

在线评测教学辅助系统设计

肖红玉, 贺辉, 陈红顺

北京师范大学珠海分校 信息技术学院, 广东 珠海 519087

摘要: 为满足程序设计类课程实验教学及程序设计竞赛的需要, 对国内外现有的在线评测系统进行研究, 以 ACM/ICPC 竞赛模式为蓝本, 采用 Python、Mysql、Django 1.8、HTML5、jQuery 1.11、Bootstrap 3.5、Nginx1.8 等技术, 设计并实现了基于 RabbitMQ 高级消息队列和多级安全沙盒的在线评测辅助教学系统。基于 Celery 并行分布式框架解决了系统大规模评测的并发性需求, 可根据评测负载动态调整评测节点个数, 系统具有高容错性, 个别评测节点的意外故障不影响系统整体运行。通过 linux 用户权限机制、全虚拟化技术、Linux Container (LTX) 和 setUID 实现的多级安全沙盒, 保证了评测环境的公平性和安全性。系统主要为日常实验教学代码评测而设计, 共有 4 个角色, 与面向竞赛的评测系统相比, 增加了教学管理、作业管理、统计分析管理模块。系统已经成功应用在 C/C++/Java/数据结构等课程的日常实验教学中, 运行结果表明, 该辅助教学系统对促进教学质量、提升学生学习兴趣非常有帮助。

关键词: 在线评测; 实验教学; 教学辅助系统; 沙盒

中图分类号: TP302 文献标识码: A

Design and Application of Online Judge Teaching Assistant System

XIAO Hong-yu, HE Hui, CHEN Hong-shun

(College of Information Technology, Zhuhai Campus, Beijing Normal University, Zhuhai 519087, China)

Abstract: To meet the requirement of experimental teaching and programming contest of programming courses, an online judge teaching support system was designed and implemented based on the ACM-ICPC competition model, the RabbitMQ advanced message queue and multi-level safe sandbox after analyzing the existing online judge systems. The concurrency needs of large-scale online judge were satisfied based on Celery. The number of judging nodes was automatic adjusted according to the system load, the system has high fault tolerance, and accidental failures of some nodes was tolerated. By using user permission mechanism of the Linux, full virtualization, Linux Container, setUID, the multi-level sandbox faired and secured the judging environment. The system is mainly designed for code judging experiment teaching, and has four roles. By increasing the teaching management, schoolwork management, statistics management module to meet the requirement of experimental teaching. The system has been successfully applied in C/C++/Java/data structure courses daily experiment teaching. The results show that the system is very helpful to promote the teaching quality and enhance students' learning interest.

Key words: Online Judge; Experiment Teaching; Teaching Assistant System; sandbox

C/C++/Java/数据结构等程序设计课程是高校理工类专业的核心课程, 课程培养目标: 培养学生掌握程序设计的思想和方法, 掌握基本的程序设计过程和技巧, 初步具备分析问题和利用计算机求解问题的能力^[1]。这类课程对实践开发能力、逻辑性思维要求比较高^[2]。提

收稿日期:

稿件编号:

课题编号: 2014A030310415(广东省自然科学基金-博士启动); GDJY-2015-C-b048 (广东省教育研究课题);
作者简介: 肖红玉 (1976-), 女, 江西信丰人, 博士生, 副教授, 研究方向: 分布式水文系统、软件工程; 贺辉 (1979-), 女, 博士, 副教授, 研究方向: 智能空间信息处理; 陈红顺 (1982-), 男, 博士, 副教授, 主要研究方向: 遥感与 GIS 应用、信息系统集成开发。

高学生实践开发能力最有效的方法就是重视实验教学，督促学生在课堂内外编写大量代码。在线判题系统可以实时评判学生提交的源代码，不仅大大减轻了教师的负担，对提高学生学习热情也很有帮助^[3]。ACM 国际大学生程序设计竞赛 (ACM/ICPC)是目前全球公认水平最高影响最大的编程赛事，许多高校开发了基于 ACM/ICPC 的程序设计在线评测系统^[4-9]，孟学多^[3]等基于多核平台实现了在线判题系统；蒋辉等^[10]从满足学生实践的角度设计了在线评测系统；庄奇东等^[11]从页面和数据库缓存、服务器架构等方面讨论了在线判题系统的优化；黄洪波等^[12]讨论了大规模程序评判系统的设计与实现。但这些系统主要为 ACM 比赛服务，不适合直接用于课程实验教学。也有些高校以 ACM/ICPC Online Judge(OJ)系统为蓝本设计了在线评测教学辅助系统^[13]。本文对现有的在线判题系统进行了研究分析，综合各系统的优点与特点，结合我校程序设计类课程实际教学需要，设计了一个面向实验教学兼顾竞赛的在线评测教学辅助系统，与面向比赛的 OJ 系统相比，教学辅助系统增加了课程管理、作业管理子系统和教学中心。此外本文设计了 4 级沙箱安全机制保证系统的安全，基于 RabbitMQ 和 Celery 设计了生产者-消费者模型的高效并发机制。

1 系统总体设计目标

基于 Linux/Windows、Python、Mysql、Django 1.8、Nginx1.8、HTML5 + jQuery 1.11 +Bootstrap 3.5 技术，以 ACM/ICPC 竞赛模式为蓝本，设计并实现在线评测教学辅助系统。系统可以对用户提交的 C/C++/Java 源代码进行编译、运行、实时评测打分并给出明确的评判结果，为老师提供了课程管理、排课管理、题库管理、作业管理、助教管理、教学资源管理等功能，是日常教学、考试、竞赛的一体化平台。学生借助该平台可以完成日常课程作业、查看用户排名、站内检索、下载教学资源等。

2 系统总体架构

系统采用多层分层架构，分为：表现层、业务逻辑层、评测模块、数据访问层。层间关系如图 1 所示。

表现层：接收用户提交的数据，并以图表的形式向用户展示请求的数据，主要用于实现与用户的交互。

业务逻辑层：包含应用于业务对象的规则，系统的业务规则在这部分编写，主要完成以下功能：完成系统业务规则和逻辑的实现，从表现层接受用户请求，根据业务规则处理请求，通过数据访问层完成与数据库的交互任务，将处理结果返回给表现层。

数据访问层：主要完成与数据库的交互即 CRUD (Create、Retrieve、Update、Delete) 操作。数据访问层为业务逻辑层提供服务，根据业务逻辑层的要求从数据库中提取数据或者修改数据。

评测模块：对学生提交的程序代码进行编译、运行。如果源代码在编译过程出现错误，直接返回 CE (Compile Error)，结束判题。编译成功，则读取预定的输入数据并运行程序，获取程序运行时间、内存和 CPU 使用情况，如果发现这些数据超出题目的设定值，则结束判题，并给出对应的提示。如果程序顺利执行，则把输出重定向到临时的输出文件中，并和预先给定的标准输出对比，两个文件的内容一模一样则判定学生提交的代码是正确的，并返回 Accept；否则返回 WA (Wrong Answer)，评测过程如图 2 所示。

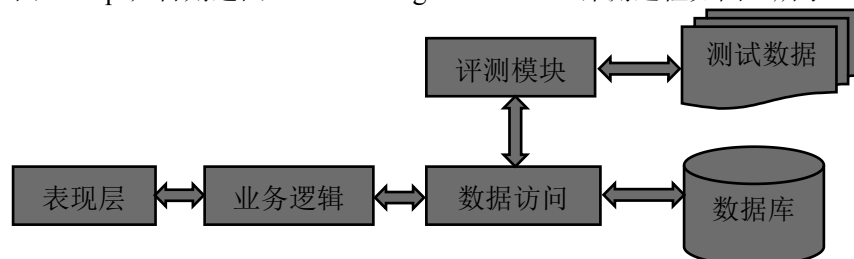


图 1 系统体系结构图

3 功能设计

系统共有管理员、老师、学生、助教 4 种角色,不同角色拥有不同的操作权限,系统功能结构框图如图 3 所示。

管理员：系统管理和权限管理子模块是管理员独享的子模块，系统管理子模块的功能包括：设置系统判题机是否启用、网站名称等初始参数设置；用户管理；网站公告管理。权限管理子模块对各用户的使用权限进行设置和管理。

教师：教师是使用系统的主体，是系统使用的引导者和督促者，拥有除系统管理子模块之外的其他所有功能子模块。

1. 题目管理子模块：该子模块的功能主要是：题目分类管理，目前系统设置了 4 级分类；题目的 CRUD（Create 增加、Retrieve 查询、Update 修改和 Delete 删除）管理，增加题目时需要输入题目标题、题目描述、输入输出描述、样例输入输出、时间和内存限制、题目提示和参考程序，其中时间限制、内存限制可以采用默认值；可以设置是否隐藏题目，被隐藏的题目只对老师可见；本人发布的题目只有自己可以修改、删除，其他老师只能查看、布置作业时选用；每道题都可以查看评测历史。

2.测试数据管理子模块：该模块主要完成测试输入输出数据的 CRUD 操作，为了提高系统的易用性，本文还设计了自动生成测试输出数据的功能。每组测试数据可以灵活地设置分值以及是否启用。

3. 教学管理子模块：教学中心包括课程管理（创建、编辑、删除课程）、课程通知（发布、修改、删除课程通知）、教学资料（上传、删除教学资料）、排课管理（设定上课的起止教学周、上课的具体时间、生成选课码）、学生管理（批量导入上课学生名单、添加单个或多个、批量设定学生上课时间、学生调课管理）、助教管理（添加、修改、删除助教）。

4. 作业管理子模块：发布作业、修改作业、批改作业、删除作业。发布作业时首先输入作业基本信息（作业名称、最高成绩、备注、设置评测语言），接下来从题库中选择题目添加到作业中，设置每道题的分值以及是否必做，设置作业的访问权限，人工批改作业（可批量修改），作业代码打包下载。

5. 统计分析子模块：作业统计分析、个人刷题情况统计分析、全站排行榜等。

学生：学生是使用系统的另一大主体，学生的主要功能是提交代码以及各种查看功能，包括查看评测结果、题目详情、课程公告、网站公告、教学资料、作业、排行榜、统计分析结果等。

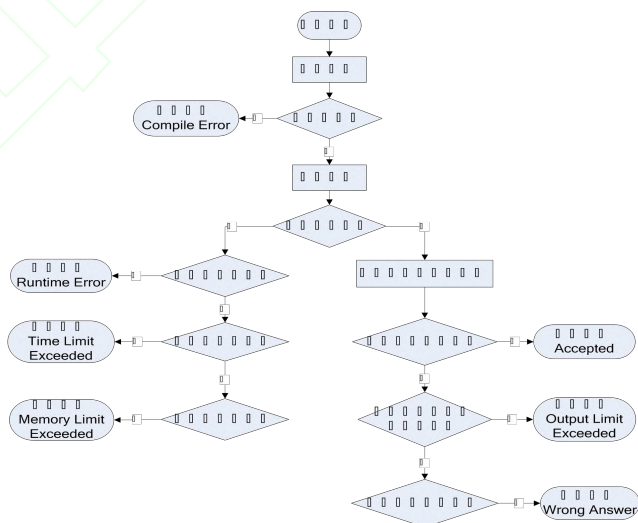


图 2 代码评测流程图

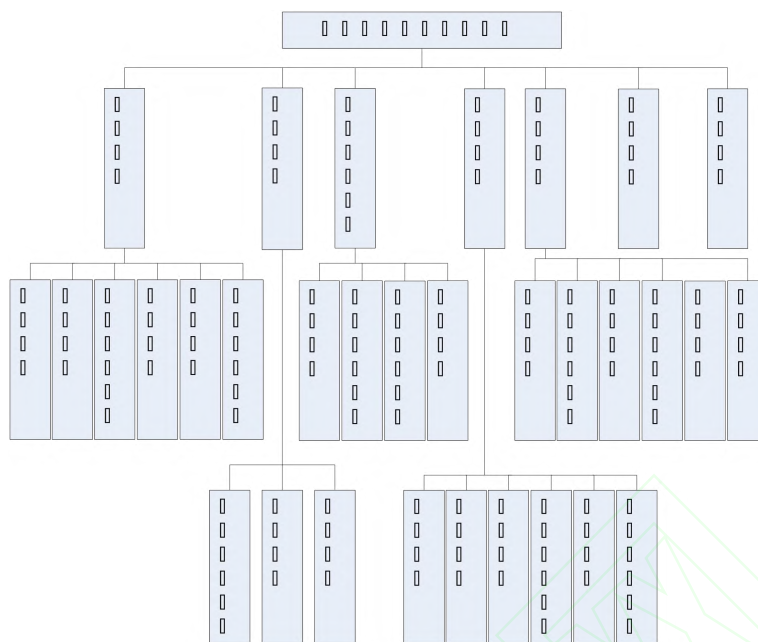


图 3 系统功能框图

4 数据库设计

采用 MySQL 数据库，系统运行过程中产生的所有数据都需要持久化保存到数据库中。根据系统功能需要，共设计了 39 张表，考虑到文章的篇幅，本文只列出了作业管理和教学管理 2 个子模块的相关表，各表含义如表 1 所示。

表 1 数据库设计

模块	表名	表的含义
作业	asgn_asgn	作业表
	asgn_asgnproblems	作业题目与设置
	asgn_asgn_problemset	作业题目集关联表
	asgn_asgnaccesscontrol	作业权限控制设置
	asgn_asgn_access_control	作业访问权限关联
	asgn_solution	作业解题情况
	asgn_solution_judge_status	作业解题情况与评测状态关联表
	asgn_asgn_judge_status	作业与评测状态关联表
	asgn_asgnvisitrequirement	调课请求表
	asgn_stureport	实验报告表
	asgn_asgn_black_list	作业访问黑名单
教学	education_course	课程表
	education_coursemessage	课程消息表
	education_edudepartment	学院信息表
	education_edumajor	专业信息表
	education_arrangement	课程排课信息
	education_arrangement_students	排课选课情况
	education_course_arrangements	课程排课关联表
	education_repository	教学资源库
	education_course_repositories	课程与教学资料关联表
	education_course_assistants	课程助教关联表

5. 安全设计

提交的程序是用户编写的，可能有危害系统安全的行为，常见的恶意为有^[12]：

- (1) 使用非法指令停止系统，如在提交的代码中嵌入 shutdown 等指令。
- (2) 试图删除系统关键文件。
- (3) 大量占用硬盘资源，在代码中反复执行写文件操作，试图消耗掉服务器的硬盘空间。
- (4) 可能耗尽服务器 CPU 资源的死循环。
- (5) 申请大量内存或线程。
- (6) 故意出错，尝试暴露检测代码。

以上恶意为不一定是故意的人为攻击，可能是用户不成熟的编程水平造成的，作为辅助教学系统，用户大部分是大一刚接触编程的学生，死循环是很常见的错误之一，对此在评测机内部对运行时间和内存占用会进行判断，如果超过了题目的预设值，则启动进程将此评测任务 kill 并释放其占用的资源。

对于资源滥用之外的不安全行为，本文提出一种多级安全 Sandbox(沙盒)的权限控制机制，Sandbox 在计算机领域主要指一个严格受控的环境，程序在该环境运行时的状态、所使用或访问的资源都受到严格的记录和控制。Sandbox 根据指令对全系统进行模拟，对程序的访问资源、执行环境按照规则赋予一定的访问权限。程序只能在 Sandbox 里面进行相关操作，限制了恶意程序的危害，每个程序在自己的受保护的 Sandbox 之中运行，不会影响到其他程序的运行，同样，这些程序的运行也不会影响操作系统的正常运行^[12,14]。本文提出的多级安全 Sandbox 机制共分 4 层，分别是 setUID Sandbox、Linux Container 容器 (LTX)、全虚拟化虚拟机和操作系统。为了保证服务器的安全，评测机不是直接运行在操作系统之上，而是运行架设在操作系统之上的全虚拟化层，一方面用户提交的 shutdown、删除文件等非法操作不会影响底层真实系统的正常运行；另一方面，全虚拟化与物理硬件和服务器操作系统隔离，为评测机提供了完整的运行环境，也便于虚拟机的无缝迁移和部署。全虚拟化带来安全保障的同时带来了服务器硬件结构和运行性能的高要求，鉴于评测机需要的系统资源不多，在生产环境中本文将服务器的计算资源划分给多个评测机实例，出于防止相互干扰、易于部署等因素的考虑，各评测机之间需要互相隔离，在考察了相关文献后，本文使用 LXC 技术实现各评测机之间的隔离。LXC 是操作系统级别的轻量级虚拟化手段^[15]，提供了在单一可控主机节点上支持多个相互隔离的服务器容器同时执行机制。LXC 提供了一个拥有自己进程的和网络空间的虚拟环境，与传统的硬件抽象层次的虚拟化技术相比，LXC 具有更小的虚拟化开销、快速部署、不用单独为内核打补丁的优势，评测机运行在各自的 LXC 环境中，互相之间相互独立互不干扰。

用户提交的源代码可能对服务器造成攻击，所以编译代码不能跟评测机放在同一个运行环境，以免源代码的缺陷影响评测机的工作。为了保证评判结果的公正和客观，所有源代码要保证运行环境和策略完全一致，所以在每个源代码被评测前，评测机需要为它准备 Sandbox 运行环境，并且该环境是可重入的。全虚拟化和操作系统级虚拟化的初始成本较高，考虑到能耗比和实际运行的需求，本文设计了基于 Linux 的 setUID Sandbox，在 Linux 中每个进程都会有 uid，不同 uid 进程之间相互隔离，各自都有自己独立的权限，互不干扰。对于不可信的程序，在启动该程序时可以为它分配一个随机 uid，基于上述原理，本文将被源代码降权为随机用户，从而实现了源代码与其他进程的相互隔离。

6 并发性设计

评测机编译和运行用户提交的源代码需要创建和销毁子进程，这个过程需要花费一定的时间，为了评判结果的公平，1 个评测机 1 次只执行 1 个评测任务，并使用单个逻辑处理器，为了加速系统的响应时间提高系统的并发性，本文采用 RabbitMQ-Celery 设计了基于生产者-消费者模型的并发机制，以满足大规模教学评测的吞吐量需求。

RabbitMQ 是基于 AMQP (Advanced Message Queuing Protocol, 高级消息队列协议) 的开源消息队列系统, RabbitMQ 可以根据消费者对消息的处理情况灵活调整消费者权重, 将消息优先分配给处理时间短的消费者, 从而实现系统的负载均衡。Celery 是基于 Python 开发的分布式任务队列, 支持使用任务队列的方式在分布的机器/进程/线程上执行任务调度, 总体架构包括 4 部分: 任务生产者、任务代理、任务消费者(评测机)和任务执行结果存储。在本文中, 生产者是与用户交互的网页前端, 消费者是评测机, 任务代理由 RabbitMQ 负责。用户提交的源程序封装成 RabbitMQ 消息, 在封装时附上相关的消息头, 如使用的编程语言、评测任务类型等。RabbitMQ 在分配消息时根据已声明的任务调度器投递给对应的评测机, 所以, 不同的评测机可支持不同的编程语言或者评测任务。从而, 在实际评测时可根据系统具体的负载情况, 动态地增减对应编程语言或评测任务的评测机, 适应系统用户规模的变化。另外, Celery 支持分布式部署和横向扩展, 可以在多个节点增加评测机的数量来增加系统的高可用性和并发性。

7 在日常教学中的实际应用

平台目前已是我院程序设计类课程(C/C++/Java/数据结构)日常实验教学的辅助平台, 平台访问地址 <https://oj.bnuz.edu.cn/>, 已建设实验题库 5 个: C 程序设计、数据结构、Java 程序设计、C++程序设计、ACM 日常训练。图 4 是某题(题号是 1275)的评测统计情况截图。目前, 系统平均在线活跃用户数约 2000, 已成功完成评测任务几十万次, 这些数字还将随着系统的持续运行不停增长。

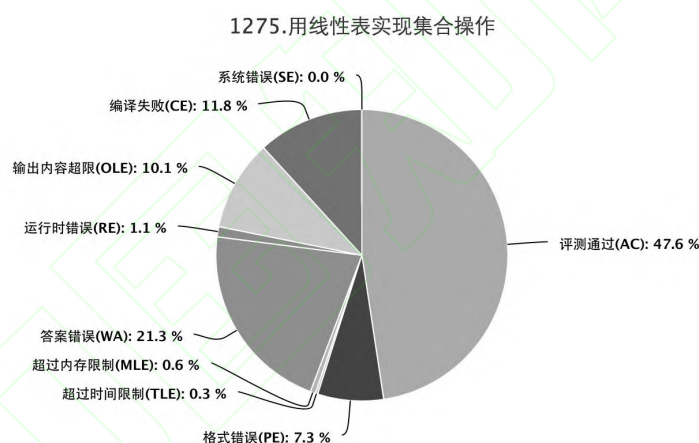


图 4 平台实际应用情况部分截图

8 结束语

本文以 ACM/ICPC 在线评测系统为蓝本, 设计了程序设计实验教学在线评测辅助系统, 系统对学生提交的程序源代码实时评测, 具有人工评判所无法比拟的优点, 系统有别于传统意义上面向比赛的 OJ (Online Judge, 在线评测) 系统, 拥有更适合课程教学的教学中心和作业管理子系统, 对程序设计类课程如 C/C++/Java/数据结构等的实验教学起到了很好的辅助作用, 有效地提升了教学质量。本文设计的基于 setUID、LXC、全虚拟化、linux 操作系统权限机制的多级沙箱模型, 不但可以隔离非受信的源代码, 并且降低了维护成本, 兼顾了系统运行效率。为了适应大规模并发评测的需求, 本文设计了基于 RabbitMQ-Celery 的生产者-消费者并发机制, 在实际评测时系统可自动调整评测机的数量以适应不同规模的教学需求。系统已成为我院程序设计课程辅助教学的有利工具, 对提高教学质量、提升学生学习兴趣很有帮助。

参考文献:

[1]何钦铭, 颜 晖, 苏小红, 等. “程序设计基础”课程教学实施方案[J]. 中国大学教学, 2010(5): 62-65.

- [2]葛文庚, 蔺莉. 程序设计基础课程教学模式研究与设计[J]. 电子设计工程, 2012 (20): 44-46.
- [3] 孟学多, 俞雪永, 颜 晖. 基于多核的在线判题系统的设计与研究[J]. 计算机时代, 2011 (7): 7-9.
- [4] 韩建平, 刘春英, 胡维华. “课内外贯穿, 竞赛教学融合”的程序设计教学模式[J]. 实验室研究与探索, 2014, 33 (6): 169-176.
- [5] 刘 楠, 孙国道, 田贤忠. ACM 在线评判系统设计与实现[J]. 计算机时代, 2012, (2): 34-35.
- [6]谢 迪, 李文新, 郭 炜. “百练”:一个程序设计技能训练与水平测试平台[J]. 合肥工业大学学报(社会科学版), 2008, 22(4): 172-175.
- [7]张浩斌. 基于开放式云平台的开源在线评测系统设计与实现[J]. 计算机科学, 2012, 39 (11A): 339-343.
- [8]曾棕根. 源程序在线评测系统技术改进[J]. 计算机工程与应用, 2011, 47 (4): 68-71.
- [9]车明洙, 纪洪波.一种基于 ACM 程序设计竞赛在线评测系统解决方案[J].微型机与应用, 2010 (4): 11-13.
- [10]蒋辉, 汪大菊. 在线评测系统的设计与实现[J]. 计算机与现代化,2012,198(2): 111-115.
- [11]庄奇东, 王键闻, 张楠, 等. Online Judge 系统的优化[J]. 计算机系统应用, 2012, 20 (8): 115-121.
- [12]黄洪波, 宋鸿陟, 彭红星, 等. 大规模程序评判系统的设计与实现[J]. 计算机工程与设计,2016,37(3):825-831.
- [13]韩君泽,钟美,刘东升. 程序设计在线评测辅助教学系统的设计与实现[J]. 内蒙古师范大学学报(自然科学汉文版), 2010, 39 (5): 473-476
- [14] 陈丹伟, 唐平, 周书桃. 基于沙盒技术的恶意程序检测模型[J]. 计算机科学, 2012, 39 (6A): 12-14.
- [15]吴佳杰. 基于 LXC 的 Andriod 系统虚拟化关键技术设计与实现[D]. 杭州: 浙江大学, 2014.