellow)

(2011 年 Dijkstra 奖获得者)

提出两个重要规约: 弱列表规约与强列表规约
证明了 RGA [Roh:JPDC11] 满足强列表规约

猜想: *Jupiter* [Nichols:UIST95] 协议满足弱列表规约

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

我们证明了如下**猜想 @PODC'2016** [Attiya:PODC16]

实现复制列表的 *Jupiter* 协议 [Nichols:UIST95] **满足弱列表规约** [Attiya:PODC16].^{*ab*}

^{*a*}Wei:PODC-BA2018.

^{*b*}Wei:OPODIS2018.

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

Reviewer expertise

4. Expert

该类 (OT 类) 协议的首个严格证明

To my knowledge, the paper presents the first ever rigorous proof of an operational transformation-based protocol. This is quite an achievement: most of existing OT protocols, starting from the first one by Ellis&Gibbs, have been shown incorrect. This is because these protocols are very hard to understand, and the present paper contributes to establishing rigorous theoretical foundations of operational transformations. Hence, I consider the result in the paper very important.

该论文中的结果非常重要

证明方法 “is neat”, 很自然

The technique of establishing an intermediate protocol where clients maintain additional information is neat. This seems like a natural way of relating the outputs of different operations, which is required by the list specification of Attiya et al.

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

出于各种原因, *Jupiter* 协议有众多变体, 晦涩难懂、关系纠缠不清

- ▶ 经常不加证明 [Ressel:CSCW96]
- ▶ 证明是错误的 [Imine:ECSCW2003]
- ▶ 勘误也是错的 [Oster:TR2003]

目标: 理清它们之间的关系、验证它们的正确性

(一): 协同编辑系统中 *Jupiter* 协议族的正确性与精化

发现: 变体的动作一致, 采用的数据结构不同, 维护的“信息量”不同

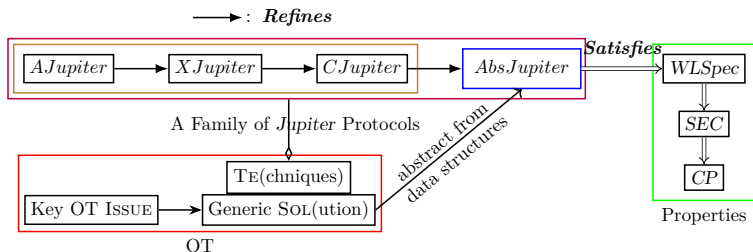


Figure *Jupiter* 协议族的数据精化^a

(二): 共识算法 TPaxos 的推导、规约与精化

PaxosStore: High-availability Storage Made Practical in WeChat

Jianjun Zheng[†] Qian Lin^{†*} Jiatao Xu[†] Cheng Wei[†]
Chuwei Zeng[†] Pingan Yang[†] Yunfan Zhang[†]

[†]Tencent Inc. ^{*}National University of Singapore

 [Tencent / paxosstore](#)

 Watch

91

 Unstar

907

 Fork

220



Figure 分布式存储系统 PaxosStore [Zheng:VLDB2017]^a@VLDB

全面支撑微信业务:

用户账户管理、通讯录、即时通讯、社交网络、在线支付

对于如此重要的系统，它的核心协议一定要是精确无误的!

(二): 共识算法 TPaxos 的推导、规约与精化

TPaxos: PaxosStore 实现的 Paxos 协议变体

1. 看上去与经典 Paxos 差别较大, 难以理解
2. 缺少形式化规约, 自然语言与伪码存在未充分阐明之处

按照这个伪代码的确你说的这样, 调换位置可以算是一个异常路径的优化。不过实际情况这个异常路径走到的可能性不是很高。

rockzheng(郑建军)

3. 缺少数学证明与形式化验证

message processing. In PaxosStore, the Paxos protocol depicted as **Algorithm 1** is implemented in about 800 lines of C++ code, with robustness proven by its successful deployment in WeChat production.

动机: 为 TPaxos 提供形式化规约与验证

(二): 共识算法 TPaxos 的推导、规约与精化

我们的贡献^a:

1. 论证如何从 Paxos 推导 TPaxos:
TPaxos 是 Paxos 的自然变体
2. TPaxos 的 TLA⁺ 规约:
发现未充分阐明的微妙之处
提出新变体 TPaxosAP
3. 验证 TPaxos 与 TPaxosAP 的正确性
(动作) 精化技术
提出新的“投票”机制

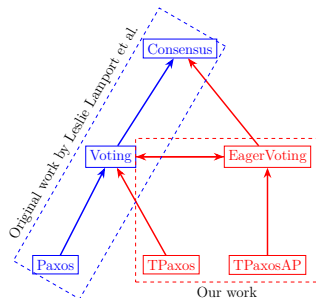


Figure 精化关系图

^aWei:JOS2020.

(二): 共识算法 TPa_{xos} 的推导、规约与精化

目前PaxosStore在微信大规模实施运营的过程中，除了你提到的保证正确性带来的挑战。还有更多的是来自真实系统的挑战，比如我们实现的Paxos算法是不考虑拜占庭故障失败，但实际中却总会遇到，包括但不限于数据盘损坏、数据回退、人工误操作删除数据等，这些情况需要去考虑怎么处理。另一部分挑战是来自在线系统对可用性的高要求，以及高可用情况下的性能表现。

我们非常乐意可以跟学术界有些交流，也欢迎黄教授推荐同学来我们这里实习工作。

rockzheng(郑建军)

更多来自工业界的真实问题:

“我们实现的 *Paxos* 算法不考虑拜占庭故障失败，
但实际中却总会遇到”

希望: 今后能与微信部门交流合作，研究解决这些真实问题

(三): 复制数据类型规约框架

目标: 为复制数据类型建立统一的规约框架

Replicated Data Types: Specification, Verification, Optimality

Sebastian Burckhardt

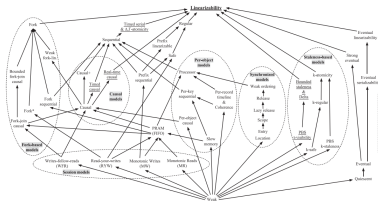
Alexey Gotsman

Hongseok Yang

Marek Zawirski

Figure规约框架

[Burckhardt:POPL14]



Figure多种规约 [Viotti:CSUR16]

已有规约框架，为何再继续研究？

我们有两个主要动机

(三): 复制数据类型规约框架

动机一: 已有框架有特定的目标场景, 没有涵盖很多经典一致性规约

$$(vis, ar)$$

ar : 约束过强, 不能表达“非收敛的”经典一致性规约

我们的扩展: (vis, ar_l, ar_g)

(三): 复制数据类型规约框架

动机二: 发现了通常被忽视的数据类型操作“纯与不纯”的问题

$Pop = Peek + RemoveTop$ is not *pure*

(三): 复制数据类型规约框架

动机二: 发现了通常被忽视的数据类型操作“纯与不纯”的问题

$Pop = Peek + RemoveTop$ is not *pure*

“such operations can always be *separated* into a query and an update which is *not* a problem ...” [UC:IPDPS15]

我们发现: 并非如此!

依赖“简单拆分假设”的工作需要被重新审视

(三): 复制数据类型规约框架

动机二: 发现了通常被忽视的数据类型操作“纯与不纯”的问题

$Pop = Peek + RemoveTop$ is not *pure*

“such operations can always be *separated* into a query and an update which is *not* a problem ...” [UC:IPDPS15]

我们发现: 并非如此!

依赖“简单拆分假设”的工作需要被重新审视

这是一项最近的工作, 技术部分已基本完成

科研方面: 论文情况

已发表 (第一单位、第一作者):

1. RVSI@SRDS'2017
(CCF B)
2. Jupiter@PODC-BA'2018
3. Jupiter@OPODIS'2018

在审论文:

1. JupiterRefine@TSE (第一作者)
2. PARO@TPDS (通讯作者)
3. TPaxos@ 软件学报 (通讯作者)
4. CRDT@ 软件学报 (通讯作者)
5. ASC@TC (其它作者)

继续关注重要的系统、重要的协议
加强与高水平学者以及工业界的交流与合作

科研方面: 参与/主持项目

项目来源	项目名称	个人经费/总经费 (万元)	参与类型
青年科学基金 (2018 年 01 月-2020 年 12 月)	面向分布式系统的复制数据类型 理论与技术研究	25/25	主持
国家重点研发计划 (云计算和大数据专项) (2017 年 10 月-2021 年 09 月)	可成长的智能化网构软件范型 理论、方法与技术研究	50/999	参与
总计 (万元)		75/1024	

科研方面: 获得奖励

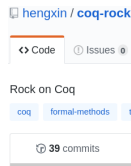
2017 年 CCF 优秀博士学位论文奖

博士论文: “分布数据一致性技术研究”

人才培养方面

团队建设、学术积累 (TLA⁺、Coq 讨论班)

- ▶ 本科毕业设计: 4 名
- ▶ 协助指导硕士生: 2 名
- ▶ 协助指导博士生: 3 名



coq-seminar

本讨论班旨在共同学习 Coq.

以 *Software Foundations* 为主要学习材料.

- *Software Foundations*, Vol 1 - Logical Foundations
- *Software Foundations*, Vol 2 - Programming Language Foundations

以在实践科研中熟练使用 Coq 为目标.

Spring, 2019

No.	Topic	Lecturer	Date	Material
1	FP and Induction	谷青松	2019-05-05 (周四)	sf1-8: Basics v + Induction v
2	Lists and HD Functions	江晋	2019-05-13 (周四)	sf1-8: Lists v + Poly v
3	Tactics	祝浩	2019-05-21 (周四)	sf1-8: Tactics v
4	Logic in Coq	崔昱良	2019-05-28 (周四)	sf1-8: Logic v
5	Inductively Defined Propositions	谷青松	2019-06-06 (周四)	sf1-8: IndProp v
6	Total and Partial Maps	曹泽华	2019-06-13 (周四)	sf1-8: Maps v
7	The Curry-Howard Correspondence	唐保清	2019-07-02 (周二)	sf1-8: PropObjects v
8	Induction Principles	祝浩	2019-07-23 (周四)	sf1-8: IndPrinciples v

Autumn, 2019

服务方面

- ▶ (2018 年 8 月) CCF 2018 年第九届优博论坛报告
- ▶ (2018 年 11 月) 《CCF 通讯》邀稿^a: PODC 会议介绍文章
- ▶ (2018 年 12 月) 青年学者论坛报告
- ▶ 参与本科生开放日面试
- ▶ 参与研究生毕业论文复审

^a感谢 CCF 分布式计算与系统专委会

总结

魏恒峰 (hfwei@nju.edu.cn)

聘期合同要求	工作情况
教学: 承担一门课程	问题求解课程 五个学期; 共 164 学时 (2019 级本科生“我心目中的好课程”)
科研: 4-6 篇高水平论文	发表 3 篇 (含 1 篇短文) 在审 4 篇 (2017 年 CCF 优秀博士学位论文奖)
人才培养	负责或协助指导学生 9 人次 (学术积累: 组织 TLA ⁺ 与 Coq 讨论班)
主持/参与 多个基金项目	主持 1 项; 参与 1 项 个人可支配总经费 75 万元



Hengfeng Wei (hfwei@nju.edu.cn)

