服务计算技术与系统教育部重点实验室
集群与网格计算湖北省重点实验室

并行與分布式計算通訊

BING XING YU FEN BU SHI JI SUAN TONG XUN

2018年第2期 总第33期 2018年06月

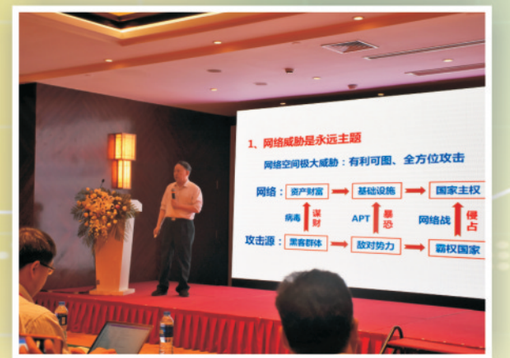
http://grid.hust.edu.cn

并行與分布式計算通訊

地 址：武汉市华中科技大学东五楼一楼
邮 编：430074
电 话：（027）87541924 或 87543529
传 真：（027）87557354
E-mail：wwu.hust@hust.edu.cn
网 址：http://grid.hust.edu.cn

美国弗吉利亚理工大学 楼文菁教授

美国亚利桑那州立大学 张彦超教授



新西兰怀卡托大学 高国隆副教授

澳大利亚莫纳什大学 廖启瑞博士

致实验室2018届毕业生

盛夏光年，毕业的骊歌已然奏响，又到了这个挥手告别的季节。忙碌而夯实的科研时光已在不经意间悄悄流逝。宽敞明亮的实验室、情谊融融的寝室楼，奋勇拼搏的运动场，每一个角落都辉映着鲜活明快的青春年华与温暖美好的独家记忆。在这个梦想开始的地方，你们由最初的迷茫沮丧到对研究方向的驾轻就熟，经历了学科思维方式的形成，感受到了论文专利成功发表的喜悦。这里见证着你们锲而不舍的知识求索，也见证着你们攻坚克难破茧成蝶的华丽蜕变。今天，你们即将告别这片成长的沃土，启程去往远方追寻美丽的梦想。临行之际，请让老师们向圆满完成学业、即将踏上新征程的你们表示最热烈的祝贺！

“张舫：感谢你两次远赴腾讯的付出！”，“陈湃：虽然你有时畏惧困难，但你最终总能做到！”，“刘仁山：懵懂的男孩逐渐成熟了，前行的路上加油！”，“陈吉：勇于面对困难，你将会做得更好！”，“段卓辉：感谢你的付出，接下来一起干一票大的吧!”，“吕青侠：愿你如刀锋战士，未来的路上一往无前！”，“肖威：不忘初心，砥砺前行！”，“汤智阳：愿你勇敢前行找到属于自己的美好未来！”，“林敏豪：朝着心中的目标，诚挚地飞吧！”，“田泽亮：智商和情商‘双一流’，相信你会收获事业的成功和社会的认同，看好你哦！”，“王沛：愿你再接再厉，顺利完成学业，走向事业！”，“肖德山：祝福你在新的工作岗位上开辟新天地！”，“周阳：人在旅途，一路风景；留下美好印记，成就一片天空。”，“叶阁焰：普通话说得越来越好了，加油，IT高手！”，“张杨松：充分发挥你的潜力，明天的舞台一定更精彩！”，“王磊：三石为磊，夯实基础，继续进步！”，“李肖瑶：踏实勤奋、好学上进，我们也是你的‘家人’，常回‘家’看看。”，“金琰祺：期待你继续有志于在学术界不断前进和发展！常回‘家’看看！”，“肖逸凯：去工业界释放自己的潜能吧，心有多大、舞台就有多大！”，“黄东锋：必将在工业界大有用武之地！期待你的好消息！”，“梅超：在腾讯云除了继续炫酷发型还要争取多长点肉。”，“胡佳焕：祝未来工作岗位中，取得更加深入和丰硕的成果。”，“刘志毅：以后到头条必须要找你蹭饭。”，“柳密：密神你要让阿里大牛们重新思考人生”，“熊壮：感谢你为863项目做的突出贡献，在华为云加油！”，“陈洋：你的本事可以为网易数据中心节省不少成本喔。”，“唐晓兰：工作中一定要注意细节，没有人会为你承担错误。”，“陈俊：好好体验小米的团队风格，如此有思想的你如果能再外向点，一定是个极品技术男！”，“肖柏昀：柏昀（白云），学霸更需要霸气。”，“陈肖肖：祝你在未来工作岗位上勇往直前，看好你哦。”，“陶晟：去工业界的舞台释放打磨自己的技术，放飞自己的潜力吧。”，“胡清月：清新女孩，为人踏实、做事放心，愿你永远保持一颗纯真的心，一路顺利。”，“樊荣，聪明的小伙，有想法，反应快，再细心一点你会发展的更好。”，“钱立祥、李阳阳、艾明、肖威：希望大家在今后的工作中继续保持钻研精神。祝福大家在新环境中闯出新天地！”，“行方家：愿研究生阶段努力汗水换来的收获和快乐，化作以后工作和生活中的动力，更上层楼。”，“廖良翌：愿你未来，人生中收获水到渠成的成功。”，“胥清清：大大咧咧的铿锵玫瑰，望在工作岗位继续灿烂！”，“李高峰：低调踏实，典型的华科人，祝在工作上怒放！”，“董泽照、王世振、杨嘉玮：愿你们积极地在每次人生挑战中找到机会，愿你们未来的前途光明似锦！”，“刘若雪：祝福你在今后的工作中平安顺遂，多想着实验室哈。”，“石子倩：所有的委屈和担当都将成为你未来成长的财富。加油，老师看好你。”，“阮臻：以后工作中，多一些思考和格局，让你的人生更加灿烂。”，“王祖剑：祝福你能够带着豪爽的大笑走上今后的生活和工作，一路欢畅。”，“彭欢，思行：你们都是个性鲜明的实力派，继续努力，祝你们有一个美好顺利的未来。”，“李永强：感谢三年来默默的付出，祝前行的路上洒满阳光！”，“李浪：踏过荆棘，方见彩虹！”，“王晨龙：进步非常明显，相信你以后一定能干出一番事业。”，“邓俊：相信通过这三年的锻炼你能在未来的工作岗位上发挥特长，干出成绩。”，“夏妍：做事认真踏实，相信你在以后的工作岗位上能游刃有余。”，“李文珂：多才多艺，谢谢你给组里所做贡献，祝你未来一路风和日丽。”，“王斐：大家的斐神，祝前程似锦。”，“赵鹏：总可以给大家惊喜的鹏组，未来工作顺利，家庭美满。”，“柯志祥：组里的表哥，技术过硬，前途无量。”，“张驰：用勤奋铸就未来，祝你事业有成，生活幸福。”，“何涉平：把握当下，造就自己。”，“李文垚：不要让慢成为你的风格，时不我待，只争朝夕。”，“邓之珺：很欣赏你做事踏实，待人热情礼貌的方式，相信你一定成为大才女！”，“赵健：能力很强，多锻炼一下团队领导能力，前途不可限量！”，“刘本熙：欣赏你的系统底层研发的功底，期望未来在核心创新技术方面有你的一席之地”，“卢宇：你兼具男生的能力和女生的做事细心风格，相信你未来一定能走得很远！”，“刘元栋：愿你在新的征程中继续做一个成熟、稳重的‘高中生’。”。

各位同学，毕业不是学习的结束，而是迎接未来征途的崭新开始。前行的道路上，或许一帆风顺，或许布满荆棘。“浩渺行无极，扬帆但信风”，请力求知行合一以立足社会，坚持终身学习以探寻真理，乘风破浪，勇往直前。愿你们承载着老师们的祝福牵挂，珍惜韶华、砥砺前行，用百倍的信心，书写新的人生篇章。亲爱的同学们，你们在这里的一言一行将化为珍藏的永恒记忆，请记得无论身处何方这里永远是你们温暖的家！

肖　江

二〇一八年六月

主　编：金　海
本期执行主编：肖　江
编　委：陈汉华、代炜琦、丁晓锋、耿　聪、顾　琳、胡　侃、华强胜、黄　宏、蒋文斌、廖小飞、刘方明、刘海坤、刘英书、陆　枫、吕新桥、马晓静、羌卫中、邵志远、石宣化、王多强、吴　松、肖　江、谢　夏、徐　鹏、余　辰、于东晓、袁平鹏、章　勤、张　宇、赵　峰、郑　龙、郑　然、邹德清
（按姓氏拼音排序）
责任编辑：吴　未

地　址：武汉市华中科技大学东五楼二楼
邮　编：430074
电　话：（027）87541924 或 87543529
传　真：（027）87557354
E-mail：wwuhust@hust.edu.cn
Homepage：http://grid.hust.edu.cn
（此刊仅供内部交流学习）

gVisor：谷歌最新发布的用于提供安全隔离的轻量级容器运行时沙箱

（李屹诺 整理）

谷歌发布了一种新型沙箱gVisor(项目主页：https://github.com/google/gvisor)，可以用于为资源占用较少、不需要运行完整VM的容器提供安全隔离。gVisor的核心是一个Golang编写的开源用户空间内核，与现有的容器技术相比，其设计做了不同的权衡，它实现了Linux系统外层的主要部分。该项目包含集成了Docker和Kubernetes的OCI运行时 runsc。

据gVisor项目的GitHub README介绍，gVisor是一个作为普通非特权进程运行的内核，支持大多数的Linux系统调用。就像在VM中一样，在gVisor沙箱中运行的应用程序有自己的内核和虚拟设备，与主机和其它沙箱区分开来。通过拦截应用程序系统调用并作为客户内核运行，gVisor提供了强隔离边界，可以将其视为半虚拟化的操作系统，与完整的VM相比，资源占用更灵活，固定成本更低。不过，这种灵活性牺牲了性能和兼容性，对于频繁进行系统调用的工作负载，gVisor的性能可能会差一些。gVisor不是简单地把应用程序系统调用重定向给主机内核，而是实现了大多数内核原语（信号量、文件系统、管道、内存管理等），并基于这些原语实现了系统调用处理程序。虽然gVisor 实现了Linux系统API的一大部分（目前200个系统调用），但有几个系统调用和参数还不支持（包括/proc和/sys文件系统的某些部分），因此，并不是所有的应用程序都可以在gVisor内运行。

gVisor运行时通过 runsc （run Sandboxed Container 的缩写）集成了Docker和Kubernetes，遵循OCI运行时API标准。runsc运行时可以和runc互换，后者是Docker的默认容器运行时。在Kubernetes中，大多数资源隔离发生在pod层，这让pod特别适合作为gVisor沙箱的边界。Kubernetes社区目前正在规范化沙箱pod API，但是，现在已经提供了试验性支持。runsc运行时可以通过cri-o或者cri-containerd项目在Kubernetes集群中运行沙箱化的pod。这两个工具会把Kubelet的消息转换成OCI运行时命令。

gVisor使用Go语言编写，选择它是考虑到它的内存和类型安全性。需要注意的是，gVisor目前只能在x86_64 Linux kernel 3.17+上构建和运行，而且在沙箱内只支持x86_64二进制文件（即不能运行32位二进制文件）。

HTC推出区块链手机，是颠覆Or最后一搏？

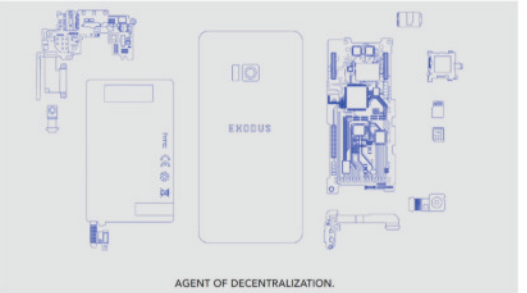
（包庆华 整理）

HTC推出新款手机啦！不过作为一家老牌手机制造商，发布新手机应该并不是什么新鲜事，不过这并不是一款普通的手机，这是一款最近很热的“区块链手机”。2018年以来，区块链热迅速占领全球，成为科技界最热门的话题之一。手机作为一种汇集各种尖端技术的技术实施对象，通常会成为最新科技的试验田。理所当然，区块链也会被人思考可不可以用到手机上，“区块链手机”这种概念由此而生。HTC区块链手机能干啥？

据了解，HTC的这款区块链手机代号为Exodus，这名字似乎很有深意，英文意思是《圣经》中的“出埃及记”，讲述了以色列人在耶和华（上帝）美妙的引领他所选之民以色列人下离开在埃及为奴。看来HTC对这款区块链手机寄予厚望。

主打加密货币。Exodus将提供一种系统级钱包功能，内置安全硬件以支持数字加密货币和去中心化应用。根据HTC的设想，他们未来将以Exodus手机为节点形成原生区块链网络，支持用户之间的加密货币交易。

HTC Exodus的设计图纸也被曝光。从图纸上看，这款新机采用单摄像头设计，延续了HTC一贯的ID风格。



目前Exodus手机暂定支持BTC、ETH及其它主要加密货币，同时，HTC也在考虑支持用户以加密货币购买Exodus手机，不过目前售价还未公布。

Exodus并不是第一款区块链手机，它还有多个竞争对手，像糖果、富士康、长虹、联想等。

富士康Finney可以说是目前最火的“区块链概念”手机。这款手机是由富士康与以色列科技创新公司Sirin Labs合作推出的。这款手机将在今年10月份开卖，并将登陆 美国、英国、韩国、越南和日本5个国家，订单已超10万台。根据 Sirin Labs的介绍，Finney区块链智能手机可以对比特币等加密货币进行安全的存储和使用，并且用户在交易加密货币的时候不用缴纳交易费。

根据目前获得的信息，HTC Exodus手机应该主打安全方向，希望这款手机能如其名。区块链手机究竟只是鸡肋，还是会像乔布斯推出首款iPhone那样颠覆整个手机行业，也许只有随着区块链技术的成熟才能慢慢得到答案。

来源：http://www.8btc.com/htc-blockchian-phone-exodus

面向物联网开发的小型 hypervisor

（卞盛伟 整理）

随着物联网规模呈指数级增长，物联网开发者需要支持各种不同的硬件资源、操作系统、软件工具/应用程序。这是一个巨大的挑战，因为许多互联的物联网设备在资源上受到各种限制，例如运行内存空间，CPU核数。虚拟化技术有助于满足这些广泛的需求，但是当前现有的虚拟化方案无法为物联网开发提供同时满足尺寸、灵活性和功能的适当组合。

在2018年的Linux嵌入式大会上发布的ACRN是一款灵活的、轻量级的参考hypervisor，以实时性和关键安全性为设计的出发点，并且通过开源平台为精简嵌入式开发进行优化。传统数据中心适用的hypervisor代码过于庞大，不支持关键安全性业务，同时运行开销过大，这些对嵌入式开发确实必不可少的。专有的解决方案通常价格昂贵且难以提供长期的产品支持，因此需要一款开源的hypervisor方案来应对嵌入式物联网开发中存在的独特需求，ACRN因此应运而生。

目前ACRN只有25K行代码，英特尔开源技术中心为项目的发布贡献了源代码。

ACRN由两个关键部件构成：hypervisor和ACRN设备模块。ACRN Hypervisor是一个Type 1 的hypervisor，可以直接运行于裸机之上。ACRN 设备模块是针对虚拟设备仿真的参考架构实现，它提供丰富的I/O虚拟化支持，目前计划支持音频、视频、图形和USB。随着社区的发展，预计会有更多设备虚拟化功能加入。

ACRN Hypervisor直接运行于裸机之上，在其上可以运行一个基于Linux的服务操作系统（SOS），然后可以同时运行多个客户操作系统，以便整合工作负载。ACRN Hypervisor为Service OS创造了第一个虚拟环境，然后启动Guest OS。Service OS运行本地设备驱动程序来管理硬件，想Guest OS提供I/O mediation。

Service OS以系统最高优先级的虚拟机运行，以满足时间敏感要求和服务质量的要求。Service OS目前可以运行Clear Linux，但是ACRN也支持其他Linux的发行版或者专有RTOS作为Service OS或Guest OS。

为了保持ACRN Hypervisor代码库尽可能小且高效，大部分设备模块的实现驻留在Service OS，用来提供设备共享和其他功能。目的是保证在资源受限的设备上实现小尺寸、低延迟的代码库优化，为物联网开发平台构建虚拟化特性功能。通过这种方法，ACRN填补了用于大数据中心的hypervisor和直接硬件分区hypervisor之间的空白，是各种物联网开发的理想选择。

新型人机交互

（梅辉耀 整理）

人机交互是指通过计算机输入、输出设备，以有效的方式实现人与计算机对话的技术。键盘，鼠标，屏幕等我们经常使用的设备都属于人机交互的成果，为我们使用计算机提供了很大的便利性。随着可穿戴设备等新型技术的发展，新型的更方便的人机交互技术成为未来发展的一个重要部分。在2018年4月举行的ACM CHI上，计算机系人机交互课题组阎裕康等发表了基于物理经验的虚拟物体手势获取技术研究，带来了人机交互的新技术。

手势识别

机器通过摄像头识别用户不同的手势，来解读用户表达的信息，并作出反应。手势识别在VR，AR等领域有着重要作用。阎裕康等创新性地提出虚拟物体选择的自然手势策略。用户使用抓取真实物体的手势选取对应的虚拟物体，简化了搜索物品的过程。通过用户实验，验证了该技术是一种高一致性，准确，低学习成本的物体选取技术。该方法具有极低的学习成本，用户甚至可以自发地发现所需手势。该技术可以有效替代菜单，为VR游戏，VR办公等应用提供支持。

盲空间物体抓取

在实际生活中，我们经常会遇到这样一种场景：我们顺手把一件东西，比如一本书放在我们后面的桌子上，在需要使用它时，我们并不需要回头，只是把手向后面的桌子上我们记忆中的位置摸索一下即可拿到这本书。然而在VR空间中，我们平常所养成的这种记忆和经验完全不同，导致我们很难仅凭记忆和经验抓取到不在视野范围中的物体。清华大学阎裕康等人在CHI'2018上提出了一种通过对目标获取过程建立用户的心理模型和运动模型，实现了通过头戴式虚拟现实的盲获取技术。让用户无需看到虚拟目标，单纯基于空间记忆和自体感知能力就可以抓取目标。该方法可以有效克服VR视野的限制，减少注意力开销，降低用户疲劳，同时能将输入速度提高16%。

EdgeX Foundry: 物联网急需的解决方案

（陶志恒 整理）

最新的Linux基金会项目EdgeX Foundry旨在为物联网公司创建一个通用框架，以提高物联网生态系统的互操作性和健康增长。

物联网在工业和企业领域扮演着非常关键的角色。航运公司正在使用它来跟踪集装箱。它被用于大型仓库以更好地利用空间。它被用于工厂，建筑工地和矿山，以提高工人的安全性。有很多用例。它在消费领域也逐渐成熟，尽管它主要由于各种原因而受到安全问题的困扰，主要是缺乏商业模式。从工业物联网到企业和消费者物联网的一个常见挑战是缺乏标准的基础平台或框架。物联网公司正在做他们自己的事情，重复努力，编写他们自己的代码，他们自己的协议，创建碎片和互操作性挑战，因为他们做了上述所有事情。这实际上阻碍了物联网的健康发展和应用。

Linux基金会希望解决这个问题。超过50家公司聚集在Linux基金会的名为EdgeX Foundry的协作项目之下。就像任何其他Linux基金会的合作项目一样，这个项目的目标是通过将利益相关方聚集在一起，同时允许每个供应商在公共基础之上创建自己的差异化产品，来简化和标准化工业物联网边缘计算。Linux基金会物联网高级总监Philip DesAutels在接受采访时说，EdgeX Foundry的核心软件组件是戴尔的Project Fuse，该公司向Linux基金会开放并捐赠。

EdgeX Foundry旨在将这些优势带入物联网领域：

1. 最终客户可以快速轻松地部署物联网边缘解决方案，并可灵活地动态适应不断变化的业务需求；

2. 硬件制造商可以通过可互操作的合作伙伴生态系统和更强大的安全和系统管理来加快扩展速度；

3. 独立软件供应商可以从与第三方应用程序和硬件的互操作性中受益，而无需重新建立连接；

4. 传感器/设备制造商可以使用SDK编写具

有选定协议的应用程序级设备驱动程序，并从所有解决方案提供商处获取;

5. 系统集成商可以使用即插即用配件与他们自己的专利发明相结合，加快上市速度。

由于EdgeX Foundry的重点是工业物联网：当被问及它能怎样帮助消费者物联网时，他表示消费者与工业物联网之间的界限模糊不清。“当我们考虑消费者范畴时，我们看到了围绕它们形成生态系统的独立产品，”DesAutels说。“Comcast和Verizon出售大量家庭自动化产品。这些产品包括安全系统，智能门锁，烟雾警报器 – 为人们解决实际问题的产品。这种情况下的消费者范畴看起来很像一个小工业或小企业问题。”物联网实际上超出了我们所能理解的范围。DesAutels举了一个能够举办摇滚音乐会和百老汇演出的戏剧IoT公司的例子，他们的平台必须近实时。一切都必须同步 – 实时动作和音乐。

物联网设备正在为企业节省数百万美元的生产损失。“Weir Group是一家150年历史的工业泵制造商，在该领域拥有超过15万项资产，现已转向戴尔技术公司对该领域的设备进行数字化仪器测试而非猜测，来了解这些关键资产的情况，”物联网解决方案和合作伙伴关系戴尔总监Jason Shepherd说，“如果一台泵发生故障，停机时间很快就会造成超过数百万美元的生产损失，而紧急服务事件的成本很高，一些服务行程需要搭乘直升机。”DesAutels表示，我们还将在能源管理系统等场景中看到BYOT（Bring your own things），在这些场景中，可以带上消费级设备，并与建筑系统进行整合，以监测空气质量等等。工业和消费物联网之间将存在很多交叉。处理所有这些情况的最佳方法是创建一个通用框架，即EdgeX Foundry。

然而，面向用户的设备并不是物联网唯一的组成，它实际上只是冰山一角，真正的冰山是后端，数据中心和运行这些设备服务的云。这就是为什么像Cloud Foundry这样的平台即服务项目也是EdgeX Foundry的一部分。面向物联网设备的用户位于网络的边缘，这使得事情变得复杂。您可能正在云中运行大规模分析，同时需要在边缘实现较小的分析，因为您不想将所有的数据发送到云，您只需将一些更高级别的信息发送到云中处理。

Shepherd给出了一个很好的例子，“Weir 建立了与Edge Gateway的连接，并利用传感器数据的边缘分析来预测泵何时会发生故障。云中的分析使他们的客户能够查看跨所有资产的运营趋势，同时保存数据，如果他们愿意的话，最后，Weir使用所有这些新数据来优化产品开发决策并更有效地进行竞争。”“Cloud Foundry的标准化模型是云计算的一个标准化的基础架构，EdgeX Foundry的标准化模型是边缘计算的一个标准化的基础架构。”DesAutels说。

EdgeX Foundry的目标并不止于创建这个框架，它超越了这个框架。认证和规范性是它将关注的另外两个关键领域。Linux基金会拥有许多物联网/云相关项目，包括Cloud Foundry。作为同一组织的一部分，不仅允许这些项目在更深层次上相互协作，还允许这些项目利用Linux基金会在其他领域的日益增长的专业知识，包括规范性，认证和教程。该项目将采用典型的Linux基金会方式进行管理：将有一个技术指导委员会来推动代码并提供项目的技术指导，然后将有一个理事会来推动业务决策。

阿里的开源项目Dubbo 3.0 即将到来

（张望成 整理）

2018年1月8日，Dubbo 创始人之一梁飞在Dubbo 交流群里透露了 Dubbo 3.0 正在动工的消息，并且在2月15号，Dubbo也宣布正式进Apache孵化器，这使得Dubbo最近受到了广泛的关注。

Dubbo是阿里巴巴公司开源的一个高性能优秀的服务框架，使得应用可通过高性能的 RPC 实现服务的输出和输入功能，可以和 Spring 框架无缝集成，Dubbo是国内比较早的、影响较大的开源项目，包括阿里巴巴、京东、当当网、去哪儿网、网易考拉、微店等电商平台都有其成功

应用案例。由于其开源版本很早已经不再维护，便逐渐淡出人们的视野。而这次其3.0版本即将开工的消息意味着Dubbo的满血复活。

据创始人梁飞透露，Dubbo 3.0开工的原因主要是因为其新特性，接下来我们了解下Dubbo 3.0的改动的说明及特点。

第一，新的 Dubbo 内核与 Dubbo 2.0 完全不同，但它兼容 2.0。Dubbo 3.0 将以 Streaming 为内核，而不再是 2.0 时代的 RPC，但是 RPC 会在 3.0 中变成远程 Streaming 对接的一种可选形态。

第二，将会有一个新的内核接口：Streaming docking（简称Streaming）。一切服务治理将围绕这个内核接口进行扩展，支持HTTP/2，同时REST接口也会受到一等公民支持。不过Dubbo 3.0在通讯方面的改动不大，主要是在服务治理和编程模型上的改动。

第三，Dubbo 3.0的新特性将去掉一切阻塞。Dubbo 3.0以“一切同步”为第一目标，在对 IO 密集业务的处理上，它能够提高机器利用率，使得一半机器的成本被节省下来。创始人梁飞还表示，Dubbo 3.0 技术选型重大变更的驱动因素，也就是降低成本，因为在将系统服务化后，全业务线的机器都在等待返回数据，负载压不上去，机器浪费严重。

第四，Dubbo 3.0 将支持可选 mesh，多加一层 IPC，这主要是为了兼容老系统；而内部则会优先尝试内嵌模式。代理模式 Ops 可独立升级框架，减少业务侵入，而内嵌模式可以带业务测试、部署节点少、稳定性检测方便。同时，可以将Dubbo 3.0启动为独立进程，由dubbo-mesh 进行 IPC，路由、负载均衡和熔断机制将由独立进程控制。

目前Dubbo 3.0已经正式投入全职开发梯队，并且在今年的大年三十，成功进入了Apache的孵化项目中，初步 Runtime 已在今年的3月底在线上应用投入使用。这都使其在最近备受关注，Dubbo 3.0的新特性驱动因素就是为了减少企业成本，相信一旦Dubbo 3.0发布，其将会受到企业的广泛应用。

以太坊智能合约漏洞频出 基于UTXO的公链是否是更安全的区块链平台？

（张玮炜 整理）

近日，EDU 智能合约出现重大漏洞，任意账户中的EDU Token可被转走，这些漏洞造成了来自项目方地址中的30亿万枚EDU Token被偷走，并被陆续转手在火币上进行抛售，引发市场进一步恐慌，价格持续走低 。

目前基于以太坊的Token合约风险主要有两种：

一是自己写的智能合约出现漏洞，被人攻击。以太坊作为一个记录DAPP执行结果的区块链，对于基于合约层的智能合约，需要依赖开发者自己来保障其安全逻辑，一旦合约编写者一时粗心，便可能造成一失足成千古恨的后果。例如18年4月22日BeautyChain出现重大安全漏洞，价值几乎归零，BEC凭空蒸发了10亿美元，事后证明是程序员忘记做溢出检查。

二是合约调用的底层合约出问题，间接产生风险。合约层是一台封装了可以执行图灵完备脚本语言的虚拟机，可以通过编写脚本语言作为智能合约部署到以太坊区块链中。去年，一个叫做“devops199”的开发者触发了以太坊Parity钱包漏洞，漏洞出在钱包调用的一个library（以太坊的底层库函数，可以理解成公用合约）上。因为以太坊的合约调用是把地址当指针使用，每次执行用户编写的智能合约都要调用library。“devops199”把底层库函数破坏的同时相当于把所有合约的指针都破坏了，从而使得所有钱包里的资金被永久冻结。

那加密货币复式记账所需的UTXO模型的区块链项目有哪些，在安全性上又有哪些优势？

比特币的UTXO模型经过了多年较为稳定的运行和测试，性能和安全性都有较大的优势。而以太坊的账户模型理解起来比较容易，但是需要考虑更多复杂的临界情况来防止双花攻击和重播攻击。另外做资产交易的量子链和比原链也采用了UTXO模型。

其中使用扩展UTXO模型BUTXO的比原

链，把资产作为新的UTXO形式进行交互，不仅支持无限种类的资产，而且拥有了图灵完备的智能合约。同时杜绝了以太坊帐户模型所存在的数据溢出，或者各种其它漏洞的出现，兼具灵活性和可控性。

（链接：http://www.8btc.com/smartcontract-eth）

关联型数据的查询利器：图数据库

（李贤亮 整理）

关联型数据是一种常见的数据模型，具体表现在数据与数据之间存在着复杂且动态变化的关联性，而数据的存储，查询与分析往往又依赖于数据之间的连接关系，现实社会中社交网络，金融网络，电子商务中都存在着大量的关联型数据。图数据库是管理与分析关联型数据的利器，其基于图模型建立的高效存储查询引擎，在学术界和工业界都有广泛的研究和应用。

关系型数据库无法很好的处理对关联型数据的查询分析，这种二维表结构在处理关系操作时需要执行非常耗时的join操作，例如在社交网络数据中查询某个人朋友的朋友从而进行关系推荐，关系型数据库首先需要通过外键检索到某人的朋友集合，然后再次通过外键获得朋友的朋友集合。这种通过join数据库连接的方式会导致对整表的扫描，浪费大量的I/O和计算资源。

图数据库能很好的解决关系查询应用。在图拓扑中，每个节点能以O(1)的时间访问其邻居节点，所以上述问题在图模型中可以表示为一个从起始节点开始的两层遍历操作，这个过程中查询操作仅仅扫描了需要访问的那一部分数据，系统资源得到充分利用，算法复杂度低。

当前主流的图数据库可以大致分为两大类：1）基于图模型定制的原生存储，查询和计算引擎。2）基于已有非图模型的开源系统及组件，但针对图查询进行功能适配以及性能优化。

Neo4j和TigerGraph属于上述中第一类图数据库。Neo4j是在DB-Engines中活跃度排名第一的图数据库系统，通过将图的节点和边分开存储在不同的一维表结构中，指定节点的边通过指针将所有的邻居边串起来，形成双向链表结构，这样便可满足节点在O(1)时间内访问到所有的邻居列表。Neo4j默认采用自家的Cyber查询语言，Cyber查询语言基于子图匹配规则，在给定一个子图模式以及部分锚定信息的情况下，在数据库中匹配到所有满足模式的子图结构。TigerGraph原名GraphSQL，是一种建立在分布式平台下的商业图数据库系统，因其良好的可扩展性、吞吐量和易用性受到业界广泛关注。TigerGraph采用GSQL查询语言，本身继承了SQL语言的诸多规则特性，易于学习。

第二类图数据库系统数量众多，Titan, Redis- Graph均属于这一范畴。Titan默认采用Apache Cassandra作为存储引擎，而Redis-Graph则是基于Redis数据库，这类系统都会具备key-value的存储结构，可以方便的将图结构映射到key-value存储模型上，具体的将与key节点相关联的所有边信息打包，并以value的形式进行关联，从而可以方便的实现对图结构数据的访问和修改。在这种情况下图数据库系统也可以看作是基于已有组件构建的应用程序，由于已有组件本身已近具备很多特性如原子性和高可用性，可以方便的实现了数据库要求的如ACID特性。值得指出的一点是，Titan采用的Gremlin查询语言，是Apache 顶级项目TinkerPop所定义的规范查询语言，目前已经被许多图数据库系统所采用。其特点在于采用一种给予路径查询的方式，如查询某人的祖父，即可表示为某人父亲的父亲，书写起来十分自然。

图数据库在业界也有着广泛的应用，如智能推荐系统，金融诈骗检测，智能交通物流等。不难发现，对于这类需要对复杂网络进行深入分析的应用需求，图数据库本身有着巨大的优势，随着数据规模的不断扩大，关联型数据的分析需求也正在逐步增加，图数据库也将扮演着更加重要的角色。

问渠那得清如许，为有源头活水来

2017年6月5日，具有国际影响力的ICDCS 17(The 37th IEEE International Conference on Distributed Computing Systems)在美国亚特兰大召开。CGCL/SCTS实验室硕士生艾明的论文“Distributively Computing Random Walk Betweenness Centrality in Linear Time”被该会议录用，艾明在会议上作了陈述性发言。一路走来，这其中浸透着艾明奋斗的汗水，也得益于华强胜、于东晓老师的精心指导。

论文主要研究了图介数中心度的分布式计算问题。介数中心度(Betweenness Centrality)为图的重要性质，在网络分析领域受到了广泛关注。一个点的随机游走介数中心度表示该点被所有其它点对间随机游走所访问的次数。基于此，论文提出了一个分布式随机近似算法，该算法能够以常数倍误差计算所有点的随机游走介数中心度。同时，该文还证明了提出的分布式算法达到了近似最优。

本刊对艾明进行了采访，并以第一人称进行记录。

时光飞逝，转眼之间我的硕士生涯即将结束，回首这三年的生活，既有失败的辛酸，也有成功的喜悦。独处中思考，讨论中启发，摸索中进步，在披荆斩棘的路上有过跌跌撞撞，也收获了累累硕果。总结过去可以使人更快成长，在此分享我的科研历程，希望给大家带来帮助。

一. 初来乍到

在2014年底，在确认保研本校华中科技大学的研究生后，我机缘巧合地跟随了同样也是初入CGCL&SCTS实验室华强胜老师进行本科的毕业设计。由于这也是华老师刚到实验室的第一个年头，在学校没有指导一个学生，因此我有幸成为他重点的指导对象。但是此时大四保研后的轻松生活让我懈怠，对于突如其来的研究生活并没有做好充分的思想准备，加之我对图算法方向的不熟悉，因此当时的研究进展缓慢。

一晃已是2015年初，与华老师之间的讨论如期而至。由于对于论文中的细节我是知其然不知其所以然，这次的讨论也在我对问题一知半解的情况下结束了。此次讨论，让我更深刻明白“科研就如同一面镜子，你怎样对它，它就会怎样对你”。一方面我对自己的现状不满意，另一方面我又不服输于现状，因此我暗暗下定决心，一定要打破这种懒散的状态，不能在继续懈怠下去。

之后的一个月中，我不断研习论文，一遍不懂读两遍，两遍不懂读三遍，直到看会为止。对于文章中的某些细节，我甚至可以做到倒背如流的地步。黑暗过后，黎明终会到来。科研亦是这样的道理，虽然研究的过程是乏味枯燥的，但是随之而来的将会是收获知识的兴奋与喜悦。第二次的讨论也在轻松愉悦的氛围中结束，此时我更加明白“一份耕耘，一份收获”。

二. 要求甚解

在看过好几篇论文过后，虽然这些文章内的内容都能弄懂，但是对于如何做科研，如何有自己的想法，还是一窍不通，我意识到此时

不能仅仅局限于文章之内，还应该放眼文章之外，与之相关的内容均“要求甚解”，特别是一些理论基础。合抱之木生于毫末，九层之台起于累土。没有坚实的理论基础，再美轮美奂的大楼也是将倾之厦，一触即倒。一本《算法导论》，一本《Distributed Graph Algorithms》，在课余时间我研读了数遍相关内容，建立好了基本的知识框架。

在充分理解了最短路径介数中心度的定义与求解方法，以及基础分布式图算法过后，2015年10月，在全组人员的帮助下，我们终于完成了第一篇论文，并投稿了IPDPS 16。由于理论知识的不完备，该文章只是提出了一种求解方法，并没有证明该方法的好坏，虽然该文章得到了weak accept的意见，但依然还是被拒之门外。而此时我们发现，要想证明结果的好坏，下界求解是必不可少的一环。

下界的求解与《通讯复杂度理论》有着密不可分的关系。在我们小组研究过该理论之后，与老师之前得到的结果相结合，得到了一篇新的关于求解最短路径介数中心度的文章。并于2015年年底投稿了ICDCS 16，并顺利被该会议录用，证明了我们研究思路的可行性。也证明了做科研，不能不求甚解，遇到不懂的地方，哪怕是再小的细节，也要追根溯源，把它弄懂、吃透、领悟，只有融会贯通之后，才能发现自己的“新大陆”。

三. 源头活水

在文章被ICDCS 16录用过后，我并没有停下脚步。在该文章的评审意见中，有评委指出除了最短路径介数中心度以外，还有另一种形式的介数中心度，被称为随机游走介数中心度。该介数中心度的定义与最短路径介数中心度不同，因此在算法设计上也完全不同。

新的机遇伴随着新的挑战。开始研究随机游走介数中心度后，仿佛进入了一个新的领域，以往的方法不再使用，转而需要使用到另一套知识体系。由于该问题的定义与求解主要使用到线性代数的知识，而之前研究的内容更加偏向于图算法，因此实际上研究的方法大相径庭。在与华老师以及于东晓老师讨论过后，基本确定还是在原本的图处理模型设计算法，只不过需要以线性代数的形式分析。在着手论文的分析后，我发现论文中某些地方与研一所学的课程《矩阵论》有着千丝万缕的联系。于是基于《矩阵论》的分析，我提出了一个分布式随机游走介数中心度的上界。

而对于文章中的下界问题，我发现传统的分析方法并不能应用到此下界的求解当中。在与小组成员讨论过后发现，一般的下界证明运用的是定量分析方法，而此方法并不适用。因此我转换思路，使用定性分析方法去分析下界，得到了理想的结果。2017年3月，当被通知文章被ICDCS 17录用时，之前的辛酸与忐忑都抛诸脑后，剩下的只有成功的喜悦，现在回想起来，当时奋斗时的情景还恍如隔日，时刻激励着我。

四. 写在最后

在零零碎碎分享完自己的科研经历过后，以下分享一些我个人的科研感悟：

（1）厚积而薄发，勿急于求成。科研需要一定知识和实践的长期积累，才能有后面的得心应手，所谓厚积薄发就是这个道理。刚刚接触上手的时候可能会觉得科研难如登天，这就好比最开始登山的时候，巍峨的大山就能令人望而却步，但是只要一步一个脚印，有一股坚忍不拔的钉子精神，在不久之后就能到达山顶，感受到一览纵山小的豪迈。平时也应当注

意知识的积累，不能认为有些内容与自己无关就完全不关注，其实某些内容是触类旁通的，这也是实验室经常让我们参加各类专家讲座的原因，可能专家们与我们研究方法不尽相同，但是无论是在研究方法亦或是研究思想上，都将会给我们带来启发。

（2）行百里半九十，勿半途而废。在科研中我们常常发现，有时自己有一个想法，但是做到后来发现遇到瓶颈，找不到突破口做不下去就放弃了，然而一段时间过后，就有一篇与自己研究类似的文章被发表出来。其实有时候我们的想法是对的，研究也进行到中途，只是在山重水复时选择了放弃。须知别人可能也是在数次尝试过后才柳暗花明，而我们缺的只是最后的那临门一脚。

（3）行文欲周全，勿顾此失彼。有时候我们写出的文章自我感觉不错，但是在审稿人眼中却并不如意，常常是因为审稿人能看到你自身没有看到的一面。所以在写文章时，尽量做到面面俱到，想到审稿人可能注意到的地方，加以完善，才能提高文章被接收的概率。

致谢

论文能被ICDCS录用，首先要感谢华强胜老师与石宣化老师对我的支持，另外还要特别感谢于东晓老师在百忙之中抽空指导与帮助我，还应该感谢实验室的小伙伴们对我的鼓励与启发。最后，衷心的祝愿STCS/CGCL大家庭发展越来越好，大家在科研的道路上一帆风顺！

艾　明
2015级硕士研究生
研究方向：大数据
Email：aim@hust.edu.cn

言　论（转载）

人生的低谷，寒冷、黑暗、无奈、彷徨，这些不算什么，只是你在它们的磨砺下应该还剩下最初的那份坚强，那不被动摇和充满斗志的心，记住那些寒冷，记住那些黑暗，记住那些无奈与彷徨，记住那些给予你帮助的人，因为那是你成功时的资本，也许在你“黑云压城城欲摧”之际，迎接的是“甲光向日金鳞开”那片光明。

（https://m.weibo.cn/status/4240712352337932?sourceType=weixin&from=107B295010&wm=20005_0002&featurecode=newtitle 包庆华）

烟雨蒙蒙的早晨，他对你说：“走吧，出海去。”你摇摇头，深邃的眼光紧盯着前方。你说：“此行可能会碰上风雨，或许一无所获，看看再说吧。”你担心自己装备先进的小船会无端受损，得不偿失。

而他，拥有的，只是一弯木，一页白帆，一身勇气。

阳光明媚的早上，你正整装待发，却有人在你身边不停赞赏。赞赏你美的小船，还有你美的身姿。你被包围着，被夸赞着。他们甚至让你拍照，说你是海边难得的装饰。你应酬着，不知不觉中有点飘飘然，也忘了时间的流逝。

很快，他来了，带着征服者的微笑从海上满载归来。

你有些吃惊，有些羡慕。但你没有冲动，依然以深沉的样子，思虑着。

船，依然静静卧在海边……。仿佛一个卧着的问号，给经过的游人以思考和启迪……。

人啊，条件优越时，往往顾虑太多，没有拼搏的动力和勇气！常常在犹豫和徘徊中虚度光阴。

家庭条件不好的孩子，会发奋苦读，心理一个念头：努力学习，将来改变家庭，让父母过上好日子。他们往往学习成绩骄人。

工作岗位不是很满意的人，会不安现状，想方设法进行改变，或寻找更好的，或努力表现，企求超越。他们往往抓住了机会，成为佼佼者。

日子过得优裕了，就容易流于平庸。人平庸了，就难以出类拔萃。

但愿海边静卧的船，能让安享于优越生活的人，有所触动……。

（李贤亮 https://user.qzone.qq.com/834658870?source=aiostar）

2018届毕业生成果展示

支持大页和大容量缓存的层次化异构内存系统

姓　　名：陈　吉
研究方向：内存计算
导　　师：廖小飞
指导老师：廖小飞
Email：1327757613@qq.com
QQ：1327757613
联系电话：15927297680
毕业去向：深圳微众银行

大数据应用的涌现，促使计算机提供更大容量的内存。传统动态随机存储器(Dynamic Random Access Memory, DRAM)容量受限，且功耗大，无法满足当前应用的需求。非易失性存储器(Non-Volatile Memory, NVM)存储密度高，没有静态功耗，但读写操作的延迟大、写功耗高，写耐受力差，通常与DRAM一起构造大容量的异构内存。大数据应用在地址转换过程中通常也面临着TLB(translation lookaside buffer)缺失率过高的性能瓶颈。使用大页可以大幅提高TLB的覆盖范围，有效降低TLB缺失率。在平行结构的异构内存上支持大页存在着NVM大页迁移开销大以及细粒度数据迁移复杂的问题，而在层次结构异构内存上支持大页不存在大页内部细粒度数据迁移的困难。

针对上述问题，提出了支持大页和大容量DRAM缓存的层次化异构内存SSLDC系统。SSLDC系统在DRAM和NVM之间使用直接映射，同时在DRAM缓存中使用4KB的大粒度数据块管理粒度，最大程度减少了元数据存储开销，使得元数据能直接放入片上高速缓存中，精简的元数据使SSLDC系统可以支持大容量的DRAM缓存。为防止异构内存间大粒度的数据交换对带宽造成压力，提出了针对DRAM缓存缺失时会从NVM中取数据这一操作的缓存过滤机制，只有超过阈值的热数据才能被缓存到DRAM中，减少了异构内存间的数据交换，从而减轻了带宽压力。此外还提出了一个基于内存实时访问信息的动态阈值调整策略，以灵活适应应用访存特征的变化。

实验表明，SSLDC系统与使用大页的纯NVM内存系统和CHOP系统相比分别平均有69.9%和15.2%的性能提升，并且与使用大页的纯DRAM内存系统相比也平均只有8.8%的性能差距。

基于堆叠DRAM的缓存管理机制

姓　　名：陈　湃
研究方向：内存计算
导　　师：廖小飞
指导老师：吕新桥
Email：1902552434@qq.com
QQ：1902552434
联系电话：15927062770
毕业去向：华　为

随着计算机处理器核数的不断增加，静态随机访问存储器因功耗和技术限制，已经无法满足系统越来越大的缓存需求。虽然动态随机访问存储器的容量足够大，但是由于它的高延迟特点，一直以来仅仅被用作主存。然而，最近的3D 堆叠技术大大提高了动态随机访问存储器的性能，使得访问延迟降为原来的一半甚至三分之一。

堆叠DRAM 有望作为系统最后一级缓存来

满足应用日益增长的缓存需求，从而为促进计算机应用技术的发展带来了新契机，但这种缓存目前还存在硬件复杂度高、可扩展性差、缓存利用率低等挑战。

对于DRAM 缓存来说，命中率和访问延时是两个相互冲突的优化目标。堆叠DRAM 缓存系统（SODA）提出了一种全新的组织结构，设计实现了路定位器和新奇的数据布局策略，避免了请求的串行化访问问题，同时，SODA 还提出了基于空间局部性的策略来降低路定位器的存储开销，增强了缓存系统的可扩展性。对于DRAM 缓存，高度组相连结构是过度设计的，因此，SODA 采用两路组相连进行组织，实现了命中率和访问延时之间的有效权衡。此外，针对主存控制器中的各种请求，SODA 还提出了请求的调度管理机制，使得系统可以有效、稳定的运转。

实验结果表明，相较于当前最为流行的三种基于块的DRAM 缓存系统，SODA在性能、命中率以及访问延时等方面都展现出了很大优势，命中率相较于Alloy-cache提升了8.1%，相较于LH-cache、Alloy-cache、ATcache 系统，平均访问延时分别降低23.1%，13.2%和8.6%，性能分别提升了17%，12.8% 和8.4%。

基于决策树的GPU应用性能评估方法研究

姓　　名：胡清月
研究方向：高性能计算
导　　师：郑　然
指导老师：郑　然
Email：940226261@qq.com
QQ：940226261
联系电话：15171421256
毕业去向：华　为（杭州）

GPU因其卓越的并行计算能力被应用到许多领域，但GPU架构的复杂性使得程序员很难充分利用GPU上的所有计算资源和存储带宽。而GPU内存层次的复杂性和多种形式的并行处理，使得找到GPU应用程序的性能瓶颈非常具有挑战性。目前存在的一些分析工具只提供大量程序员无法很好理解的数据，需要很强的专业知识对其进行进一步分析。而使用数学建模方法来进行性能分析的方法通常过于复杂，需要依赖各种监控工具来获取分析数据，耗时很长，不易于用户使用。

针对上述问题，提出了一种将决策树算法与理论分析模型相结合的性能评估方法GPUPerfML，来快速准确地定位GPU应用的性能瓶颈，并指导应用程序优化。利用决策树算法对通过nvprof工具获取的应用性能特征数据进行学习和训练，构建出能够表征应用性能特征重要度的决策树。基于决策树的特征选择，可以从大量特征中提取对应用性能影响较大的性能特征集合。同时，将这些特征根据对性能的影响程度进行排序，能鉴别出影响性能的几个关键因素、以及它们的重要程度。提出的基于程序特性的理论分析模型根据GPU架构及GPU 应用运行情况的相关性，将性能问题分为计算相关、内存相关、同步相关三大问题，并建立了性能特征与这些性能问题间的映射关系，从而快速直接地识别出GPU应用的性能瓶颈。通过理论分析来指导决策树的构建和特征分析，保证和提高了该方法用于性能评估的准确性。

该性能评估方法被用来分析四种常用且具有代表性的GPU应用程序（Matrix Transpose，Parallel Reduce，BFS和SPMV），识别其性能瓶颈并据此对其进行优化。实验结果表明，该性能评估方法能准确鉴别出不同应用程序的性能瓶颈问题，判断出的影响程度与实际相符，而且据此进行的优化达到了12%~144%的性能提速。

基于部分数据的大型数据图挖掘优化

姓　　名：李永强
研究方向：图计算
导　　师：邵志远
指导老师：邵志远
Email：1377830123@qq.com
QQ：1377830123
联系电话：18207113532
毕业去向：出门问问

图结构数据能够很好地描述现实世界中实体与实体间的联系，它被广泛应用于各个行业的数字化建模中。随着大数据时代的到来，这类图结构数据的规模呈指数级增长，在这些数据上进行的计算导致了非常大的时间和能耗开销。因此，对大规模图结构数据进行近似计算，不仅是受物质条件（计算的效率和能耗）所制约所导致的结果，也将是历史（数据膨胀）发展的必然。现有的在大规模图结构数据上进行的近似计算方法，往往对图结构数据的类型和算法应用具有较高依赖性。因此，研究高效的近似计算策略成为大规模图结构数据挖掘领域的重要方向。

基于随机张量模型（Stochastic Kronecker Model）的近似计算策略是通过挖掘图结构数据的初始矩阵来提取该图数据的基本特征，并使用该初始矩阵进行算法的近似计算，该策略的最重要部分是求取图结构数据初始矩阵的算法MomentSA。MomentSA是一种基于矩估计的快速挖掘初始矩阵的方法，即通过最小化图结构数据的特征属性的真实值与数学期望之间的距离。首先，MomentSA采用URNS（Uniform Random Node Sampling）定理进行图属性的扩展，避免了采用添加数据的方法带来的破坏图属性或者复杂度高的问题。然后，构造关于图结构数据的特征属性的真实值和数学期望之间关系的目标函数。最后，MomentSA采用ADAM（Adaptive Moment Estimation）算法进行参数求解，可以使参数快速地收敛于最优解。

实验结果表明，在相同的实验环境下，MomentSA处理速度比现有的挖掘方法提高了二到四个数量级。同时，MomentSA求取的初始矩阵最为精确，即使用求取的初始矩阵合成图结构数据的属性值与现实世界的图结构数据的属性值之间的差值最小。

基于Pending Period的数据竞争检测机制

姓　　名：林敏豪
研究方向：程序动态分析与调试
导　　师：廖小飞
指导老师：廖小飞
Email：394536367@qq.com
QQ：394536367
联系电话：13469976917
毕业去向：网易游戏

随着计算机技术的不断发展，并行编程作为一种有效的提高程序运行效率的手段，已经越来越融入到日常的软件研发项目中了。并行编程虽然为程序的快速运行带来诸多的便利，但是同样会带来许多问题与麻烦，数据竞争就是并行编程过程中经常出现，但是却难以根治的一类问题。当前一些数据竞争检测工具虽然能够有效的检测出数据竞争，但是却存在开销过大，扩展性差的问题，使得其难以在实际的生产中应用。

基于pending period的数据竞争检测机制实现了基于物理时钟的数据竞争检测机制，该机制利用物理时钟的一致性，代替原有的逻辑时钟记录事件偏序顺序的方式，解决维护一致性高开销的问题，提高了程序的可扩展性。与此同时，与原有记录锁编号用于区别不同锁的方式，在基于pending period的数据竞争检测机制中不需要对不同的锁进行区分，进一步提高了检测效率。此

外，基于pending period的数据竞争检测机制采用局部变量保存变量的pending period，有效减少了大量访问相同变量时，频繁访问修改全局变量导致的额外开销，提高了系统性能。

实验结果表明，相较于采用逻辑时钟的FastTrack数据竞争检测工具，提出的方法在性能和可扩展性方面都表现出很大优势，时间开销平均减少53.72%。想较于混合了逻辑时钟和lockset的ThreadSanitizer检测工具，P减少了41.05% 的开销。与同为基于重叠检测的算法IFRit相比，也减少了21.01%的时间开销，同时提升了被检测出的数据竞争的数量。同时该方法也可以采用采样算法减少开销，在10%的采样率下，开销相较于同是采样算法的Pacer，性能提高了31.28%。

性能敏感的异构内存分配器设计与实现

姓　　名：刘仁山
研究方向：内存计算
导　　师：吕新桥
指导老师：廖小飞
Email：lrs1993@qq.com
QQ：597413954
联系电话：15071439243
毕业去向：拼多多

随着计算机应用的发展与“大数据”的兴起，应用程序需要越来越多的内存，同时由于功耗和工艺的限制动态随机访问存储器（Dynamic Random Access Memory, DRAM）已经无法满足应用程序对内存的需求。由新兴的非易失性内存（Non-Volatile Memory, NVM）和DRAM组成的平行化异构内存系统，因其同时具有NVM大容量的特点和DRAM低读/写时延、低写能耗、寿命长的优势，近年来已经引起了越来越多的关注，为计算机体系结构的发展带来了新的契机。

在平行化异构内存系统中只有将应用程序的数据放置在合适的内存介质中，才能充分利用NVM和DRAM各自的优势。现有的研究主要着重于利用页面迁移的机制来获取更好的性能和更优的能耗效率。然而这些机制都依赖于在线的页面监测，而这种监测机制需付出很大的性能开销，并且以页面为粒度进行数据迁移操作会浪费一定的DRAM带宽并降低DRAM的利用率。

本文提出了一种基于对象的内存分配和迁移机制OAM（Object-level memory Allocation and Migration）。OAM首先利用离线访存分析工具获取对象的访存模式，然后使用性能/能耗模型来指导应用程序的数据分配，同时还会为有需要的数据对象提供迁移操作。基于OAM中所提供的用于平行化异构内存系统内存分配和数据迁移的编程接口，应用程序的源码可通过OAM所提供的静态代码工具自动修改。当程序开始执行后，操作系统根据OAM所提供的分配建议以及当前系统中DRAM的可用量共同决定对象的放置。

实验结果表明，与Linux系统所提供的page-interleaving 数据放置方案相比，OAM在性能能耗积上表现出很大的优势，能平均降低61%应用所需的执行时间。与当前常用的页面迁移算法CLOCK-DWF相比，在SPEC CPU2006和PARSEC 3.0上OAM能够显著减少42%的数据迁移量，并平均提升9%左右的系统性能。

多GPU环境下数据交换及优化机制研究

姓　　名：刘元栋
研究方向：高性能计算
导　　师：金　海
指导老师：郑　然
Email：1531364016@qq.com
QQ：1531364016

联系电话：15927371641
毕业去向：网易游戏

随着通用计算图形处理器的发展和普及，越来越多的科学计算和工程应用使用GPU来加速其性能。但是对于一些非规则应用，其非规则内存访问模式大大限制了性能的提升。目前，一些研究方案通过拓展GPU硬件结构或采用编译技术来解决非规则内存访问的问题，然而这些方法只适合处理静态非规则内存访问，对于访存模式在程序运行后才能知晓的动态非规则访问，作用于编译期的编译技术就不再适用了，而硬件方法成本太高而且不通用。因此，如何使用软件方法解决动态的非规则内存访问仍然是一个严肃的挑战。

针对目前非规则应用使用GPU加速效果不理想的问题，本文提出用数据重组和多CUDA流解决动态非规则内存访问的软件方法。数据重组技术将需要非规则访问的数据重组为规则访问的数据，消除非规则内存访问。而索引重定向技术则将GPU线程重定向为访问重组后的规则数据。为了提高数据重组的效率以及避免额外传输重组的数据，重组数据的操作交由GPU来完成。同时为了消除数据重组产生的额外开销，数据被均分为多个数据块，分别放入不同的CUDA流中，利用Hyper-Q技术，将数据重组的CUDA流和数据处理的CADA流并发执行。同时，缓存重组的数据以供多个非规则内核以及多次迭代过程重复使用，以减少冗余的数据重组。

选取共轭梯度法、调查传播法和分子动力学3个代表性的非规则应用在Tesla K40c GPU上进行了性能测试。实验结果表明：使用该方法消除非规则内存访问后，GPU内核的内存数据传输次数下降了23.1%到62.9%，同时，GPU内核的性能提高了13.5%到35.3%。

面向动态大图异步增量计算的优化机制

姓　　名：肖　威
研究方向：图计算
导　　师：华强胜
指导老师：吕新桥
Email：513921558@qq.com
QQ：513921558
联系电话：13212764668
毕业去向：华　为

在动态图处理系统中，保持图计算结果实时更新至关重要，增量计算能重用历史计算结果来保持计算结果实时更新。然而，现有的异步增量图处理系统仍受限于欠佳的图计算收敛速度,这是因为它们忽略了一些重要顶点(对图收敛速度很重要)对状态传播效率的影响，而且它们忽略了顶点之间的固有连接性，数据的随机分布使得各数据分区之间的通讯开销大大增加，同时也增加了随机访问，降低了系统性能。

面向动态大图异步增量计算的优化机制在原异步增量计算系统上设计实现了一个新的优化框架，着重解决现有异步增量图处理系统缺乏有效的图划分方法与调度方法的问题。首先，该优化框架提出了一种基于重要顶点的动态图划分方法，由于图中的少数重要顶点在迭代计算过程中容易被高频访问，该方法采用优先集中放置重要顶点及其关联的边的策略，以数据块管理各个分区，使得迭代计算具有较好的局部性。然后，在基于重要顶点的动态图划分方法的基础上，该优化框架提出了一种基于数据块的优先级的调度策略，该调度策略优先选择更具有传播影响力的数据块进行处理，能有效提高顶点状态传播效率。图迭代计算过程中，通过使用该优化框架，分区内的重要顶点将拥有更好的局部性，顶点状态传播效率将会增加。

实验表明，面向动态大图异步增量计算的

优化机制能在原系统上正确稳定运行，有效地减少迭代计算更新次数，增加局部性，减少图计算收敛时间。相对于原系统，顶点总体迭代更新次数减少了30%，总体计算时间减少了35%。

分布式深度学习中基于组策略的弱一致性方法研究

姓　　名：叶阁焰
研究方向：分布式深度学习
导　　师：谢　夏
指导老师：蒋文斌
Email：386921016@qq.com
QQ：386921016
联系电话：13260682035
毕业去向：腾　讯（深圳）

随着人工智能的发展以及神经网络模型规模和数据集规模逐渐增大，分布式深度学习已成为学术界和工业界的研究热点，研究一种能高效训练出成熟的神经网络模型的方法具有重要意义。论文对已有的典型一致性算法（包括Sync-SGD、Async-SGD和Stale-SGD）进行了较为深入的分析，结合异构集群中潜在的分组特性，提出了基于组策略的弱一致性方法——分组同步随机梯度下降算法（Grouping Synchronous SGD，Grouping-SGD），对应的并行机制称为分组同步并行机制（Grouping Synchronous Parallel，GSP）。

论文基于遍历收敛理论（Ergodic Convergence Theory），分析得出：Grouping-SGD算法可以在保证模型最终的收敛准确率前提下，大幅提高模型收敛速度。Grouping-SGD 算法中采取了四项项优化措施，包括：（1）分组同步优化，使用 K-Means 算法将性能相近的工作节点分为一组，为了提高收敛效果，同一分组内的工作节点同步更新模型参数，为了提高整体训练速度，不同分组之间异步更新模型参数。（2）参数服务器排列优化，对参数服务器按其更新速度从快到慢排列，让它们分别负责神经网络中从低层到高层的模型参数的更新，可以提高参数服务器整体的更新速度。（3）基于梯度延迟程度的权重因子调整优化，在权重因子作用下，可以促进模型收敛。（4）训练数据“按劳分配”优化，为训练速度快的工作节点分配更多的训练数据。

在两个不同配置的集群上分别对当前常用的数据集Cifar10、Cifar100和ImageNet进行实验，对比了几种一致性算法（Sync-SGD、Async-SGD、Stale-SGD和Grouping-SGD）的性能。实验结果表明，Grouping-SGD 算法的收敛速度最快，它是一种高效的优化算法。

基于Spark的高效并行频繁项集挖掘算法的研究与实现

姓　　名：张　舫
研究方向：机器学习
导　　师：廖小飞
指导老师：廖小飞
Email：345721580@qq.com
QQ：345721580
联系电话：15527952358
毕业去向：深圳腾讯

随着现实生活中数据量的急剧增长，数据挖掘在各个研究领域都引起了极大的关注，特别是游戏领域。频繁项集挖掘是一种非常流行的数据挖掘技术，在许多重要的数据挖掘任务中发挥着重要作用。然而随着大数据的迅速发展，人们对数据中的有价值信息的需求却越来越大，如在大量游戏数据中，人们希望发现用户游戏中的成长历程来给用户更好的游戏体验，而硬件条件却越来越无法满足人们快速对有价值信息挖掘的需求。换言之，在无法改变硬件条件和数据量的前提下，现有的频繁项集

挖掘算法已经不能在有效的时间内满足人们对有效信息的渴望。因此，研究和实现高效的并行频繁项集挖掘算法成为现在数据挖掘领域的重要方向。

提出了一种高效的并行频繁项集挖掘算法PNPFI，该算法基于Prepost算法和Spark平台进行实现。PNPFI算法在Spark平台实现了节点间互不依赖地并行，同时提出了一种新的N-list求交集算法，该算法通过提前判断结果是否符合阈值来提前停止N-list求交集，大大减少了内存和时间的消耗。为了进一步减少部分冗余N-list求交集过程，PNPFI基于N-list提出了一种新的概念P-Subsume，通过P-Subsume可以直接与项结合产生部分频繁项集，而无需N-list求交集操作，大大缩短了算法运行时间。此外，考虑到算法实用性，PNPFI还提出一种负载均衡策略，通过预测项负载来均衡划分事务集，使集群能够达到负载均衡。

实验结果表明，相对于经典并行算法和最近提出的并行算法，PNPFI在性能和内存开销上都表现出较大优势，性能最大提升70%，平均提升39%；内存消耗最大可减少96%，平均可减少71%。

众核模式下深度学习模型细粒度并行方法研究

姓　　名：张杨松
研究方向：深度学习并行化
导　　师：蒋文斌
指导老师：蒋文斌
Email：1107211209@qq.com
QQ：1107211209
联系电话：15927665120
毕业去向：北京字节跳动科技有限公司

近年来，深度学习在计算机视觉、语音识别及自然语言处理等领域发挥着十分重要的作用。深度学习模型通常要进行反复的迭代式训练，需要消耗大量的时间与计算资源，因此引入并行化方法对于深度学习意义很大。现有的并行化方法主要包括数据并行和模型并行。这两种模型级别的粗粒度并行方式通常是通过增加训练单元的数目来获得加速，不能充分挖掘模型自身的可并行性与及硬件资源。

针对上述问题，提出了一种GPU众核平台下深度学习中基于层的细粒度并行方法（FiLayer），具体包括层间并行与层内并行。前者旨在让模型中的相邻层并行处理，后者旨在并行化网络层内部的计算。为了实现层间并行，设计并实现了数据流水线算法。该算法先将网络模型按层划分成多个部分，其次将训练数据划分多个分片，最后以流水线的方式让多个网络层并行计算多个分片数据。为了实现层内并行，设计并且实现了并行化卷积算法，包括并行化卷积前向计算算法和并行化卷积后向计算算法。并行化卷积算法将卷积层的计算划分成多个部分，分发给不同的部件进行计算。CUDA流技术被应用到了FiLayer的实现中，保证了FiLayer在物理上真正的并行执行。相对于现有的基于模型的粗粒度并行方法，FiLayer充分挖掘了模型自身的可并行性和训练单元上的计算资源，在保证网络模型训练精度的前提下，提高模型的收敛速度。

上述方法实现在Caffe系统上，拓展后的系统命名为LP-Caffe。采用四个有代表性的标准数据集（MNIST, CIFAR10, CIFAR100及ImageNet）在三个不同的硬件环境上对所提出的方法进行了性能评估。实验结果表明，FiLayer 在GPU上能帮助Caffe获得1.58x-2.19x的加速。

基于线上线下数据的商业规划评价模型

姓　　名：陈　俊
研究方向：城市计算
导　　师：余　辰

指导老师：余　辰
Email：junc@hust.edu.cn
QQ：2455004028
联系电话：15927424295
毕业去向：小　米

城市计算是融合城市产生的大量多源异构的空间数据来解决城市面临的挑战。综合城市的交通数据、地理数据、环境数据、社交网络数据、O2O（Online to Offline）数据等等，采用数据挖掘、机器学习和可视化技术，城市计算为城市的诸多问题提供解决方案或支持。

在城市计算中，结合异构数据集来评估新商业点的规划对于投资的成功至关重要。传统方法主要解决线下地理位置的排名问题。然而，随着新业务模式–线上到线下的兴起，人们的行为受到移动互联网的影响，新的业务模式逐渐改变了传统行业。因为在移动互联网上可以看到如此多的信息，人们有了前所未有的巨大选择空间。因此，新商业点的规划也应考虑线上的商业策略。而来自O2O平台的这些新数据集则为此提供了重要知识。

因为城市数据每天的获取量可达Gigabyte/Terabyte级别，首先针对海量空间数据预处理的速度优化，将目前单机算法进行分布式并行化。并行化就是考虑在分布计算加速时如何保障算法和单机情况下的结果保持一致。然后，在预处理空间数据的基础上，选择酒店规划作为分析对象，提取不同的特征来进行详细的可视化和分析。提取的特征可以分为三类：地理特征，商业策略特征和移动性特征。接着，通过融合不同的特征，设计模型来评估新酒店的规划。根据实验结果，所提出的特征和模型不仅可以解决传统的地理位置排序问题，而且可以根据商业策略简单预测酒店的交易量。它为移动互联网时代新商业点的规划提供了更全面的评估。

网络功能虚拟化的部署和流调度

姓　　名：陈肖肖
研究方向：网络功能虚拟化的资源调度研究
导　　师：陆　枫
指导老师：顾　琳
Email：gailchen@foxmail.com
QQ：1285255670
联系电话：13129937819
毕业去向：华　为

随着网络的不断演进，网络功能虚拟化（NFV）成为一种用以减少网络服务提供商的购置成本和维护成本的潜在解决方案。网络功能虚拟化解锁了网络功能软件与硬件的依附性。通过将虚拟网络功能（VNF）运行在通用硬件上，NFV使网络变得更加灵便可控制，从而减轻了传统网络扩展和维护的昂贵成本等问题。但是，仍然还需要解决很多的技术问题，特别是如何在分布式的数据中心中高效的部署VNF和调度网络流的问题，即虚拟网络功能服务链的部署问题。虚拟网络功能服务链部署问题的优化目标是最小化部署成本和通信成本，通过多级选址问题可以证明该问题是一个NP难问题，其无法在线性时间内求得解。对于不考虑虚拟机资源限制的情况下，即无资源限制的虚拟网络功能服务链部署问题，采用低复杂度的基于松弛策略的算法（RLX算法），进而求得VNF部署方案和数据流调度方案。对于有资源限制的情况下，即有资源限制的虚拟网络功能服务链部署问题，还需要考虑不同网络功能（NF）的不同资源需求的问题，先通过松弛策略获得初始VNF部署方案和初始流调度方案，然后采用贪婪策略对VNF进行装箱处理，最后将装好箱的VNF部署在合适的虚拟机上（REP算法）。最后，通过大量的仿真实验来分析RLX算法和REP算法的性能。将RLX算法与最优解OPT和最短路径算法SP的结果对比，并将

REP算法与已提出的链式部署算法CDM和单实例算法SV进行比较，结果表明，RLX算法和REP算法能够有效地降低各种场景下的总体部署成本和通信成本。

面向延迟敏感型应用的数据中心功率利用率提升方法研究

姓　　名：陈　洋
研究方向：云计算
导　　师：吴　松
指导老师：吴　松
Email：1540734639@qq.com
QQ：1540734639
联系电话：15527304439
毕业去向：网易游戏

数据中心已成为全球信息化建设的基础支持设施。功率供应是建造数据中心的最大限制之一，是数据中心昂贵的资源。然而部署了大量延迟敏感型应用的数据中心功率利用率低下问题一直困扰着其管理者。由于延迟敏感型应用对性能要求较高，数据中心管理者不得不为该类应用预留大量功率以满足其性能要求，这也成为数据中心功率利用率低下的主要原因之一。降低应用的功率预留量可以有效提升功率利用率，然而预留功率不足可能会带来不可接受的应用性能损失。

针对以上问题，提出了面向延迟敏感型应用的数据中心功率利用率提升方法，解决了两个问题，1）在必须满足应用性能要求的情况下，为服务器预留多少功率最为合理？2）在数据中心额定功率限制下，如何将服务器部署其中，可以有效提高数据中心功率利用率？针对问题1中预留功率不足时可能造成的性能损失，提出一种定量衡量性能损失的方法—细粒度微分方法，该方法基于微分任务量不变的原理，可细粒度精确评估给定预留功率阈值后可能的性能损失，能够在给定的性能要求限制下，得到最合适的预留功率。针对问题2，采用“装箱问题”的形式，将各台服务器按照最合适的预留功率部署在数据中心中，能够在给定的额定功率限制下，保证数据中心的功率峰值控制并有效提升数据中心服务器部署密度和功率利用率。

对真实数据中心中25000多台服务器的历史数据验证结果表明，所提细粒度微分方法在评估应用的性能损失上准确度很高，误差仅为5%。同时，与常用数据中心服务器部署方法对比，所提方法在给定性能要求下，可将服务器部署密度提高12.6%，功率利用率增至1.12倍。

基于容器的移动云平台系统定制技术研究

姓　　名：胡佳焕
研究方向：云计算与移动计算
导　　师：陈汉华
指导老师：吴　松
Email：hhunger@qq.com
QQ：1027814295
联系电话：13164690179
毕业去向：百度在线网络技术（北京）有限公司

近些年，随着各种智能手机、可穿戴设备的飞速增长，移动设备已成为人们日常生活以及工作中不可或缺的一部分。移动云计算作为一种逐渐成熟的技术，其不仅为移动终端用户提供了更加丰富的应用服务模式，同时也满足了日益复杂的移动应用在计算能力、存储容量、安全性等各方面对移动终端提出的多样化的需求。移动云平台作为移动云计算中的核心角色，也正扮演着越来越重要的作用。

当构建一个可行的移动云平台时，一件重要的事情是为移动应用程序提供一个可供执行的运行环境，例如可供Android应用运行的云端环境。尽管已经有很多研究工作在这方面做出了巨大的贡献，但是它们忽视了一个重要的问

题，即移动云平台存在着定制Android应用运行环境的需求。由于现有的系统定制技术仅针对特定的硬件平台或特定的驱动程序，而几乎没有考虑到云端平台多样化的应用场景，同时也缺乏统一的定制方法以及有效的升级和维护机制。因此，传统的系统定制技术已再不适用于大规模且多样化的云端应用场景。

基于容器的移动云平台系统定制技术提出了一种统一且有效的方法来定制移动云平台的Android应用运行环境。该方法提供了一种基于容器的解决方案来定制Android系统组件，同时为不同的云端应用场景提供了可行的Android应用运行环境。在此方法的指导下，基于容器的移动云平台系统定制技术开发了一个名为AndroidKit 的自动定制工具集来生成特定的Android操作系统组件。通过这个工具集，可以启动名为AndroidX的新的Android虚拟机实例。这些AndroidX由AndroidKit生成的系统镜像组成，从而可以被很容易地定制和组合以满足云端各种应用场景的需求。

移动设备D2D通信下的节能研究

姓　　名：金琰祺
研究方向：移动计算与绿色计算
导　　师：刘方明
指导老师：刘方明
Email：richie.jyq@gmail.com
QQ：595370122
联系电话：15623554321
毕业去向：招商银行武汉分行

随着移动通信技术和移动互联网的高速发展，即时通信类软件（例如微信、QQ、WhatsApp等）已经成为移动用户最常用的应用之一。为保持移动用户的在线状态，该类应用需要在后台向服务器频繁发送心跳包。在传输心跳包时，移动设备（如智能手机）与基站之间需要频繁建立与释放通信信道。心跳包的频繁发送消耗了大量的信令资源，是造成运营商基站“信令风暴”这一痛点问题的重要因素之一。同时，这种频繁的小包传输也会导致移动设备频繁激活数据通信模块，造成移动设备大量的电量消耗。

为了解决心跳包给运营商和移动设备带来的信令与电能的双重消耗，应用5G时代下“端到端”（Device-to-Device，D2D）多设备智能互联通信技术开发了一种新型的群体心跳包互助转发系统。其基本思想是：首先，通过合理挑选和激励蜂窝网络中的部分移动设备作为转发设备，并适时以高能效的D2D通信方式收集周围邻近的移动设备心跳包；然后，由转发设备将收集到的心跳包进行统一批量发送，从而大幅减少蜂窝网络内众多移动设备的信令消耗。转发请求设备在D2D连接之前，该系统通过搜索到转发的设备信号强度和负载情况决定是否进行D2D连接。转发设备在转发心跳包时，该系统能够基于心跳包的周期和失效时间进行合理的传输调度，从而在满足延迟约束的条件下最小化基站信令和移动设备电能消耗。在安卓智能手机上，开发了基于该系统的真实App，实测验证了该系统能够减少50%以上的信令消耗并节省高达36%的电能。

基于FPGA的高性能网络功能加速平台

姓　　名：李肖瑶
研究方向：网络功能加速
导　　师：刘方明
指导老师：刘方明
Email：calmisi975@gmail.com
QQ：358571803
联系电话：18627743294
毕业去向：英特尔亚太研发中心

网络功能虚拟化将网络功能从昂贵、固化

的专用网络设备解耦到通用服务器上，以软件的方式部署和运行，极大地提升了灵活性。然而，软件网络功能在进行深包处理时，需要占用大量的CPU核心数才能达到线速度。另一方面，由于FPGA支持高并发并且可编程，使用FPGA加速深包处理是一种可行并且具有发展前景的方案。但是，将整个网络功能部署到FPGA上会造成不必要的资源浪费，因为FPGA中的可编程逻辑十分有限且成本高昂。此外，当需要更改网络功能时，开发人员需要耗费数小时生成新的FPGA程序，这使得网络功能的快速部署难以实现。

针对上述FPGA在软件网络功能性能加速中面临的关键挑战，提出一种基于动态硬件库（Dynamic Hardware Library, DHL）的FPGA-CPU协同设计框架。该框架旨在保证软件网络功能平台的高性能和高灵活性：（1）将深包处理过程实现为FPGA中的加速模块，并将这些加速模块抽象成硬件函数库，并为开发人员提供一套DHL编程API，实现同类多个网络功能的复杂逻辑能够统一部署在FPGA中加速，而简单逻辑仍协同在CPU中运行，从而实现多个网络功能在FPGA-CPU架构中灵活、快速、高性价比的部署，具有高通用性和易编程的优势；（2）通过结合无锁通信队列、用户态I/O、NUMA感知的内存分配、批处理和轮询等一系列实用优化技术，最优化DHL整体框架的网络性能。（3）实验结果表明，DHL框架极大地降低了软件开发人员使用FPGA的编程难度，同时相比于纯CPU方案吞吐率更高，延迟更低，相比于纯FPGA方案FPGA的资源利用率更高。

共享单车网络中基于云计算的高效数据迁移方法研究

姓　　名：刘若雪
研究方向：云计算
导　　师：陆　枫
指导老师：陆　枫
Email：rxliu@hust.edu.cn
QQ：313591652
联系电话：15172509310
毕业去向：武汉市科技局

随着共享单车系统的发展，车辆间的信息交换也有着大量需求，例如及时知晓道路状况、交通事故、或者与其他单车的距离位置信息等等，需要构建一个共享单车网络来交换数据。共享单车这种新型系统存在随机性和灵活性，导致车辆间的网络通信有了更大的难度，阻碍了其智能交通的发展。并且单车数量庞大，伴随产生了大量数据，而单车移动设备的数据计算能力十分有限，不足以支持实时消息的处理。

将待处理的应用数据迁移到附近的车载设备上进行处理，并进一步转发传输，消息通过共享单车车载自组网络传递到某个特定节点，可以帮助车载移动设备高效处理大量信息并及时交换消息。基于单车车载自组织网络提出一种高效数据迁移策略，首先建立应用处理和数据迁移系统模型，然后利用李雅普诺夫优化技术设计在线分布式系统控制框架，框架目的在于最小化单车网络传输的开销，控制框架主要研究了等待传输的积压队列稳定性和系统开销之间的平衡。根据经济学中边际成本理论，进一步将系统开销分解为通信开销和处理开销两部分，并结合控制理论，利用学科交叉方法提出队列稳定性、通信开销、计算开销之间的三方均衡优化策略。

为了测试该在线分布式优化控制方法的高效性，采用大量仿真应用程序，以真实场景中的消息传输需求为标准，模拟迁移计算效果。实验结果表明，该在线最优控制框架能够满足共享单车通信网络的多种设计需求，具有稳定

性、最优性以及创新性，对于共享单车移动网络通信有着重要的意义。

基于流特征的双重范式流数据处理

姓　　名：刘志毅
研究方向：分布式流数据处理
导　　师：吴　松
指导老师：吴　松
Email：zhiyiliu@hust.edu.cn
QQ：532328694
联系电话：15527952808
毕业去向：北京字节跳动科技有限公司

在流数据处理中，包含有五个特征，即实时性、突发性、持续性、易失性和变化性。当前主流的流数据处理框架分为两大类模型。数据流模型通过采用单条记录处理的方式以期望获得更低的延迟，仅能涵盖实时性、持续性与易失性。任务流模型通过数据批量的方式处理以期望获得更高的吞吐量，仅能涵盖突发性、持续性和易失性。这就造成了面对复杂和可变的流数据处理环境下，选用流处理框架时两难的境地，最终导致不能满足流数据处理的所有特征。

对当前流数据处理两种模型进行分析，提出了一种双重范式模型。通过将数据流模型和任务流模型在一个流数据处理程序中同时运用，以结合两种模型的优势，满足流数据处理的所有特征需求。双重范式模型用更细粒度的视角看待流处理程序。在操作层面，通过分析数据量大小，将受数据洪流影响的操作采用任务流模式执行以获得更大吞吐量并满足突发性，而未受数据洪流影响的操作采用数据流模式执行以兼顾低延迟并满足实时性，并在程序运行过程中动态调整执行模式，达到高吞吐量、低延迟的流数据处理，满足变化性特征。同时，双重范式模型包含可重放统一数据集，将数据与执行模式解耦，解决数据管理问题，满足持续性和易失性特征。

低延时的数据流处理系统的研究

姓　　名：柳　密
研究方向：云计算与分布式系统
导　　师：吴　松
指导老师：吴　松
Email：2270566005@qq.com
QQ：2270566005
联系电话：15623473694
毕业去向：浙江杭州阿里巴巴阿里云

数据流处理（DSP，Data Stream Processing）应用通常是建模为有向无环图：操作符与它们之间的数据流。尽管现在的数据中心的节点都是通过Gigabit以太网或者InfiniBand 连接，但是跨越进程的操作符间的通信开销依然远远高于一个进程内部的操作符间的通信开销。在实践中发现，操作符间的延时占到DSP应用的总处理延时的86%以上。因此，操作符间的通信对DSP应用的总处理延时有重大的影响。尽管如此，目前关于优化操作符间的通信方面的研究工作相对较少，都主要侧重于减少节点间通信量而忽略了节点内部的进程间通信（IPC）的开销。即便是目前应用最广泛的IPC框架在一个节点内部依然会有多次开销很大的内存拷贝操作以及不必要的等待时间，从而带来高延时。

通过对当前流计算系统的分析，提出了TurboStream的设计和实现。TurboStream是专为解决操作符间通信所带来的高延时的问题而设计的新型DSP系统。为了降低操作符间的延时，引入了两个功能上互补的组件：（1）改进的IPC框架。它内部集成了一个OSRBuffer。OSRBuffer是一个专门设计的面向DSP的堆外环形缓冲区。当在一个节点内部的操作符之间传输消息的时候，它可以减少内存拷贝操作并降

低单个消息的等待时间。（2）粗粒度调度器。它在调度之前会根据操作符实例间的数据依赖关系和运行时的通信量信息合并操作符实例，再将合并后的操作符实例分配到节点以减少节点间的IPC通信量。

鉴于JStorm在工业界的广泛应用以及其在低延时方面的优异表现，TurboStream的原型实现是基于JStorm。通过实验证明，改进后的IPC框架将节点内部的IPC的端到端的延时降低超过45.94%。此外，与JStorm相比，TurboStream 将DSP的平均处理延时降低了83.23%。

基于Unikernel的边缘计算卸载云平台研究

姓　　名：梅　超
研究方向：云计算
导　　师：吴　松
指导老师：吴　松
Email：meicorl@hust.edu.cn
QQ：756621494
联系电话：13349873655
毕业去向：腾　讯

针对传统虚拟机由于体积庞大，资源开销巨大而难以应用在边缘计算等资源受限场景下的问题，我们提出基于特性化内核技术——Unikernel构建超轻量级计算卸载云平台的方法。Unikernel是一种新兴的超轻量级内核虚拟化技术，相对于传统虚拟机，Unikernel具有启动快、体积小、资源开销少、安全性高等众多优势，可以满足边缘计算环境下对服务器运行时的苛刻要求。通过提出通用型Rich-Unikernel，并基于此提出内核离线编译与应用在线加载的方法，实现了将应该快速Unikernel 化并共享基础镜像以省去耗时的重复编译工作，最后将部分计算卸载场景下常用的Android系统库移植到Unikernel中，以此实现了在Unikernel中为Android 应用卸载代码提供支持。利用Unikernel的众多优势，实现了在毫秒级响应终端卸载请求，同时只占用极少的服务器资源。

实验结果表明，通过Unikernel构建的计算卸载云平台，其运行时启动速度比传统基于虚拟机的运行时快25倍以上，磁盘开销和内存开销都仅为基于虚拟机的运行时的10%左右，同时减小了云端运行时的能耗开销，极大地提高了云平台的效率并减小了云平台的负担。

cloudlet环境下基于用户能耗优化的代码迁移机制研究

姓　　名：石子倩
研究方向：云计算、任务迁移
导　　师：吕新桥
指导老师：陆　枫
Email：805985501@qq.com
QQ：805985501
联系电话：13545119517
毕业去向：腾　讯

随着移动应用数据的井喷式增长，移动设备端能耗问题日益突出，通过将移动设备端任务迁移到附近的cloudlet上执行可以有效节省能耗。但在真实应用场景中，cloudlet与用户之间的连接具有间断性，不合理的任务迁移策略会导致任务迁移效率低下和能耗浪费。因此，基于cloudlet环境下的真实任务迁移场景展开研究，提出高效节能的任务迁移算法对节省移动设备端能耗具有重要的理论意义与现实意义。

高效节能的任务迁移算法旨在用户可容忍延迟时间内，以最小化移动设备端能耗为目标，动态地决策任务迁移与否。首先，基于cloudlet资源局限性提出了基于半马尔科夫过程的cloudlet接入机制，计算出了cloudlet拒绝用户接入请求的概率，使用户能够最大化利用cloudlet端资源。接着，基于用户移动随机性推导出了任务迁移成功概率。然后，基于全局预测计算出了

用户移动过程中可迁移到cloudlet端执行的任务数量，并进一步推导出了移动设备端最优CPU运行频率。最后，提出了基于概率的动态任务迁移算法--ETOA，在用户移动过程中通过比较任务迁移成功概率与阈值概率动态地决策任务迁移与否，以达到最小化移动设备端能耗的目的。

测试结果表明，基于半马尔科夫过程的cloudlet接入机制使得cloudlet端的资源利用率始终保持在93%以上，满足了移动用户端的资源需求。且实验数据同时表明，任务迁移算法ETOA能在保证任务完成率的基础上有效节省能耗，对比其它算法能够节省高达50.5%的消耗。

网络功能虚拟化下流处理加速机制研究

姓　　名：陶　晟
研究方向：网络功能虚拟化
导　　师：胡　侃
指导老师：顾　琳
Email：523717980@qq.com
QQ：523717980
联系电话：13006311497
毕业去向：网易游戏

网络功能虚拟化（NFV）是将传统的基于专用硬件的网络功能部署到通用服务器上，从而达到减少成本，提高灵活性与可管理性的目的。NFV目前存在的两个主要问题就是如何合理地调度网络流来优化网络性能以及如何使用硬件加速技术来提高网络性能，从而达到满足高速网络需求的目的。在调度算法方面，现有的NFV网络流调度算法主要是静态的，无法适应动态变化的网络环境。在系统设计方面，现有的NFV系统大都缺乏硬件加速的支持。

针对现有工作的局限性，一方面，通过使用李雅普诺夫优化框架分析以网络系统总效用为优化目标的动态网络流调度问题，提出了一种基于数据包队列长度的动态网络流调度算法。该调度算法根据每一时刻的数据包队列长度，分别采取接纳控制决策，传输控制决策和计算控制决策来调度网络流。另一方面，设计与实现了基于DPDK编程框架的NFV网络流处理系统。该系统采用了基于提出的动态网络流调度算法的新型处理模式，并且应用了基于CUDA 编程框架的GPU加速技术来进一步提高系统处理性能。

仿真实验结果表明提出的动态网络流调度算法能够最优化网络系统平均总效用，并且在网络流流量超过网络系统承载能力时使网络系统保持稳定状态。同时，通过调节公平因子能够在公平性和吞吐率之间取得不同的权衡。系统测试结果表明相比传统的处理模式，新型处理模式能够获得更高的网络流总吞吐率。此外，根据不同的网络功能类型，使用GPU加速相比仅使用CPU处理网络流能够获得几倍到十几倍的加速比。

基于时间敏感的多源异构数据的特征描述及预测研究

姓　　名：肖柏昀
研究方向：普适计算
导　　师：丁晓锋
指导老师：余　辰
Email：xiaobaiyun@hust.edu.cn
QQ：834724608
联系电话：17771842126
毕业去向：招商银行武汉分行

特征描述作为上下文推理预测的先行步骤，它对原始数据进行一定的语义抽象并为后续推理预测模型提供直接的输入数据，所描述的特征样本数据的优劣将直接影响到最终推理预测的结果。虽然目前已有许多关于特征描述方法的研究，但都是针对图像识别、音频识别等特定应用领域，智慧交通场景下具有时间敏

感性的多源异构时空数据的相关研究尚且较为欠缺。因此依托于智慧交通场景的上述特性，尝试对出租车搭乘量进行预测并由此展开特征描述方法的研究是十分有必要的。

通过对出租车搭乘量的时序规律进行分析以及充分考虑到交通场景下的时间敏感性，提出了一种能够生成特征样本数据的特征描述方法，这种特征描述方法能够将时序预测问题转换为机器学习中的监督学习问题，从而提供一种普适的预测方案。首先，从时序角度出发考虑时效趋势特征、时效周期特征以及全局周期特征，其中全局周期特征需要进行规范化的压缩描述；随后，考虑即时天气信息，将参考时间与预测时间在天气状况上的相似度作为特征变量用于描述即时特征；最后，将时序特征与即时特征有机结合，总结出完整的特征描述方法。

在珠海市出租车真实轨迹数据上，按照特征描述方法生成对应特征样本并利用脊回归模型进行预测实验，以预测的性能指标作为特征描述方法有效性的反馈。进一步地，针对性地组合不同类别特征并进行各自的回归预测实验，从而对比并分析不同类别特征在预测时不同作用。

基于软件定义网络的虚拟网络功能服务链在线调度技术

姓　　名：肖逸凯
研究方向：软件定义网络和网络功能虚拟化性能优化
导　　师：刘方明 教授
指导老师：刘方明 教授
Email：fierralin@hust.edu.cn
QQ：1049419760
联系电话：13129953083
毕业去向：北京字节跳动科技有限公司

随着软件定义网络(Software Defined Network, SDN)和网络功能虚拟化(Network Function Virtualization, NFV)技术的发展，SDN和NFV整合网络(SDN/NFV)也在学术界和业界引起广泛关注并取得巨大发展。一方面SDN/NFV对网络技术的发展做出了巨大贡献，如软件定义数据中心、5G网络等；另一方面它们也遇到了许多性能瓶颈问题。

在SDN/NFV平台如数据中心，一个网络服务往往要经过一系列有次序的虚拟网络功能(Virtual Network Function, VNF)处理，这种有次序的网络功能组成的链状服务也被称为服务链(Service Function Chain, SFC)。而网络请求的到来和离开普遍具有频繁且随机的特点，网络流量也相应地迅速实时地变化。在这种场景下，现有的SFC调度方案往往会导致VNF或网络节点产生性能瓶颈，进而造成SFC的性能下降。因此需要对SFC进行高效的在线动态调度，降低资源占用及操作开销，同时提升接收网络请求流量。首先，为了解决上述问题，将SFC调度问题建模为一个基于马尔可夫决策过程的多目标优化问题。然后，针对SFC在线调度中网络流量变化快、频率高的特征，提出一个高效的基于深度增强学习(Deep Reinforcement Learning, DRL)的序列化回溯及特征学习的方法(DRL with Serialization Backtracking and Feature Learning, DRLSF)实现自动化、持续化地在线调度。其中，序列化回溯方法的提出是为了有效地解决SFC放置中动作空间太大的问题。而序列化回溯方法的引入会破坏SFC序列化特征，因而提出SFC特征学习方法，通过将后续未放置的VNF加入当前状态实现SFC序列特征的学习。

通过仿真实验，验证了DRLSF不仅能够在训练中快速收敛并能高效快速响应网络请求，而且相较于当前先进的研究成果，DRLSF提升了平均20.44%的接收网络请求流量，并提升了平均16.83%的网络服务提供商利润。

基于在线学习的磁盘故障预测技术

姓　　名：熊　壮
研究方向：智能运维–磁盘故障预测
导　　师：胡　侃
指导老师：吴　松
Email：820726714@qq.com
QQ：820726714
联系电话：15527597597
毕业去向：华为–CloudBU

基于在线学习方法，设计实现的具有自适应性的在线故障预测系统，能避免预测性能的衰退问题，提高故障预测模型的实用性。运用在线学习方法对磁盘故障进行预测会面临以下挑战：1）如何在线地对陆续收集的SMART样本进行标记？2）如何克服正负样本不平衡对预测模型性能的影响？针对前者，提出了样本自动在线标记方法。该方法根据磁盘故障状态，实时地对SMART样本进行标记，作为训练数据输入到在线学习模型进行模型更新。其次，对online bagging方法进行改进，通过使用两个不同参数的泊松分布函数分别作用在实时的正负样本之上，使得负样本相对正样本具有更小的概率被选中而实际地用于模型更新。考虑到在线随机森林 (Online Random Forests, ORFs) 具有可并行性、低内存需求以及更好的预测性能等优点，在线磁盘故障预测系统基于ORF实现。

实验表明基于ORF的在线预测模型能够快速逼近离线随机森林 (Random Forest, RF)，并在低误报率 (False Alarm Rates, FARs) 前提下实现稳定于93-99%间故障检出率 (Failure Detection Rates, FDRs)。相较于进行离线更新的RF模型，ORF模型能够在实现近似相同的FDR时维持更低的FAR，并且无需每隔一段时间重新训练一个新的模型。因此，基于ORF的在线学习方法适用于实际的长期使用。

云环境下面向HPC应用的容器间通信机制优化研究

姓　　名：张　驰
研究方向：云计算与虚拟化
导　　师：王多强
指导老师：王多强
Email：zhangchixtacbn@qq.com
QQ：554242516
联系电话：13387659337
毕业去向：百　度

针对同主机容器间现有网络通信机制存在的效率低和开销大的问题，考虑到HPC（High Performance Computing）应用中包含了大量的数据交换和对通信开销敏感的特点，提出了一种基于共享内存的容器间通信框架LiICC（Linux module for Inter-Container Communication）。通过在Linux内核的INET层和TCP层之间添加了一个路由层，当检测到不同容器之间开始创建TCP连接时，代之创建了一个LiICC连接，并采用基于共享内存的通信信道来替换默认的网络TCP协议。LiICC提供了对现有TCP网络程序的支持，现有应用不需要进行改写，可直接运行在配置有LiICC的容器当中，提高了容器间进程网络通信的效率。

隐私保护下的视频运动物体检测

姓　　名：曹思行
研究方向：多媒体安全
导　　师：金　海
指导老师：马晓静
Email：892939495@qq.com
Qq：892939495
联系电话：17600185776
毕业去向：欢聚时代

视频数据处理在大数据时代得到了广泛的应用，视频数据中涉及的用户隐私泄露问题也随之日益严重，视频数据的信息利用与隐私保护之

间的矛盾突显而出。为了保护用户隐私，在对视频数据加密时往往会破坏视频数据的结构，导致视频压缩效率的降低。保护隐私的同时，如何在高压缩效率下利用和处理视频数据信息具有重要的理论和现实意义。因此，隐私保护下的视频信息利用成为了一项值得研究的课题。

在对视频数据的各种应用中，视频运动物体检测是一个较常见的应用。而隐私保护下的视频运动物体检测算法的目标就是在保证视频内容不泄露的同时，能较为准确地检测视频中的运动物体。现有方法在对隐私保护下的视频进行运动物体检测时，会遇到硬件条件不易满足，算法复杂度较高，视频压缩效率不高等问题，亟需解决。

算法分为加密压缩和检测两部分。在加密压缩部分：算法采用压缩效率较高的HEVC视频压缩标准对视频数据进行压缩，同时对数据进行加密，从而保证视频数据的安全。在检测部分：算法首先在压缩域判定编码树块的状态以检测每帧中存在的运动物体，随后利用同一物体在不同帧的时空相关性将前后帧的同一运动物体连接起来，再对视频中存在的误连物体做切割。正确检测出运动物体后，用卡尔曼滤波算法生成运动物体轨迹。最后再利用运动物体轨迹去除视频中的背景和噪声。

实验结果表明，此算法能够实现隐私保护下的视频运动物体检测，并且所需硬件条件较容易满足，算法复杂度不高，检测较为准确。

基于区块链的高安全轻量级钱包设计与实现

姓　　名：邓　俊
研究方向：系统安全
导　　师：邹德清
指导老师：代炜琦
Email：1310002607@qq.com
QQ：1310002607
联系电话：17607186396
毕业去向：中科曙光（武汉）

经过近几年的发展，数字货币在金融科技领域越来越受欢迎，其总市值于2018年1月曾达到8000亿美元（其中比特币就占有率超过了30%）。不同于传统的钱包，数字钱包更加依赖于钱包的私钥，因此钱包私钥的安全性也愈发的重要。当前已有的数字钱包种类繁多，其中，硬件数字钱包是安全的，但是这很不方便，因为用户需要携带额外的物理设备；软件数字钱包很方便，但是安全性无法得到保证。同时，上述钱包都必须同步完整的区块链，然而大多数当前的移动设备不具备存储整个区块链的能力。为了解决上述问题，可以利用简单支付验证(Simplified Payment Verification, SPV)技术进行交易。然而现有的的数字钱包没有办法来保护SPV过程。因此，这对区块链中轻量级数字钱包的保护是一项值得研究的课题。

为了解决数字货币钱包的安全问题，基于Trustzone的安全区块链轻量级钱包(A Secure Blockchain Lightweight Wallet based on Trustzone, SBLWT) 实现了一种高效且实时的钱包私钥、地址以及同步和交易验证过程的保护机制。它比硬件数字钱包更便携，比软件数字钱包更安全。首先，SBLWT通过构建安全的启动序列保证SBLWT先于普通操作系统启动，并通过不可屏蔽中断技术保证安全的切到可信执行环境中。其次，将钱包的公私钥以及钱包地址的生成过程置于安全执行环境中，同时私钥被存储在安全存储中。然后，将本地存储的信息加密存储在非安全存储中，并在安全执行环境中完成钱包的同步和交易验证过程。最后，通过在安全执行环境中构建安全的输入输出驱动保证可信的人机交互。

为了验证SBLWT的有效性，将SBLWT部署在RASPBERRY PI 3 MODEL B开发板上。其

中，SBLWT的核心代码不到3000行，不会大量的增加安全环境的可信基；同时关键模块平均总运行时间低于20ms，不会过多的增加钱包的时间开销。

基于多特征分析的漏洞自动化识别研究

姓　　名：邓之珺
研究方向：漏洞挖掘与识别
导　　师：郑德清
指导老师：郑德清
Email：Elaine@hust.edu.cn
QQ：506012274
联系电话：15527287870
毕业去向：腾讯安全平台部

为适应快速的版本发布周期，软件开发商大多遵循持续集成和持续交付的开发模式，且越来越依赖开源库来快速完成开发任务。然而，由于大多数开发人员缺乏安全专业知识，大量的安全漏洞没有被正确地识别，或者没有公开披露。这些未识别的漏洞让开发商的产品陷入被恶意攻击的风险中。因此，迫切需要一个高效的漏洞自动化识别系统，来识别开源库中未知的漏洞并确保安全的软件开发过程。

各大软件开发商大多借助缺陷追踪系统完善其软件开发过程，通过审核缺陷追踪系统中的缺陷报告可以高效地检测漏洞。然而，现有的面向缺陷追踪系统的漏洞识别方案大多只考虑了缺陷报告的单一特性，导致其漏报和误报严重。

基于多特征分析的漏洞自动化识别方案SBRer（Security Bug Report Identifier）有效地解决了上述问题。SBRer提取了缺陷报告中多种有效信息，包括预定义非文本字段（元特征），缺陷报告的文本内容（文本特征）和缺陷补丁文件的代码属性（代码特征）。基于这些特征，SBRer通过自然语言处理技术和机器学习算法构建了一个自动化识别漏洞的模型。此外，SBRer还充分考虑了数据的不平衡现象，提高了漏洞识别精度。

实验结果表明，SBRer在识别漏洞时精确率和召回率分别达到了99.4%和79.9%。与现有的自动化识别方案相比，SBRer在保持高精确率的同时，将召回率提高了22.9%~175.5%。

FaaS云计算中基于SGX的安全保护机制研究

姓　　名：董泽晗
研究方向：系统安全
导　　师：羌卫中
指导老师：羌卫中
Email：2283296285@qq.com
QQ：2283296285
联系电话：15927039826
毕业去向：百　度

函数即服务云计算是当前新兴的一种云计算模式，具有无需维护、规模自动伸缩以及事件驱动等特点。一方面，函数即服务云计算面临着与传统云计算模式相同的安全问题，如服务提供商非法获取计算过程中隐私数据、攻击者恶意篡改应用代码等。另一方面，由于函数即服务云计算平台允许用户部署任意代码以及服务运行时容器环境的弱隔离性，用户的恶意代码会破坏服务运行时，从而导致云端环境面临的安全挑战更加严重。因此，需要设计一种针对函数即服务云计算的安全保护机制。

基于SGX保护的函数即服务云计算框架，利用Intel软件保护扩展SGX保护云服务运行时和API网关等重要组件。通过在服务运行时中引入基于SGX的双向沙箱机制，可以防止服务提供商非法获取计算过程中隐私数据，并且保护运行时不被用户的恶意代码破坏。在双向沙箱的基础上，增强了对沙箱内函数模块的权限监控以防止越权执行，并且优化了双向沙箱的执行周期以提高服务运行时性能。针对应用代码

面临被攻击者恶意篡改的威胁，设计并实现了基于SGX认证机制的应用代码完整性验证方法。最后，为防止API网关计算过程中隐私数据被窃取，利用SGX隔离API网关中的重要模块。

实验结果表明，基于双向沙箱的服务运行时可以保护计算过程中隐私数据不被泄露并且防止用户的恶意代码进行破坏。应用代码完整性验证可以有效地检测到代码是否被篡改。基于SGX的API网关隔离保护机制可以防止攻击者窃取网关计算过程中隐私数据。双向沙箱对服务运行时执行时间的影响较小。

云环境信息流分析与攻击防御研究

姓　　名：刘本熙
研究方向：安　全
导　　师：金　海
指导老师：邹德清
Email：lbxhust@163.com
QQ：727934274@qq.com
联系电话：15997408048
毕业去向：阿里巴巴

在云计算技术广泛应用的同时，云环境的安全也面临着诸多挑战。其中，代码复用攻击是一种应用普遍、危害巨大的攻击方式。通过代码复用攻击，攻击者能劫持云环境中的服务器和应用，这给云服务商和用户带来了巨大威胁。云环境信息流分析与攻击防御系统IFAADS（Information Flow Analysis and Attack Defense System）针对代码复用攻击的特点，在内存揭露阶段防御代码复用攻击。内存揭露必须利用程序中的代码页或代码指针，它包括两种类型：基于读取的内存揭露和基于执行的内存揭露。为了阻止不同类型的内存揭露，IFAADS系统基于数据流和控制流两类信息流进行了综合性的分析和检测，并利用隔离区机制保护程序中的代码页和代码指针。在编译阶段，IFAADS系统对程序进行分析，为利用隔离区保护代码页和代码指针提供必要信息，并对程序和共享库进行修改，使它们能够支持隔离区机制；在运行阶段，IFAADS系统对程序中的数据流和控制流进行检测，利用隔离区保护程序中的代码页和代码指针，阻止攻击者进行内存揭露。IFAADS系统利用了x86-64平台上高效的Intel MPX（Memory Protection Extension）指令，并对检测指令进行了优化，减少了性能开销。在IFAADS系统的保护下，攻击者无法通过内存揭露获取gadget。测试结果表明，IFAADS系统能够有效阻止基于读取的内存揭露和基于执行的内存揭露并防御代码复用攻击，并且在SPEC CPU2006基准测试集上只引入5.8%的平均性能开销。与现有方法相比，IFAADS系统能够在提供完善保护的同时，仅引入较小的性能开销。

基于动态分级的网络服务安全共享机制研究

姓　　名：卢　宇
研究方向：网络空间安全
导　　师：邹德清
指导老师：邹德清
Email：595129882@qq.com
QQ：595129882
联系电话：13517242129
毕业去向：行吟信息科技（上海）有限公司

随着网络的高速发展，传统网络架构的不足逐渐暴露出来，已无法满足持续扩大的流量需求。为了探索新的网络架构，SDN（Software Defined Network，软件定义网络）通过将控制平面和数据平面分离，使网络的控制能力和可扩展性得到了空前的提升，为网络架构带来了新的理念和活力。另一方面，NFV（Network Function Virtualization，网络功能虚拟化）技术旨在采用虚拟化技术将传统网络功能设备软硬件解耦，以提升设备的部署、管理效率。SDN

与NFV结合的趋势越来越明显，但新的网络架构也带来了新的安全问题，如何能有效地管理SDN以及NFV的开放接口，安全地共享SDN/NFV所提供的网络服务，目前少有相关研究。

针对多用户SDN/NFV网络的特点，基于动态分级的细粒度权限管理框架设计了新的权限管理方法，能够有效防止控制器API（Application Programming Interface，应用程序编程接口）被滥用，保护网络服务的安全共享；提出了与用户权限紧密结合的网络划分方法，以满足用户动态变化的网络需求。框架的权限策略描述语言将用户权限抽象为设备权限、API权限以及域权限三个层级以便能够在合适的粒度上定义用户行为边界。框架由权限管理器和实时调解器两个组件组成，权限管理器主要管理权限策略的输入与解析，为网络管理员提供灵活、准确配置用户权限的接口，并且当用户权限变化导致拓扑变化时，检测用户网络连通性以保证资源设备可达；实时调解器根据权限要求验证所有API调用的合法性，管理维护用户的网络拓扑视图。

在RYU控制器上实现的原型系统及实验结果表明，细粒度权限管理框架能够确保所有的API调用都符合权限要求，从而有效地保护控制器安全，并且只会对控制器造成启动时间上3ms的额外开销以及核心API不到0.1ms的处理延迟，是可忽略的。

监控视频密文运动检测技术研究

姓　　名：彭　欢
研究方向：网络空间安全
导　　师：徐　鹏
指导老师：马晓静
Email：826886605@qq.com
QQ：826886605
联系电话：13164674849
毕业去向：中国舰船研究设计中心

云视频监控系统因其普遍性、可扩展性和稳健性而受到越来越多的关注。但是，云端存储的监控视频包含很多个人或者组织的隐私信息，这些信息面临着被窃取的危险。因此云端应该保护监控视频免受隐私泄露，目前常采用的做法是在监控视频上传到云端之前对其进行加密；另一方面，利用云的计算能力来处理和分析监控视频是一种趋势，其中运动检测和追踪是云视频监控系统的重要任务。

总之，云视频监控系统需要保护视频隐私的同时仍允许云端进行运动检测和追踪。现有技术允许在加密视频进行运动检测和追踪，其中一种解决方案在像素域进行运动检测，但是过高的计算复杂度很难满足当前应用的即时性。近来一些解决方案提出在压缩域上进行加密和运动检测，但是从加密视频中检测的运动轨迹所有细节完全被暴露。监控视频密文运动检测技术研究在最新的H.265/HEVC视频编码技术基础上提出了一种支持粒度可调整加密视频运动检测方案，对隐私保护和运动检测精度之间的冲突要求进行灵活的粒度控制。方案主要包括视频加密和加密视频运动检测两部分。视频加密算法在码流选择性加密算法的基础上添加了扰乱层次的加密，包括CTU位置扰乱和CTU内部扰乱，加密算法能够满足码流格式兼容性以及特征信息保留性，以至于能通过标准解码器解析得到统计特征信息；同时该算法可逆，以确保授权用户能够从加密视频中解密还原原始视频。加密视频运动检测算法包括信息预处理、局部方向预测和全局方向预测，最终推断运动物体的轨迹。实验结果表明系统在运动检测和跟踪方面具有较高的精度，并且实现了检测粒度可调，同时加密算法对底层视频压缩效率影响极小。

面向云辅助物联网环境的公钥可搜索加密研究

姓　　名：唐晓兰

研究方向：密码学
导　　师：徐　鹏
指导老师：徐　鹏
Email：2279777383@qq.com
QQ：2279777383
联系电话：17764007268
毕业去向：汇添富基金管理股份有限公司

云辅助物联网平台是现在流行的系统模型，它融合了云计算和物联网的多方面优点，在实际应用中，云辅助物联网平台中存储了大量的用户隐私信息。通过将用户信息加密之后再上传至云端，能有效地保护用户的隐私安全，但是对云端数据进行加密之后，服务器需要面临的一个问题就是如何对加密数据进行检索。

具有隐藏结构的公钥可搜索加密（SPCHS）在可搜索公钥加密领域第一个实现了检索时间复杂度仅和密文数量亚线性相关。但是该方案由于密文结构限制，不能实现密文的并行检索，因此不能大范围推广。

为了解决这个问题，提出了在随机预言机模型下具有语义安全性的并行化SPCHS方案。和之前的方案对比，并行化SPCHS方案消除了同一个关键字的可搜索密文之间的隐藏链式结构，在可搜索密文之间构造了一种新型的隐藏关系，所有的可搜索密文与公有头部构成了一个星型结构。一旦服务器收到关键字检索陷门之后，就能并行地暴露该关键字所有的密文和公有头部之间的隐藏关系并快速地匹配到所有的密文。因此，并行化SPCHS方案能实现快速并行地检索密文。

在实验过程中，基于PBC函数库和POSIX多线程编程技术，对两个SPCHS方案的检索效率、密文生成时间和通讯开销等方面进行了全面的对比。根据实验数据显示，并行化SPCHS方案提升了约五倍的检索效率。除此之外，鉴于物联网设备的性能存在一定的限制，并行化SPCHS方案中构造了更加高效的加密算法，实验结果表明该算法能节省40%的加密时间和90%的通讯开销。

基于中间语言分析的代码指针完整性保护研究

姓　　名：王世振
研究方向：系统安全
导　　师：羌卫中
指导老师：羌卫中
Email：2547102970@qq.com
QQ：2547102970
联系电话：13129933117
毕业去向：美团点评

代码重用攻击利用应用程序的内存漏洞劫持应用程序的控制流，并重用程序进程内存空间中的代码片段实施攻击。近年来学术界针对代码重用攻击设计了多种防御技术，包括地址空间随机化、代码指针完整性和控制流完整性。地址空间随机化和代码指针完整性可以通过内存泄漏攻击绕过。控制流完整性是一种比较完善的防御方法。但是，由于现有控制流完整性在保护控制流图前向边的不足，基于间接函数调用的代码重用攻击依然可以绕过其保护。

针对现有控制流完整性在保护控制流图前向边的不足，提出了一种基于指针分析的控制流图前向边的细粒度控制流完整性方法：P-CFI。P-CFI更加严格的保护间接函数调用，保护应用程序免受基于间接函数调用的代码重用攻击的威胁。P-CFI包括静态分析、插桩和运行时检查三个阶段：静态分析阶段基于指针分析为应用程序每个间接函数调用构建合法目标函数集，插桩阶段为每个间接函数调用插入检查代码，运行时检查阶段检查间接函数调用的目标是否在合法目标函数集中。

P-CFI基于LLVM实现了原型系统，并评估了它的功能和性能。安全分析证明，P-CFI可以

防御基于控制流图前向边的代码重用攻击。性能评估表明，P-CFI可以保护目标程序免受基于控制流图前向边的代码重用攻击的影响，时间开销在0.06%到3.25%之间。

基于细粒度动态行为检测的安全云服务构建机制研究

姓　　名：夏　妍
研究方向：云安全
导　　师：金　海
指导老师：代炜琦
Email：sum1993@126.com
QQ：2276158846
联系电话：13006376269
毕业去向：深圳前海微众银行

随着云服务结构日趋复杂，在云中发现的漏洞也越来越多，目前在部署最为广泛的基础设施云平台OpenStack中就已发现173个漏洞。这些漏洞充分证明了云服务实际上并不安全，并且用户充分信任云服务，授予云服务全部的权限。一旦云服务被入侵，攻击者可以执行任意操作，引发的安全威胁不容小觑。

基于细粒度动态行为检测的安全云服务构建机制以OpenStack为例，实现对云服务漏洞攻击的主动防御。首先，通过分析大量OpenStack设计文档与CVE官网的漏洞说明，综合考虑安全需求和风险等级，定义了OpenStack五个核心组件（KeyStone、Nova、Cinder、Glance、Swift）中的安全敏感数据和安全敏感操作，为云服务安全研究提供细粒度的保护目标。其次，根据安全敏感数据和操作列表，在云服务处理用户请求前分析执行用户请求涉及的安全敏感事件，即安全敏感数据访问事件和安全敏感操作执行事件，便于后续保护。再次，结合用户请求与用户权限检查请求的合法性，只允许云服务处理符合用户授权的请求，并在云服务运行时限制其权限，避免攻击者利用云服务额外的权限实施恶意操作。最后，当云服务运行至安全敏感事件时，根据事件的安全性需求提供有针对性的保护措施，在有效保护安全敏感数据和安全敏感操作的同时维持较小的性能开销。

测试结果表明基于细粒度动态行为检测的安全云服务构建机制可以提供针对云服务漏洞攻击的安全防护，有效保护云租户隐私数据和云服务安全，并且敏感信息分析和权限控制过程的平均运行时间约为584μs。相较于其他系统，有效防御的漏洞数目更多，产生的性能损耗更小。

面向内核提权攻击的数据完整性机制研究

姓　　名：杨嘉玮
研究方向：系统安全
导　　师：石宣化
指导老师：羌卫中
Email：807764383@qq.com
QQ：807764383
联系电话：18805569008
毕业去向：珠海西山居

研究的目标是开发一个能够阻止篡改内核敏感数据的非控制数据攻击的防御机制，PrivGuard系统。为了防御提权攻击，PrivGuard系统通过阻止对内核敏感数据的非法写操作，来对这些数据实施数据完整性策略。PrivGuard系统在没有变更现有的Linux访问控制机制的情况下，修改了系统调用入口点来监视内核敏感数据的修改。接着，检测内核栈是否耗尽，并抽取敏感数据，将其保存在内核栈上。然后使用stack canary保护储存在内核栈上的数据不被攻击者篡改。最后，使用这些保存的数据来对内核敏感数据进行完整性检测。其次，为了确保敏感数据使用和检测时的上下文一致，PrivGuard系统引入一个结构体将敏感数据和上下文绑定在一起，并且在每个系统调用执行前验证上下文是否反生改变，从而确保了敏感数

据的检查时刻到使用时刻一致性。

系统原型的实验测试结果表明，面向直接内核攻击的防御系统可以有效防御针对Linux内核的特权提升攻击。系统性能测试显示系统性能开销在可接受的范围内，对系统核心服务的开销平均值为9%。同时，应用程序测试表明系统对于用户应用程序的影响很低，应用程序的开销均少于1%。

云环境下隐私侵犯行为监控及取证方法研究

姓　　名：赵　健
研究方向：云计算安全
导　　师：邹德清
指导老师：邹德清
Email：bertcug@gmail.com
QQ：371418912
联系电话：15002770383
毕业去向：华为武研所

云计算的高拓展性与按需服务的特点极大地改变了以往的计算与存储模式，使得越来越多的企业机构将自己的服务部署在云平台中。然而云计算技术也存在潜在的安全风险，云计算数据安全与隐私保护机制的不健全使得云环境下隐私侵犯问题层出不穷。云平台下的隐私侵犯行为具有泄露痕迹易销毁、隐私数据难追踪等特点，已有的云取证方案存在粒度粗、拓展性差等问题，且忽略了隐私侵犯细节与行为特征，无法对隐私侵犯行为进行细粒度刻画。

针对上述问题，结合连续内存镜像分析与动态污点追踪对隐私侵犯问题进行多粒度的行为取证是一种有效的解决方案。第一步，从网络流量中截获可疑应用和异常流量。第二步，通过VMI技术获取目标虚拟机软件环境配置信息，依据该信息自动化生成模拟虚拟机环境，并将可疑应用或异常流量导入该环境执行。第三步，使用连续内存镜像技术对执行过程进行粗粒度取证分析。第四步，隐私信息标记为污点，通过动态污点追踪技术对执行过程进行细粒度取证分析。

测试结果表明：（1）与现有安全工具相比，系统可以检测到更详细的隐私侵犯行为。（2）系统能够有效的检测到由恶意软件或漏洞攻击导致的隐私侵犯行为，能够获取键盘记录、敏感文件和敏感内存的隐私侵犯行为的详细隐私侵犯路径。

大规模图上介数中心度分布式算法研究

姓　　名：艾　明
研究方向：大数据
导　　师：石宣化
指导老师：华强胜
Email：aim@hust.edu.cn
QQ：392601611
联系电话：13072783683
毕业去向：武汉小红书

在复杂网络分析中，中心度是一种非常重要度量方法，它能够量化一个节点相对于其它节点在网络中的重要程度。依据要求解节点重要程度类型的不同，中心度有多种不同定义，其中介数中心度（Betweenness Centrality）刻画不同社团之间节点的重要性。依据消息传递方式的不同，介数中心度可分为最短路径介数中心度（Shortest Path Betweenness Centrality）以及随机游走介数中心度（Random Walk Betweenness Centrality）。最短路径介数中心度假设图中消息仅基于最短路径传播，而随机游走介数中心度假设图中消息随机传播。

针对在CONGEST模型下计算图中所有节点的最短路径介数中心度问题，一个$O(n)$（其中n为图中节点个数）时间复杂度的分布式算法被提出。CONGEST模型是被广泛应用于分布式计算领域的算法模型，该模型假设每一轮中每一条

边至多只能传递O(log$n$)比特。同时，针对该问题在CONGEST模型下的求解，得到Ω(n/log$n$+D)的下界，其中D代表图的直径。证明了算法的高效。

针对在CONGEST模型下计算图中所有节点的随机游走介数中心度问题，一个时间复杂度为O($n$log$n$)的分布式近似随机算法被提出，该算法计算出每个点的随机游走介数中心度为该点精确的随机游走介数中心度的(1-ε)倍（ε为一个介于0到1之间的数），并且分布式精确求解该问题的下界被证明为Ω(n/log$n$+D)。证明该问题无法在亚线性时间内求解，从而证明提出算法的高效。

基于社区发现的高效在线社交网络事件流传播研究机制

姓　　名：行方家

研究方向：大数据，图计算

导　　师：陈汉华

指导老师：陈汉华

Email：1969861017@qq.com

QQ：1969851017

联系电话：13260681969

毕业去向：湖北联通

在过去的十几年里，社交网络应用迅速普及。人们几乎每一天都通过社交网络与自己的朋友进行互动，社交网络在人们的生活中占据重要位置。在大型在线社交网络系统中，用户相关的数据是以用户为单位的视图方式进行存储的，而用户及其好友的视图，通常分布在不同的服务器上，那么对于拥有大量活跃用户的社交网络，由于用户间错综复杂的社会联系，会使用户间在进行事件流通信时产生大量服务器间的通信开销。

为了解决这个问题，现有的方法通常利用社交图的结构，来减少用户间社会联系产生的多余的服务器间通信开销。目前最先进的方法，基于对在线社交网络中用户间通常存在很多共同好友这一现象的观察，通过充分的利用一种提出的中心结构来减少服务器间通信开销。为了找到最好的中心结构，这种方法需要通过迭代地去除掉权度最小的点，识别出全局最密集的子图。这样的一个过程，会导致最坏计算复杂度达到$O(n^2)$，这就使得这种方法无法扩展到真实世界的大型在线社交网络中。

为了解决传统方法不可扩展的问题，提出一种基于社区发现的高效在线社交网络系统事件流传播机制。首先，利用一种有效的基于动态距离的社区发现算法，将整个社交图划分为若干个联系相对紧密的社区。对于每个社区，设计一个启发式算法来充分利用一种中心结构。这种算法在每一轮迭代中，寻找以度最大的点为中心点的中心结构。在收集了大规模的数据集后，对新方法进行了全面的实验进行评估，实验结果表明，所提出的方法与现有方法相比，既能够大量减少通信开销，又大大降低了计算时间，是一种有效且高效的在线社交网络系统事件流传播机制。

基于程序分析优化大数据应用内存配置的研究

姓　　名：柯志祥

研究方向：内存计算，分布式系统

导　　师：石宣化

指导老师：石宣化

Email：iceke1025@gmail.com

QQ：331317953

联系电话：15172353431

毕业去向：阿里巴巴

大数据平台上基于内存计算模型的分布式数据处理系统，如 Flink和Spark，经常遭受严重的内存压力，尤其在平台内存资源紧张、且被多个用户或组织共享情况下，内存资源竞争进一步加剧。用户应用被分配的内 存空间不足，

会在运行期间产生严重的垃圾回收 (GC) 开销，而分配过量的内存会导致平台资源的浪费，因此平 台中如何为用户应用配置合适的内存成为关键性问题。

通过分析发现，是平台上的多个应用会多次共同处理某个特定的数据集，常常这些应用对数据的处理逻辑是相似的，如机器学习和图计算应用。同时，大数据应用对数据的处理流程即数据路径体现在框架的算子API与用户自定义方法(UDF)中，继而影响运行时内存的使用。针对问题与发现，提出了一种预估新提交应用的合理内存阈值的方法。这种新型的方案结合并利用了程序分析与历史应用处理数据时产生的特征信息，它基于对程序代码的分析来追踪缓存数据的数据路径，构建出数据结构kTree。当新应用提交时，将新应用的数据路径与历史应用数据路径进行匹配。当匹配成功时，利用预估模块内存中的算法来预估出用户应用的合理内存阈值。从而保证在不浪费大数据平台内存资源的情况下，充分发挥出应用最佳的性能。

提出的预估与优化方法基于Spark系统实现，对用户透明，不需要改动现有的Spark编程接口。通过一系列实验评估预估的准确性和性能收益，实验结果表明本方法预估应用的结果与真实合理内存阈值的误差比例低至4%，预估过程所产生的开销与应用真实运行时间相比可以忽略不计，平台上数据处理应用整体执行时间减少至56%，提升性能明显。

异构图模式近似匹配方法研究

姓　　名：李高峰
研究方向：图计算，图匹配
导　　师：袁平鹏
指导老师：袁平鹏
Email：peakergfli@qq.com
QQ：1553685715
联系电话：13212778039
毕业去向：阿里巴巴

随着互联网的飞速发展，大量的知识库可供公众访问。不同于传统的信息检索中使用的文档库，知识库以三元组（主语，谓语，宾语）的形式整合细粒度的信息，因此它在整合信息，加强智能搜索方面担任非常重要的角色。虽然SPARQL查询是知识图的标准接口，但是复杂的语法和知识图结构使得它不能直接为普通用户所用。因此，为知识图设计一个高效且方便用户使用的访问接口成为人们亟待解决的问题。

异构图模式近似匹配方法研究，提出了一种以图模式近似匹配的方法来访问知识图的方法。为了方便用户使用，该研究提出了以自然语言问句为访问接口。为了更加精准的挖掘出查询目的，提出了一种实体驱动的策略来构建查询图。首先识别出问句中的查询实体，然后通过基于单词之间语法关系的策略来挖掘实体之间的语义关系，基于此可以实现查询意图的图模式表达，从而将问题转化为基于知识库的异构图模式近似匹配。由于在从知识库中获取近似匹配子图的过程中会引入歧义，所以提出了用语义向量的方法来分别表示查询图中的边和知识图中的路径的语义，并基于向量之间的内积来计算它们之间的近似程度，从而筛选出正确匹配结果。

通过在QALD系列测试标准下的大量严格测试，该方法在回答自然语言的准确率和召回率以及F-1系数方面优于当前最好的系统。实验表明，该方法的在构建查询图和图匹配相似度衡量方面比当前最好的系统gAnswer有更好的效果。

基于GPU加速大规模文件系统富元数据查询研究

姓　　名：李文珂
研究方向：大数据
导　　师：石宣化

指导老师：石宣化
Email：wenkeli@hust.edu.cn
QQ：982306909
联系电话：13006378856
毕业去向：招商银行

HPC系统的富元数据被用来描述数据文件丰富的元数据信息，如产生数据文件的执行情况，环境变量以及执行的参数等。由于HPC系统富元数据的丰富性，富元数据的管理场景，如数据审核和起源查询等，需要高效查询的基础架构。最近有研究表明，基于属性图遍历模型查询的模型管理富元数据是可行的，但是由于CPU的硬件特性很大程度上被内存访问速度和计算能力限制，结合新型加速硬件优化富元数据管理成为一种有效的方式。现代GPU作为一种新型加速硬件，具有较高的内存带宽，且能提供较高的并发度。然而基于GPU实现富元数据图遍历及查询通常具有以下挑战。

本研究设计并实现GRAM系统，一个基于GPU实现的富元数据属性图遍历和查询的框架。GRAM系统的设计侧重于减少内存访问开销，并提高GPU的程序效率和GPU利用率。具体而言，富元数据图不同于普通的图数据，其附加属性导致遍历查询需要处理的数据量可能非常大，因此，GPU的高内存带宽有助于减少内存访问延迟并提高内存访问效率。除此之外，由于富元数据图处理的数据单元是独立的，因此GRAM系统可以充分利用GPU的存储资源和并行性进一步提高富元数据管理性能。

实验结果表明，GRAM系统可以有效地应用到HPC系统中的用户场景中，尤其当遍历步数较大且筛选条件较多时，GRAM系统能显著提高富元数据图遍历的效率且能大大提高元富数据管理的性能。筛选过程并行优化了GRAM 系统的并行化，基础操作合并的优化策略比原始富元数据图遍历系统节省了34%-67%的时间。

基于纯策略博弈的边流图划分研究

姓　　名：李阳阳
研究方向：流图划分
导　　师：华强胜
指导老师：华强胜
Email：yangyli@hust.edu.cn
QQ：1287836778
联系电话：13006322204
毕业去向：今日头条

在大图上开展图计算，图划分是一个至关重要的先序步骤。已有的图划分模型包括离线图划分和流图划分。在传统的流图划分模型中，当前边（顶点）需要根据先前到达顶点的划分块选择信息帮助当前边（顶点）选择最佳划分块。因此，在传统流图的划分模型中很难并行地开展划分任务。另一方面，离线图划分模型在划分过程中需要知道图的全局信息，很难适用于大规模图。

针对现有图划分模型的不足，提出了一种近似流图划分模型。并基于该模型提出了一种基于纯策略博弈的边图划分算法。具体地，通过将图的边流划分成若干批次，在每个批次内将图划分问题转化为一个博弈过程。批次内的每条边被视为博弈的玩家，每条边的划分块选择被视为其策略。从而将原图的划分问题分解为一系列寻找纳什均衡的过程。每个批次内的边选择其最佳划分块时，只依赖该批次内其它边的划分块选择，所以各个批次内的博弈过程可以并行地寻找纳什均衡。首先通过设计适当的个体代价函数和社会福利函数构造一个博弈过程，并证明了该博弈过程是一个确切势博弈，从而一定存在纯策略纳什均衡。然后基于最佳动态响应提出了一个能快速收敛到纯纳什均衡的算法。最后经过实验论证，在多个真实图数据集和随机图数据集上基于博弈的划分算法和已有的流图划分算法相比，顶点平均备份数指标下降百分比最高为31.8%，边数标准差最好情况下降44倍。

动态可扩展Cuckoo过滤器设计与应用

姓　　名：廖良望
研究方向：大数据
导　　师：陈汉华
指导老师：陈汉华
Email：253261184@qq.com
QQ：253261184
联系电话：13797070594
毕业去向：华　为

实际应用中大规模动态数据集合的涌现给近似集合成员判定技术带来了日益严峻的挑战：1）集合成员表示数据结构的容量应该支持灵活的扩展和缩小来适应集合大小的动态变化；2）集合成员表示数据结构应该支持可靠删除操作。现有的近似集合成员判定技术，例如，Cuckoo过滤器、布隆过滤器以及布隆过滤器的一系列变种均不能同时满足动态集合的上述两个要求。

为了解决这个问题，设计了动态Cuckoo过滤器。动态Cuckoo过滤器是一种支持可靠删除操作，同时支持弹性容量伸缩的结构，可以高效的进行集合的表示和近似集合成员判定。促使动态Cuckoo过滤器高效性的因素有两点：第一，动态Cuckoo过滤器所使用的数据结构具有可伸缩性，并且使用了通过基于“The power of two choice”思想的重定位技术解决了哈希碰撞带来的低空间使用率问题，使得在动态集合成员的表示中动态Cuckoo过滤器具有很高的空间效率。第二，动态Cuckoo过滤器使用一种独占的指纹来表示数据元素的存在性以此保证可靠的删除操作。

实验结果表明和目前最高效的近似数据集合成员判定技术相比，动态Cuckoo过滤器减少了75%的空间消耗，50%的构建时间，并且将近似集合成员判定速度提升了80%。通过将动态Cuckoo 过滤器应用在数据去重模块上并且测试，实验结果进一步证实了动态Cuckoo过滤器的高效性。

大规模图中围长求解分布式算法研究

姓　　名：钱立祥
研究方向：分布式图算法
导　　师：华强胜
指导老师：华强胜
Email：fancyqlx@163.com
QQ：810773948
联系电话：18655628040
毕业去向：今日头条

图是计算机科学中常见的数据结构，对图性质和结构的研究可以帮助研究人员分析网络结构、描述现实问题或进一步进行图算法的设计。在大规模图处理领域，消息传递模型被认为是最适用于图算法设计的一种模型，在这种模型中，图中点和点之间通过消息来传递信息。为了使图算法具有更强的表达能力和实际意义，研究者们将消息传递模型添加带宽限制，这种模型被称为拥塞模型。在拥塞模型中，研究者希望可以深入的讨论分布式图算法在拥塞模型中遇到的困难以及其解决方法。针对这个目标，可以通过围长问题(the weighted girth) 为切入点，讨论如何在拥塞模型中分布式计算不同图模型的围长，其中图模型包括有向图、无向图、有权图、无权图以及它们的组合。

对于无权图，围长指边的条数最小的环的长度，对于有权图，围长指所有环中边的权值之和最小的值，围长对应的环叫做最小环。一个经典的算法通过合理的调度策略启动n个宽度优先搜索来计算围长，这种方法可以在无权图上取得O(n+D)轮的时间复杂度，其中n为图中点的个数，D为图的直径，但是通过分析可以发现这种方法并没有充分利用带宽，因此结点重命名和多路复用技术被提出，通过这两个技术可以在O(D+n/B)轮时间计算出无权图的围长。对于有权图，根据最小环的特殊性质，提出了全新的思路求解围长，有效的避免了大量拥塞，使得时

间复杂度达到了O((n $\log^2(nW)$)/B+D $\log(nW)$)轮，其中W为图中边的最大权值，B为带宽。另外，通过研究在特殊网络拓扑下分布式求解围长的方法，更进一步的讨论了路由和拥塞的关系。最后，为计算有权图围长问题构造了Ω(D+n/B)轮时间的下界，这证明当W=O(n^c)，c为常数时，新的围长算法的时间复杂度达到了近似最优。

面向文档过滤的语言模型平滑方法研究

姓　　名：田泽亮
研究方向：信息检索
导　　师：赵　峰
指导老师：赵　峰
Email：823194740@qq.com
QQ：823194740
联系电话：17720500148
毕业去向：微众银行

丰富的网络资源中蕴含着海量的数据信息，帮助用户从中快速、准确的找到所需的信息是一项极具价值的任务。但是海量的数据规模以及自然语言表达带来的语义歧义性和多样性，给信息系检索带来了巨大的挑战。为了帮助用户获取有价值的信息，其中一个非常关键的问题就是如何准确地评价文档资源和用户需求之间的相关性。而传统的基于语言模型的评分方法中，只考虑单词的频率信息，缺乏语义性分析，使得单词不匹配但语义相关的文档不能被检索到，降低了检索系统的性能。因此，帮助信息系统从语义角度衡量文档和查询语句的相关性是一个亟待解决的问题。

为了衡量查询语句与文档之间的语义相关性，提出基于实体的语言模型平滑方法。实体比单词拥有更良好的定义。知识库中包含了实体较为全面的信息，而文档的语义主题可以用文档中的实体以及它们之间的关系来表示，所以这些实体在知识库中的背景知识信息较为全面的覆盖了文档主题相关信息。因此，以实体为桥梁，以维基百科中的实体信息为内容，提出了符合文档语义主题下单词概率分布的实体语义语言模型。然后提出两层次的平滑方法，结合文档无关的全局语料库信息源和文档主题相关的实体语义语言模型信息源来对原始文档语言模型进行平滑，使得平滑后的语言模型能够很好的衡量查询语句和文档之间的语义相关性，提高了检索系统的性能。

在ClueWeb09B公开数据集上进行测试。实验结果表明，相对于传统的基于的语言模型平滑方法，基于实体的语言模型方法平均能够提高10.18%的检索性能，尤其是能够检索到在单词不匹配的情况下的与查询语句相关的文档。

基于隐私保护的大图数据查询处理研究

姓　　名：王　摧
研究方向：隐私保护
导　　师：丁晓锋
指导老师：丁晓锋
Email：305172521@qq.com
QQ：305172521
联系电话：18060977698
毕业去向：阿里巴巴

图数据作为非结构化数据的代表，其在大数据分析挖掘等具体应用领域具有广阔的研究前景。然而海量的图数据中通常拥有大量的用户隐私信息，数据拥有者在将数据发布给第三方之前必须要对数据进行匿名化处理，防止个体敏感信息泄露，因此图数据的隐私保护是数据发布和数据分析处理中的一个必要环节，保证匿名化后的图数据仍旧具有高可用性具有重要意义。

针对图数据的隐私保护问题面临的突出挑战及现有研究的特点与不足，基于k匿名机制的k-dec算法在k-iso与k-auto算法的基础上通过

均匀分割，连边保留，子图同构三个步骤，提高了匿名数据的可用性，完善了匿名图的安全性，同时，算法大幅度地缩减了匿名化操作的整体运行时间。为了进一步提高匿名图的可用性，SNAM模型在对图结构分类的基础上，采用社区划分方式构建了一套完整的图数据匿名保护框架，不仅能够降低匿名图的噪声量，而且缩减了大图匿名化时间。同时，基于该匿名模型下，普通用户能够实现匿名图子图查询等具体应用。在隐私保护方案设计的基础之上，图数据可视化分析系统基于SAAS思想，采用MVC构架设计，实现了图数据在线匿名化及可视化操作，同时系统提供随机图生成，子图搜索，最短路径搜索及图表分析等功能。

实验表明，在不同的k值设定，不同的数据集规模下，k-dec匿名算法运行效率均大幅度超过同类匿名算法，而且匿名图在信息损失评估指标IL上达到了75%左右，具有极高的可用性。针对社会化网络数据集的实验中，SNAM模型相比于k-dec匿名算法降低了匿名化的执行时间，IL值也同样稳定在75%左右。匿名图的子图查询实验表明经过 k-dec算法及SNAM模型处理后的图数据仍旧具有极高的可用性。因此，该图数据匿名保护方案适合于解决海量图数据的隐私保护问题。

面向大数据平台的程序分析优化方法研究

姓　　名：王　斐
研究方向：大数据平台优化
导　　师：吴　松
指导老师：石宣化
Email：275254920@qq.com
QQ：275254920
联系电话：13164624531
毕业去向：网易杭州研究院

在大数据处理系统中，通过大数据分析技术来得到有用的信息，需要同时保证可靠性与及时性。因此，数据并行应用的性能变的越来越重要，这些应用的性能和代码逻辑与在运行时的资源利用有关。程序分析技术是一种常用的优化应用的方法，而且这项技术已经被运用到数据并行应用中，但是目前还没有一个能够高效的分析整体数据并行应用的方法。由于在大数据编程模型中涉及到大量的复杂操作，比如数据分区、数据分布和并行执行等，因此每个大数据应用涉及到大量的复杂的框架代码，即使是一个非常简单的应用。如果直接分析整个数据并行应用，是十分低效的，会浪费大量的时间，甚至不能得到有用的分析结果。

通过对数据并行应用进行观察和分析，我们得到数据并行应用的特点，即数据并行应用的运行过程具有阶段性，每个数据并行应用的作业会划分为阶段。在每个阶段中会运行多个独立的任务，任务的数目和数据分区的数目相同，每个任务会计算一个分区的数据，且每个任务对分区中的数据进行同样的操作。根据这个特点，提出了基于阶段划分的程序分析(Stage-Divided Program Analysis)，称之为SDPA，来简化和加速数据并行应用的程序分析。对于每个阶段，对其进行重新划分为一个或多个运行周期。对于每个周期内的用户定义函数，将其抽取出来并融合为一个方便分析的函数。这样避免了过程间分析且规避开不必要分析的复杂数据处理系统框架代码。

我们将SDPA实现在Spark上面，并且做了大量的实验来评测此方法的性能。实验结果表明，基于阶段划分的程序分析方法相较于直接分析数据并行应用的程序分析方法可以减少96.4%到98.8%的预处理时间和99.8的分析时间；SDPA对不同敏感度的程序分析都有很好的性能；SDPA可以覆盖Spark机器学习库中大部分应用。

共享内存环境下图查询并行处理

姓　　名：王　磊
研究方向：图数据库
导　　师：谢　夏
指导老师：袁平鹏
Email：alphaengine@foxmail.com
QQ：2503612338
联系电话：1316336516
毕业去向：字节跳动

RDF作为语义网框架的一项核心概念，因其具有简洁灵活等优点，常被用于表示图数据。SPARQL是一种标准的RDF数据查询语言，是一种常用的图查询语言。随着RDF数据的爆炸式增长，现有的一些SPARQL查询处理系统已无法在合理时间内处理复杂查询。为此，SPARQL 查询引擎应当使用并行处理技术来实现对复杂查询的高效处理。然而，当前的一些SPARQL查询并行处理技术有着查询优化效率低下且生成的查询计划不利于并行，系统并行程度较低等不足。

ParTriple旨在为多核计算机环境下的SPARQL查询提供并行处理。首先，该系统提出了一种高效，易于并行的查询计划生成算法。这种算法生成的查询计划可以使用流水线处理技术，且不同的流水线之间不需要相互等待。进一步地，这种查询计划可以更有效地削减中间结果，提高查询处理性能。接着，该系统提出了一种数据块级与模式级的两级并行执行框架，将查询执行分解成基于数据存储分块的子任务与基于中间结果分块的子任务，从操作符间并行与操作符内并行两个层面开发了系统的并行性。最后，该系统提出了三种不同的并行操作符，分别针对数据扫描，连接处理，数据洗牌等操作，进一步提升了系统查询处理的性能。

实验表明，在LUBM，WATDIV，BTC三个数据集上，对于复杂的SPARQL连接查询，ParTriple与目前流行的RDF数据管理系统 RDF-3X,TripleBit,Virtuoso相比，均有显著的性能提升。与此同时，ParTriple也展现了良好的关于线程数目的可扩展性。

动态图中核值维护算法研究

姓　　名：王　娜
研究方向：大数据
导　　师：金　海
指导老师：于东晓
Email：467502070@qq.com
QQ：467502070
联系电话：13125091230
毕业去向：华　为

紧密子图挖掘是图数据分析领域的研究热点，可用于社区发现、分析网络拓扑结构以及网络行为和功能预测等。图中顶点核值是反映顶点所在子图紧密性的重要指标，被广泛应用于紧密子图挖掘。以往的研究多集中于静态图中核值计算，在更为常见的动态数据图中如何避免重复计算，实现动态核值维护，因多边插入/删除时顶点核值变化量难以确定等难题，还未有高效算法提出。

为解决上述难题，基于边分组思想的核值维护算法得以提出，根据一定约束条件将插入/删除的边分为多个组，使得同一个组内的边插入/删除时造成的顶点核值变化量可确定，将多边更新时的核值维护问题分解为边分组和寻找核值变化的顶点两个子问题，从而高效解决动态核值更新问题。更具体地，通过证明满足“匹配”和“优边集”性质的边集在被同时处理时，能够保证顶点的核值变化量为确定值，给出基于“匹配”和“优边集”两种不同的边分组策略，得到高效核值维护算法。其中，基于匹配的算法具有更少的预处理时间，而基于优边集的算法可同时处理更多的边，减少处理轮数。

相比于传统基于单边处理算法的多边顺序

处理方式，基于分组策略的处理算法不仅极大降低核值更新的时间，提高核值更新效率，而且可以有效减少单边处理算法顺序执行过程中产生的冗余计算，降低计算成本和存储空间。此外，基于分组策略的多边处理算法允许并行执行，可通过在并行系统运行进一步提高核值更新的效率。

在真实图数据、时序图数据和生成图上的大量实验表明，基于分组策略的多边处理算法在实际环境中可高效更新核值，极大降低核值维护所需的时间，并在并行系统运行环境下具有良好的可扩展性。

实体搜索的索引及查询扩展方法的研究

姓　　名：王　沛
研究方向：信息检索
导　　师：赵　峰
指导老师：赵　峰
Email：wangpei037@163.com
QQ：2359137022
联系电话：18627270825
毕业去向：阿里巴巴

论文提出一种面向实体搜索基于语义的查询扩展方法，该方法在开放性知识库（WordNet、维基百科和YAGO3）的指导下，通过以词条为中心和以实体为中心的查询扩展过程，并分别利用语言模型和向量模型对候选项进行评估，最后采用一种线性加权的混合模型来得到最终搜索结果。文中基于语义的查询扩展方法主要包含构建语义索引和两阶段查询扩展，构建语义索引通过抽取结构化和非结构化文档中实体关联信息和词条语义信息并通过倒排表和关系链表的形式进行存储，在两阶段的查询扩展过程中，首先通过局部分析和语义映射获取词条候选集并基于支持文档集生成模型来计算扩展关系模型得到词条扩展集，然后利用语义映射生成实体候选集并使用向量模型进行评估得到实体扩展集，最后对于扩展查询生成的结果采用线性加权的方式来生成最终结果。为了验证基于语义的查询扩展方法的有效性，论文在ClueWeb09数据集上进行了测试，主要对比的是基于词条、基于实体和基于语义的查询扩展方法在该数据集上的精确度、召回率和F1-Measure值。最终，实验结果表明基于语义的查询扩展方法较另外两种方法在精确度和F1-Measure值上有明显提升，而基于词条的查询扩展方法在召回率上优于其余两种方法。

云环境下的大规模图处理资源性价比评估

姓　　名：赵　鹏
研究方向：GPU图计算
导　　师：石宣化
指导老师：石宣化
Email：zhaopeng@hust.edu.cn
QQ：364002787
联系电话：18571717966
毕业去向：北京世纪好未来教育科技有限公司

大规模图处理是大数据技术中的重要组成部分，有着广泛的应用领域。当前关于大规模图处理的研究工作集中在图处理性能的提升上，其目标是更快的图处理速度和更大的图处理规模，用户无法得知哪类图处理方案符合其利益诉求以及如何选择图处理的系统方案。云环境下的大规模图处理资源性价比评估以GPU和CPU图处理方案为研究对象，以资源性价比作为图处理方案的选择依据，为用户提供了在云环境下大规模图处理方案的性价比指导意见，帮助用户以更少的资源实现图处理的业务需求。资源性价比高的系统方案，可以消耗较少的资源来得到所需求的图处理性能，或者在相同的资源消耗下得到更高的性能，使得用户的图处理业务得到更高的收益。首先，提出

了资源性价比的正式定义和评估模型，然后结合云环境下的大规模图处理的使用场景，给出了云环境中的资源代价模型和大规模图处理的评价模型。最后，根据图处理系统的体系结构、图算法和图数据的特性和分析，提出了进行资源性价比评估的方案，并对GPU与CPU的图处理资源性价比进行了实验测量和数据分析，评估和论证了GPU的图处理资源性价比。评估实验的结果表明，在云环境下适合于使用GPU进行大规模图处理，对于性能优先和代价优先的用户，GPU的图处理资源性价比更高，用户应该选择使用GPU图处理系统；对于不关注图处理速度的用户，CPU图处理的资源性价比更高，用户应该选择部署CPU图处理系统。

电子政务流式大数据实时热点识别研究

姓　　名：李文垚
研究方向：大数据
导　　师：王多强
指导老师：陈汉华
Email：lihancock7@icloud.com
QQ：417724084
联系电话：15827003717
毕业去向：北京趣拿软件科技有限公司

随着信息化建设的不断发展、电子政务的日趋成熟，政府部门掌握着海量的数据。在对电子政务大数据的处理与挖掘中，热门数据往往是很重要的，比如，在网络舆情监测应用中，政府对网络言论中的关键词进行收集并统计，识别出实时的热门关键词。而随着计算机存储和处理能力的逐步提高，电子政务应用对大数据的处理的要求也越来越高。传统的数据热点识别方法，如Lossy Counting、Space Saving、Count-Min Sketch等，主要对数据的历史累积量进行统计，得到统计量最高的数据项，并不能达到对当前实时热点进行识别的目的，无法满足处理实时数据流时的时效性要求。

为了解决以上问题，利用少量计数器作为过滤器将数据根据频率进行区分，筛选出高频的热点数据项，单独进行记录，从而更精准地对热门数据进行统计，并支持对热门数据的输出，而中低频数据在Count-Min Sketch结构中记录。分开统计保证高频数据项与低频数据项基本不发生碰撞，同时当低频数据项变为高频数据项依然可以被识别出来。为了保证时效性，所有统计结构中均加入时间戳信息，可以获取到任意数据项在任意时间段内的计数信息，为了保证较高的空间利用率，采用线性拟合技术对带有时间戳的计数信息进行压缩。

实验结果表明，可以在处理大规模流数据时识别出实时热点，吞吐量达到18万条数据每秒，查准率、查全率、超过99%。与16年提出的Window Compact Space-Saving算法相比，吞吐量与查全率结果相似，查准率结果明显更优，热点识别的延迟更小。

基于WebGL的网络数据可视化系统

姓　　名：汪　悦
研究方向：信息可视化
导　　师：吕新桥
指导老师：袁平鹏
Email：yuewang@hust.edu.cn
QQ：844215625
联系电话：18062554115
毕业去向：京　东

近年来，随着移动互联网生态的发展，手机等移动终端硬件和网络发展提速，在这过程中积累了海量的数据。利用数据可视化技术从这些数据中提取到有价值的信息成为发展趋势。然而现有可视化工具基本处于“胖客户端”状态，可视化的规模受客户机硬件影响，并且没有对移动设备进行很好的适配等问题，严重影响了当大数

据环境下数据可视化的应用普及。研究能够针对大规模数据进行可视化展现的并且对移动端适配的工具，具有重要的理论和实际意义。

基于WebGL的网络数据可视化系统MobVis有效解决了上述的问题。 MobVis 采用节点布局计算与图形渲染分离的机制，在服务端进行网络数据的计算处理， 设计实现了基于社团检测的层次布局策略， 减少了网络数据布局计算的时间；服务端在网络数据存储方面设计实现了基于二进制的存储方式，减少了磁盘占用空间和后续网络传输时间。客户端采用 WebGL基于硬件加速渲染技术对可视化图形进行绘制，使用粒子系统和BufferGeometry技术，对大规模数据进行渲染处理，并且支持丰富的人机交互操作。最后针对移动端设备通过响应式设计技术对移动设备进行可视化的适配， 针对移动硬件的性能做出了优化，提高可视化响应的速度。

MobVis 现已在 Linux上进行了实现，支持移动端和 PC端的可视化展现，并对系统进行了实验测试。实验表明，系统在功能上对 RDF数据的查询和数据集可视化，以及可视化交互方面都有很好的体现；在性能方面，系统支持百万级别数据的可视化展现，并且渲染帧率在40fps以上，数据磁盘占用空间节省了30%以上，数据加载时间节省了 40%左右。

航空大数据的高效更新及特征分析系统

姓　　名：胥清清
研究方向：图数据
导　　师：袁平鹏
指导老师：袁平鹏
Email：xuqingqing@hust.edu.cn
QQ：906692466
联系电话：15827192864
毕业去向：成都华为

随着城市化进程的不断加快，旅客出行记录、航班计划等航空大数据与日俱增。利用航空数据进行客流量分析、售票分析、旅客分类等对航班计划调整、提供个性化服务等方面有着重要指导作用。目前大多数据分析系统利用分布式存储与计算的方法直接处理大规模数据，数据存储空间占用较大且对数据批量更新未能较好地支持，在海量复杂关联的数据处理上具有较高的时间与空间复杂度。

航空大数据的高效更新及特征抽取系统提供了高效的批量更新与特征抽取方法。由于旅客出行记录包含大量重复信息，并且实体之间具有复杂的关联信息，为减少数据的存储空间和加快数据检索的效率，系统设计以图结构方式分块存储旅客出行记录。为加快数据处理，系统在数据块层次建立并行处理模型，不同数据块上的数据操作可并行处理。此外，系统采用缓存策略减少对数据文件频繁读写，提高系统性能。系统利用图结点与边的基本特征以及图计算等方法提供高效的特征抽取方法，实现城市客流量统计、热点城市分析、航线规划等。

与当前主流的图数据存储系统进行对比，系统在插入数据量高达500万时性能是Virtuoso、PostgreSQL，x-RDF-3X的数十倍，并且任务处理的并行度可根据线程数进行扩展。在特征抽取方面，系统利用弹性分布式框架Spark对数据进行分析，提供了基于内存的海量数据处理能力，保障了航空数据高效的更新及特征抽取。

面向共享内存池的异构内存网络

姓　　名：段卓辉
研究方向：异构内存
导　　师：蒋文斌
指导老师：刘海坤
Email：zhduan@hust.edu.cn
QQ：122316931
联系电话：13627101736

毕业去向：华中科技大学

非易失内存是最近研究的热点领域，其高存储密度、非易失、低功耗等特点为打破内存墙提供了机遇。但非易失内存也存在低寿命、较高写延迟和较低带宽等一些不可避免的缺陷。将非易失内存和易失内存组合成异构内存架构成为解决上述问题的一个重要途径。随着异构内存研究的发展，可扩展性的问题逐渐变得突出。单节点异构内存结构并不具有较好的扩展性，也不是当前数据中心异构内存结构的可行解决方案。数据中心的急需一种高性能，高可扩展性的异构内存共享方案。

提出并开发了一个混合内存共享池系统：ShareHM。ShareHM通过分布式内存文件系统、Remote Direct Memory Access（RDMA）通信技术、混合索引、单节点异构内存分配和迁移策略构建混合内存池系统。在顶层，使用分布式内存文件系统来管理混合内存资源，并且通过一个混合索引来提高分布式文件系统中Key value（KV）操作的性能；在中间层，使用优化RDMA技术来构建内存数据共享机制；在底层，使用拥有内存分配策略和迁移策略的定制内核来实现更高的异构内存利用率。ShareHM紧密集成了RDMA技术和分布式内存文件系统，以减少传统分布式内存文件系统中的数据访问延迟，改进了单向RDMA技术以获得更快的网络通信，并设计更好的Remote Procedure Call访问以提高异构内存共享池吞吐量。ShareHM通过混合索引加速异构内存共享池的KV操作，使用单节点带宽感知数据放置和迁移策略来提高异构内存数据节点的带宽利用率。

对ShareHM进行了测试并与现有的分布式内存文件系统memGluster和Crail进行了对比。ShareHM的软件开销相比于memGluster大幅降低。ShareHM能较好利用网络硬件所提供的最大读写带宽。ShareHM的读写吞吐量也明显高于其他文件系统。

基于CUDA加速的医学图像可视化系统

姓　　名：樊　荣
研究方向：系　统
导　　师：郑　然
指导老师：郑　然
Email：1823325376@qq.com
QQ：1823325376
联系电话：15827380204
毕业去向：深圳华为

作为科学计算可视化的种，医学图像可视化是对医学图像序列三维展示的一种技术，能医生提供更加真实的显示效果与定量分析方法。通常可以将它归为面绘制与体绘制两类方法，其中体绘制更能表现体数据内部细节，更符合医学重建需求。而医学重建中光线投射算法效果最好，但计算量也最大，基于CPU的串行的医学光线投射算法，很难满足交互的实时性，常常需要通过GPU技术对程序进行加速。而传统基于GPU的光线投射算法采用图形绘制管线编程，同时采用一维数组进行纹理绑定，仍然存在纹理绑定效率低和采样效率低的问题。

通过设计基于CUDA并行框架的光线投射算法来缓解上述问题。采用CUDA三维纹理绑定医学体数据，并采用纹理的硬件插值功能进行数据采样，直接提高了数据绑定和访问效率。通过采用Blinn-Phong光线模型进行物体表面绘制，同时，在光线投射过程中，设计基于包围盒和阈值的光线提前终止算法，避免了无效的采集步骤，提高采样速度。最后，通过设计并实现基于Sobel算子的二维传递函数，同时引入边界增强因子，增强对绘制对象边界部位的绘制效果，突出边界显示效果，从而达到更好的组织区分度。

实验证明，基于CUDA框架的光线投射算法，在绘制图像效果不差于传统GPU算法的同时，整体绘制速度能够达到30%的性能提升，

加入Sobel算子与光线提前终止算法的基于CUDA光线投射算法较传统的GPU算法，不仅在绘制速度上有至少20%提升，在图像边界处有更好的绘制效果，同时，算法绘制的FPS能达到60，能够较好地支持展示和交互的实时性。

面向并行图处理的异构内存访存优化机制

姓　　名：吕青侠
研究方向：异构内存
导　　师：胡　侃
指导老师：吕新桥
Email：1654538659@qq.com
Qq：1654538659
联系电话：15827211730
毕业去向：上海携程

大数据应用对内存容量和性能提出了很高的要求。传统的DRAM（Dynamic Random Access Memory）由于功耗和存储密度限制，在内存扩展上面临一些瓶颈。由DRAM和非易失性内存（None-Volatile Memory, NVM）组成的异构内存系统可以提供更大的存储空间来满足对内存的需求。然而与DRAM相比，NVM具有更大的读写延迟和更高的写能耗。对于并行图处理来说，总的执行时间取决于所有处理器中最长的执行时间。当在异构内存中处理相同大小的数据时，NVM上的数据处理速度比DRAM 上的数据处理速度慢得多。传统的针对同构内存系统中的数据划分方案通过考虑每个处理器的计算负载来平衡每个处理器的执行时间，这些方案并不适用于异构内存系统。

面向并行图处理的异构内访存优化系统X-ligra旨在提供合理的数据放置策略来提高系统性能，X-ligra使用了离线放置和在线调整相结合的数据放置机制。X-ligra设计了性能开销模型用来计算访问内存的开销和计算负载开销。基于此模型，数据被划分成不同比例的数据块，然后根据放置策略将数据块放置到DRAM或NVM中，以降低异构内存环境对系统性能的影响。X-ligra还引入了轻量级的动态调整机制，它可以进行在线监测和数据迁移。根据监测信息计算运行时的负载均衡程度并动态调整应用程序的数据划分与放置。

通过使用模拟器来模拟异构内存环境，对系统进行了测试。实验结果表明，与现有系统Polymer相比，在异构内存环境中使用静态数据放置策略的X-ligra可以在几种典型的图算法实现15.3%到30.1%的性能提升。与静态数据放置方案相比，使用在线调整方案的X-ligra可以进一步实现6%的性能提升。

基于容器的电力计算云平台

姓　　名：何涉平
研究方向：Docker容器
导　　师：王多强
指导老师：王多强
Email：351211842@qq.com
QQ：351211842
联系电话：13297036520
毕业去向：华为技术有限公司

近年来，随着电力工业的快速发展，越来越多的供电企业已经将电力计算应用迁移至云平台，利用云平台的强大计算能力处理大规模电力数据。电力计算云平台大多采用虚拟机技术实现，减少了供电企业的维护成本。然而虚拟机也带来了启动慢、应用部署麻烦等问题。Docker是一种轻量级虚拟化技术，具有启动快、资源利用率高等优点，因此研究基于容器的电力计算云平台，对于提高电力计算的效率，具有重要的现实意义。

基于容器的电力计算云平台将电力计算应用封装成镜像并可以在容器中快速启动，大大简化了应用的部署和维护。首先为了满足云平台的需要，使用Docker容器技术将各类电力计

算应用及其依赖环境封装成Docker镜像，上传至Docker镜像仓库，在需要时即可方便地从镜像仓库拉取对应的镜像。其次使用Kubernetes技术搭建电力计算云平台，实现对容器集群的统一管理，并设计了一种通用的Service模型，为用户访问提供了统一的入口，可以自动地将用户的请求转发到后端的容器集群中，电力计算容器完成计算之后，云平台将计算结果数据返回给客户端。同时各类电力计算应用可以按需地以Service的形式提供给用户访问。然后，设计了合理的数据管理策略，用户上传所有的电力数据到云平台，实现电力数据的集中管理和维护。最后设计了用户管理策略，实现对各级用户的权限管理和用户数据管理。

基于OpenStack和容器的多云资源调度系统

姓　　名：黄东锋
研究方向：多云资源调度
导　　师：邵志远
指导老师：刘方明
Email：dfhuangg@gmail.com
QQ：1269566421
联系电话：13697328817
毕业去向：华为技术有限公司

随着云计算业务的发展，运营商需要在不同的地理位置部署OpenStack云为用户提供服务；当单个OpenStack云的计算资源耗尽时，需要部署多个模块化的OpenStack云来解决资源扩展问题；企业为了避免运营商锁定问题，会考虑使用多个运营商的云服务；企业发布应用时，会基于OpenStack构建容器平台，将应用打包到容器中部署。面对地理分布式的多个OpenStack云和容器云，运营商难以管理，用户对各个云中的资源也缺乏统一的资源视图。

多云资源调度系统利用API gateway技术，将地理分布式的多个OpenStack云逻辑级联成一个云，为云运营商提供统一的资源管理入口来管理多个OpenStack云，同时为用户提供资源的全局视图。由系统管理的多个OpenStack云对租户不可见，租户像使用一个OpenStack云一样使用多个OpenStack云中的资源。系统分为API gateway 模块和资源调度模块。API gateway包括Nova API gateway、Cinder API gateway和Zun API gateway，分别负责管理多个云中的虚拟机资源、磁盘卷资源和容器资源。API gateway 作为单独的Web服务，接收用户请求，然后通过资源调度模块，基于一定的调度策略，发送到合适的OpenStack云中处理。系统实现了Filter-Weighter资源调度算法和随机资源调度算法。Filter-Weighter算法根据不同的配置参数，可以达到负载均衡或最小化电能消耗等目标。

多云资源调度系统利用Python实现，最终可以在Linux平台部署和使用。实验测试表明，通过为系统配置不同的资源调度算法，可以达到不同程度的资源利用率。与单个OpenStack系统相比，多云资源调度系统在操作磁盘卷和容器时有更短的响应时间，操作虚拟机时有更高的吞吐率和更好的鲁棒性。

基于图数据处理的移动通信行业关键用户推送系统

姓　　名：阮　臻
研究方向：云计算
导　　师：陆　枫
指导老师：陆　枫
Email：henurz@163.com
QQ：543151684
联系电话：18937309977
毕业去向：百度在线网络技术有限公司

随着移动互联网的不断发展，对移动通信行业的传统业务带来了巨大的冲击，如何利用收集到的数据提升服务质量，扩大营收是运营商急需解决的问题。由于移动终端的大量普

及，市场仍保留大量移动终端的需求。换机用户为移动终端市场提供了大量的终端需求。如何发现潜在的换机用户群体对于移动运营商终端营销是至关重要的。

针对发现用户群体中潜在换机用户的难题，通过移动运营商提供的用户信息、通话行为等数据进行分析，使用特征工程等方法发现影响用户换机的因素，使用分类算法建立潜在换机用户模型。通过运营商提供的基站数据，使用社区发现算法发现筛选到的关键用户所在的社区提高运营商营销范围。之后通过对关键用户所在的社区进行用户影响力分析，分析结果可以为运营商扩大营销范围提供数据支持。

基于图数据处理的移动通信行业关键用户推送系统现已进行部署，实现了对用户群体中潜在换机用户的发现以及用户社区分析等功能。潜在换机用户模型属于分类模型，选择了平均准确率、精准度、召回率、F1-score、ROC 曲线等评估指标评估模型效果。选取Logistic Regression、SVM、C5.0、GBDT等分类算法建立的模型与本文建立模型进行对比。本系统创建模型的准确率为 0.92， F1-score为 0.61629， AUC的 值 为 0.8044，该模型的准确率高于其他算法创建模型，模型的整体效果相对其他模型较优。

面向三维重建的胶囊内镜图像拼接

姓　　名：王祖剑
研究方向：云计算
导　　师：陆　枫
指导老师：陆　枫
Email：263393242@qq.com
QQ：263393242
联系电话：18202775140
毕业去向：行吟信息科技有限公司

无线胶囊内镜在检查过程中会产生的大量图像，这些图像包含有限的信息并且存在着大量冗余，给诊断造成了困扰。需要对图像去除冗余的同时，还能够保留冗余图像的信息，使用图像拼接是一条必选之路。而图像拼接对配准有着精准的要求，现阶段的方法并不能在无线胶囊内镜图像中提供令人满意的结果。因此，找到合适的配准算法对于拼接胶囊内镜图像以方便诊断有着重要的意义。

针对胶囊内镜图像中普遍存在的弱纹理、存在杂质和多角度拍摄的问题，提出Affinedeep算法来完成图像配准。胶囊内镜图像因为弱纹理的问题而不便于提取特征点，Affinedeep算法利用卷积的方法计算图像之间的关联图，用来描述图像之间的相似度，规避了对特征点的提取，并根据结果建立图像金字塔，在塔顶确定最为相似的区域后向下溯源得到图像之间的匹配。为了解决多角度拍摄的问题，Affinedeep算法利用仿射矩阵对源图像进行变换，模拟从不同的角度位置的相机拍摄得到的图像，试图在模拟过程中包含真实的情况。使用模拟图像计算匹配，并将变换后的图像中的匹配点投影到源图像中以增加匹配数量，匹配结果经过筛选后完成配准。配准的结果用来计算图像之间的单应性矩阵，根据单应性矩阵完成对图像的透视变换以对齐图像间的重叠部分并完成拼接。

基于Docker的云仿真平台

姓　　名：周　阳
研究方向：Docker容器
导　　师：谢　夏
指导老师：王多强
Email：zhouyangyoung@163.com
QQ：329088816
联系电话：18202711703
毕业去向：杭州网易

Docker是一种面向云平台的轻量级虚拟化技术，拥有容器启动快、资源浪费低等优势。

自从2013年被公开发布以来，已经广泛应用于运维、大数据、高性能计算等诸多领域。传统的分布式仿真平台使用虚拟机来部署和运行仿真联邦。虚拟机在提供了隔离的运行环境的同时，也带来了启动慢、环境配置复杂、资源浪费大等问题。因此，研究基于Docker技术的云仿真平台是一种提高仿真服务质量的方法，对于分布式仿真领域具有重要现实意义。

为了建立基于Docker技术的云仿真平台，提出了联邦在容器环境下的运行模型。该模型下，各联邦成员分别运行在独立隔离的容器中，选用Docker的Overlay网络模式来解决联邦成员间跨主机通信问题。使用Docker数据卷存储联邦执行所需数据，避免了联邦执行失败造成的数据丢失。为了满足部分联邦的可视化展示需要，提供了本地和远程两种图形化界面显示方案。提出了将仿真资源归类并基于Dockerfile的镜像构建方法，简化了仿真资源镜像构建步骤。为了方便联邦的集群化部署与执行，基于原生Docker下的联邦运行模型，提出了Kubernetes环境中的联邦运行模型，最终建立了基于容器技术的云仿真平台，该平台包括仿真开发模块、仿真资源管理模块、仿真联邦执行模块、系统管理模块等。它拥有联邦部署快速、执行效率高、资源利用率高的优势。

联邦成员启动耗时测试表明了原生Docker中联邦成员的启动耗时远小于虚拟机环境下，云平台中的联邦成员启动耗时是虚拟机环境中启动耗时的11%至16%之间。联邦通信耗时测试说明联邦成员在云仿真平台的通信速度快于虚拟机环境，平台中通信耗时约是虚拟机环境的63%。资源消耗测试说明在云仿真平台中执行联邦不会带来过多额外的资源消耗。

基于区块链的智能合约访问控制系统

姓　　名：王晨龙
研究方向：区块链
导　　师：吕新桥
指导老师：代炜琦
Email：451506637@qq.com
QQ：451506637
联系电话：13163330931
毕业去向：去哪儿网

随着区块链技术的不断发展，许多企业和组织机构已经计划将传统业务通过智能合约的方式实现，典型的应用场景包括供应链管理、物联网等。但是，目前智能合约中还存在很多安全问题，严重影响了智能合约的应用与普及。首先，智能合约没有访问控制编程接口，每编写一个新的智能合约都需要实现访问控制功能，导致智能合约开发的成本较高，甚至有些开发者根本没有考虑访问控制的问题。其次，目前智能合约间的调用没有一套访问控制机制，也会导致一系列的安全问题。因此为区块链设计一套智能合约访问控制系统尤为重要。

基于区块链的智能合约访问控制系统解决了智能合约间访问控制的安全问题和智能合约内的访问控制接口的问题。（1）针对智能合约间访问控制的问题，系统首先引入了平台合约管理员和合约所有者角色，将智能合约部署的权限严格控制在一定范围内；其次，系统所有的管理操作都通过密钥签名，保证了规则修改权限的合法性，同时系统通过设置命令有效期的方式防止了命令重放攻击；最后系统通过新增消息类型和合约状态机状态的方式保证了合约所属关系的正确添加，通过合约间调用流程的修改保证了智能合约访问控制的正确执行。（2）针对智能合约内访问控制接口的问题，系统采用智能合约桩代码的方式提供基于属性的访问控制接口，并通过交易ID将访问权限修改记录串连起来，从而使得访问控制权限修改记录易于追踪与查询。

一种支持高并发交易的区块链激励系统

姓　　名：肖德山
研究方向：区块链
导　　师：谢　夏
指导老师：代炜琦
Email：deshanxiao@qq.com
QQ：1254582845
联系电话：15072399494
毕业去向：北京小米科技有限责任公司

区块链技术具有去中心化、不可篡改等特性，能在不同节点间建立信任。区块链技术保证节点互信的关键技术是共识机制。然而目前主流的共识机制，存在许多问题，严重影响了区块链技术的实用性。首先，共识机制没有有效地激励价值创造者，这使得比特币为首基于PoW共识机制的区块链系统被矿工垄断，普通用户难以参与。其次，这类区块链系统并发交易性能很差，无法应对真实场景的交易频次。因此设计一种支持高并发交易的区块链激励系统尤为重要。

一种支持高并发交易的区块链激励系统，能有效的激励价值创造者并保证交易系统并发量满足真实场景的需要。（1）针对目前共识机制没有有效地激励价值创造者问题，提出了基于价值量证明的激励机制，使得价值创造者可以得到系统奖励。又提出了激励系数的调整算法，使得奖励系统保持稳定。（2）针对区块链系统交易性能很差的问题，提出了快速链下交易系统，使得交易过程只需通过链下中心化服务端来完成，而无需区块链同步，提高了交易系统的并发量。

基于软件定义的漏洞防护研究

姓　　名：赵宏熠
研究方向：网络空间安全
导　　师：顾　琳
指导老师：邹德清
Email：453248399@qq.com
QQ：453248399
联系电话：15007188344
毕业去向：校内读博

在实际的工作环境中，用户往往会因为各种原因没有及时的给漏洞打补丁从而产生安全隐患，如用户忽视了安全的重要性关闭了系统的自动更新，或是服务器需要24小时运行无法安装补丁并重新启动，或是考虑软件的兼容性问题没有第一时间打补丁，因为补丁安装以后往往不能轻易的卸载，未经过详细测试的补丁可能会带来意料之外的问题。在2017年5月大规模爆发的勒索病毒WannaCry能证明这一情况，虽然微软早在数月之前就已经发布了补丁，但实际上仍有大量的用户忽视了补丁受到攻击。因此，在发现漏洞到安装补丁的这一段时间，可以从网络的角度提供对已知漏洞的防护，SDN对全网络的管理、数据流的监控能够为这一问题提供新的思路。

传统的入侵防御系统研究的优化方向主要包括检测准确性、系统带来的延迟、和防御手段的灵活性。在SDN网络中实现基于漏洞签名的入侵检测系统能够从这三个方面相较于传统方法做出提升。漏洞签名能够对所有可能触发漏洞的输入变量进行描述。由于是直接分析漏洞利用的原理，从根本上防止恶意的输入触发漏洞，因此基于漏洞特征的防御方法能够更精确的识别攻击数据包，漏报误报也更低。基于漏洞签名的规则能够与漏洞一一对应，因此能够大量减少入侵检测系统中的规则数量，减少规则匹配时带来的时间延迟。OpenFlow交换机可以通过流表对不同特征的流实施不同的防御手段，与传统屏蔽子网或是端口的方式相比具有更高的灵活性。

在实验中模拟真实的网络环境，不断产生正常网络数据包的同时，选取不同类型的真实漏洞构造攻击数据包，证了使用基于漏洞签名的检测规则能够准确的发现触发漏洞的数据包。在同样的实验环境下，将Snort自带的规则作为对比，证明了基于漏洞签名的规则检测的准确性相比自带规则更高。

Memcached剖析

（李志威 https://blog.csdn.net/qq_15439445/article/details/80488076）

1. Memcached简介

Memcached是以LiveJournal旗下Danga Interactive公司的Brad Fitzpatric为首开发的一款软件。现在已成为mixi、hatena、Facebook、Vox、LiveJournal等众多服务中提高Web应用扩展性的重要因素。

许多Web应用都将数据保存到RDBMS中，应用服务器从中读取数据并在浏览器中显示。但随着数据量的增大、访问的集中，就会出现RDBMS的负担加重、数据库响应恶化、网站显示延迟等重大影响。

这时就该Memcached大显身手了。Memcached是高性能的分布式内存缓存服务器。一般的使用目的是，通过缓存数据库查询结果，减少数据库访问次数，以提高动态Web应用的速度、提高可扩展性，如图1所示。

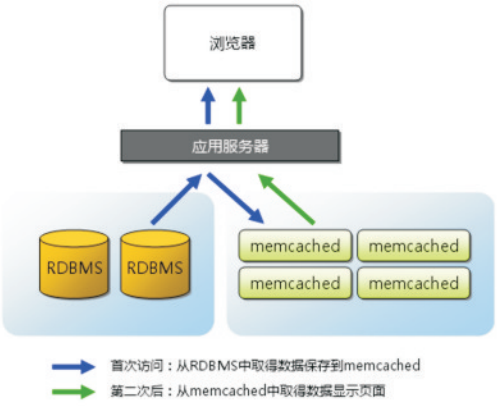


图1 一般情况下Memcached的用途

2. Memcached的特征

Memcached作为高速运行的分布式缓存服务器，具有以下特征：

（1）协议简单：Memcached的服务器客户端通信并不使用复杂的XML等格式，而使用简单的基于文本行的协议。因此，通过telnet也能在Memcached上保存数据、取得数据。

（2）基于libevent的事件处理：libevent是个程序库，它将Linux的epoll、BSD类操作系统的kqueue等事件处理功能封装成统一的接口。即使对服务器的连接数增加，也能发挥O(1)的性能。由于Memcached使用libevent库，因此能在Linux、BSD等操作系统上发挥其高性能。

（3）内置内存存储方式：为了提高性能，Memcached中保存的数据都存储在Memcached内置的内存存储空间中。另外，内容容量达到指定值之后，就基于LRU或3QLRU算法自动删除不使用的数据对象。Memcached本身是为缓存而设计的服务器，因此并没有过多考虑数据的永久性问题。

（4）Memcached不互相通信的分布式：Memcached尽管是分布式缓存服务器，但服务器端并没有提供分布式功能，各个Memcached服务器不会互相通信以共享信息，其分布式功能完全取决于客户端的实现。

3. Memcached的内存存储

Memcached采用了名为Slab Allocator的机制分配、管理内存。在该机制出现以前，内存的分配是通过对所有记录简单地进行malloc和free来进行的。但是，这种方式会导致内存碎片，加重操作系统内存管理器的负担，最坏的

情况下，会导致操作系统比Memcached进程本身还慢，Slab Allocator就是为解决该问题而诞生的。

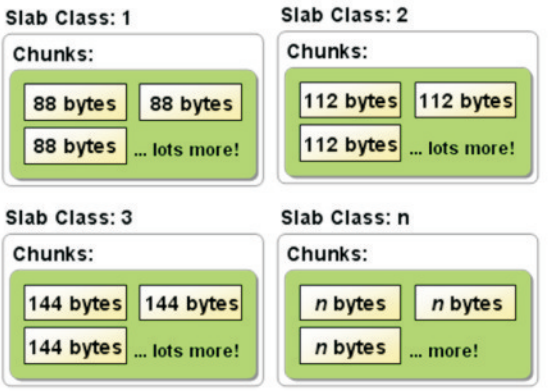


图2 Slab Alloator结构图

Slab Allocator提前将大块内存划分为1MB大小的若干个slab，然后针对每个slab再进行小对象填充，这个小对象称为chunk，从而避免大量重复的初始化和清理，减轻了内存管理器的负担。Slab Allocator内存分配的原理是按照预先规定的大小，将分配给Memcached服务器的内存吸纳分割成特定长度的内存块（chunk），再把尺寸相同的内存块（chunk）分成组，这些内存块不会释放，可以重复利用。

图3 选择存储记录的组的方法

新增数据对象存储时，因Memcached服务器中保存着slab内空闲chunk的列表，它会根据该列表选择chunk，然后将数据缓存与其中。当有数据存入时Memcached时，根据接收到的数据大小，选择最适合数据大小的slab分配一个能存下这个数据的最小内存块（chunk）。例如：在图3的分配中，100字节的一个数据就会被分配存入112字节的一个内存块中，这样会有12字节被浪费，这部分空间就不能被使用了，这也是Slab Allocator机制的一个缺点。

4. Memcached的数据驱逐机制

当新增数据对象存储时，若当前已无可用内存，Memcached需要一定机制驱逐旧数据以腾出内存空间容纳新对象，目前Memcached提供了两种数据对象驱逐策略：

（1）LRU

该策略为每个数据对象维护两个指针构建双向链表，当数据对象被访问后置于链表的MRU位置，内存不足时则从链表的LRU位置执行数据驱逐。

（2）3QLRU

该策略在（1）的基础上维护了三个LRU队列，即HOT_LRU、WARM_LRU、COLD_LRU。新分配的数据对象首先放置到HOT_LRU的MRU位置；当HOT_LRU或WARM_LRU队列中的数据对象超过预设的阈值后，将多余的数据对象转移到COLD_LRU中；如果某个位于COLD_LRU队列中的数据对象被访问，将其移动到WARM_LRU队列的MRU位置。

5. Memcached的分布式实现

Memcached虽然称为分布式缓存服务器，但服务器端并没有分布式功能，Memcached的分布式功能是完全有客户端程序库实现的，如图4所示。这种分布式是Memcached的一大特点。

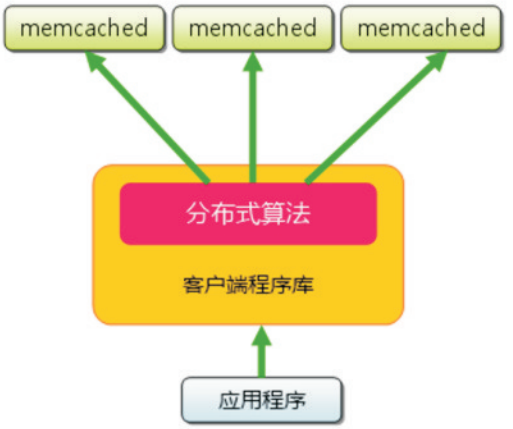


图4 Memcached的分布式

现在假设有4台Memcached服务器：node1～node4，要保存键为key1～key10的数据。首先往Memcached中添加key1，将key1传给客户端程序之后，客户端实现的分布式算法会根据这个键key1来决定保存数据的Memcached服务器。服务器选定之后，将会用选定的服务器来保存key1和对应的值。在获取数据的时候，先通过要获取的数据的key来根据客户端实现的相同的算法选择对应的数据保存的服务器，然后取出数据。这样就实现了Memcached的分布式功能。部署的Memcached服务器越多，键就会越分散。即使一台服务器挂掉，也不会影响其他服务器的运行。

目前常用的分布式算法主要有两种：余数计算分散法、一致性哈希算法。

（1）余数计算分散法

这种方法简单的说就是根据服务器的台数的余数来进行分散。首先求取键所对应的整数哈希值，然后根据哈希值与服务器台数取模后的余数来选择服务器。这种方法简单高效，而且数据的分散性也非常的好。但问题是当增加或者删除一台Memcached服务器的时候，余数就会发生巨大的变化。这样就没有办法获取和保存时相对应的服务器，从而极大地降低缓存的命中率。

（2）一致性哈希算法

这种方法首先求出Memcached服务器的哈希值，然后将它分配到0~2^32的圆上，然后使用同样的办法求出数据对象的键的哈希值，将其映射到圆上。然后从数据对象映射的点开始顺时针的查找，将数据对象保存到查找到的第一台服务器上面。如果超出了2^32仍然没有找到服务器，那么就将数据对象保存到第一台Memcached服务器上面。

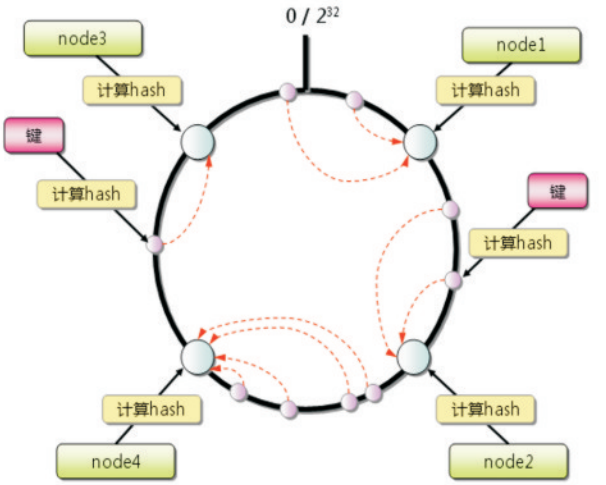


图5 一致性哈希原理图

这种方法在一定程度上在减少了修改Memcached服务器数量的时候对缓存命中率的影响。在一致性哈希算法中，只有在这个圆上，从增加服务器的那个点逆时针遇到的第一台服务器之间的键会受到影响。因此一致性哈希最大限度的抑制了键的重新分布。

另外一些一致性哈希算法也采用了虚拟节点的办法。因为使用一般的哈希函数的话，服务器的映射地点会分布的可能不太均匀，因此使用虚拟节点的思想，为每一台服务器在圆上分配100~300个点，这样就能够抑制分布不均匀，最大限度的减少服务器增加或者减少的时候数据对象的重新分布。

理解Paxos算法

（李少锋 https://www.cnblogs.com/leesf456/p/6001278.html）

一、前言

在分布式系统中，每一个机器节点虽然能够明确知道自己在进行事务操作过程中的结果是成功或是失败，但是却无法直接获取到其他分布式节点的操作结果，因此，当一个事务操作需要跨越多个分布式节点的时候，为了保持事务处理的ACID的特性，需要引入协调者的组件来统一调度所有分布式节点的执行逻辑，而被调度的节点则被称为参与者，协调者负责调度参与者的行为并最终决定这些参与者是否要把事务真正进行提交。

二、Paxos算法

Paxos算法是一种基于消息传递且具有高度容错特性的一致性算法，其需要解决的问题就是如何在一个可能发生异常的分布式系统中，快速且正确地在集群内部对某个数据的值达成一致，并且保证不论发生任何异常，都不会破坏整个系统的一致性。

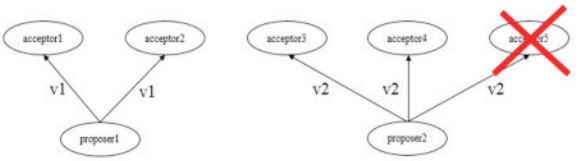
Paxos先把节点分成两类，发起提议(proposal)的一方为proposer，参与决议的一方为acceptor。

在没有失败和消息丢失的情况下，假如只有一个提议被提出的情况，如何确定一个提议，做到如下就可以保证

P1：一个acceptor必须接受它收到的第一个提议。

P1会引入一个问题，若 果多个提议被不同的proposer同时提出，这可能会导致虽然每个acceptor都批准了它收到的第一个提议，但是没有一个提议是由多数acceptor都接受的，因此无法确定一个提议。

即使只有两个提议被提出，如果每个提议都被差不多一半的acceptor批准了，此时也可能无法确定哪个提议，如下图所示



如上图所示，若acceptor5出现故障，则无法确定哪个提议。

在P1的基础上，增加如下条件：

a. proposer发起的每项提议分别用一个ID标识，提议的组成因此变为(ID, value)

b.若确定一个提议，需要由半数以上的acceptor接受，当某个提议被半数以上的acceptor接受后，我们就认为该提议就被确定了。

我们约定后面发起的提议的ID比前面提议的ID大，并假设可以有多项提议被确定，为做到确定并只确定一个值acceptor要做到以下这点：

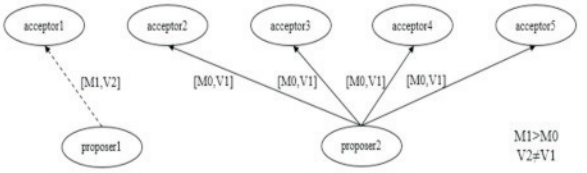
P2：如果一项值为v的提议被确定，那么后续只确定值为v的提议。

由于一项提议被确定(chosen)前必须先被多数派acceptor接受(accepted)，为实现P2，实质上acceptor需要做到：

P2a：如果一项值为v的提议被确定，那么acceptor后续只接受值为v的提议。

如下图所示，在acceptor1没有收到任何提议的情况下，其他4个acceptor已经批准了来自proposer2的提议[M0,V1]，而此时，proposer1产生了一个具有其他value值的，编号更高的提议[M1,V2]，并发送给了acceptor1，根据P1，就需

要接受该提议，但是这与P2a矛盾，因此如果要同时满足P1和P2a，需要进入如下强化



P2b：如果一项值为v的提议被确定，那么proposer后续只发起值为v的提议。

P2b约束的是提议被确定(chosen)后proposer的行为，我们更关心提议被确定前proposer应该怎么做。

P2c：对于提议(n,v)，acceptor的多数派S中，如果存在acceptor最近一次(即ID值最大)接受的提议的值为v'，那么要求v = v'；否则v可为任意值。

2.1 proposer生成提议

在proposer产生一个编号为Mn的提议时，必须要知道当前某一个将要或已经被半数以上acceptor接受的编号小于Mn但为最大编号的提议，并且，proposer会要求所有的acceptor都不要再接受任何编号小于Mn的提议，这也就是如下提议生成算法。

1. proposer选择一个新的提议编号为Mn，然后向某个acceptor集合的成员发送请求，要求该集合中的acceptor做出如下回应。

① 向proposer承诺，保证不再接受任何编号小于Mn的提议。

② 如果acceptor已经接受过任何提议，那么其就向proposer反馈当前该acceptor已经接受的编号小于Mn但为最大编号的那个提议的值。

我们将请求称为编号Mn的提议的Prepare请求。

2. 如果proposer收到了来自半数以上的acceptor的响应结果，那么它就可以产生编号为Mn、Value值为Vn的提议，这里的Vn是所有响应中编号最大的提议Value的值，当然，如果半数以上的acceptor都没有接受过任何提议，即响应中不包含任何提议，那么此时Vn值就可以由proposer任意选择。

在确定了proposer的提议后，proposer就会将该提议再次发送给某个acceptor集合，并期望获得它们的接受，此请求称为accept请求，此时接受accept请求的acceptor集合不一定是之前响应prepare请求的acceptor集合。

2.2 acceptor接受提议

一个acceptor可能会收到来自proposer的两种请求，分别是prepare请求和accept请求，对这两类请求作出响应的条件分别如下

prepare请求：acceptor可以在任何时候响应一个prepare请求。

accept请求：在不违背accept现有承诺的前提下，可以任意响应accept请求。

因此，对acceptor逻辑处理的约束条件，大体可以定义如下：

P1a：一个acceptor只要尚未响应过任何编号大于Mn的prepare请求，那么它就可以接受这个编号为Mn的提议。

2.3 提议的获取

使learner获取提议，有如下方案

① 一旦acceptor接受了一个提议，就将该提议发送给所有的learner，通信开销很大。

② 让所有的acceptor将它们对提议的接受情况，统一发送给一个特定的learner（主learner），当该learner被通知一个提议已经被确定时，它就负责通知其他的learner。主learner可能会出现单点故障。

③ 将主learner范围扩大至一个特定的learner集合，该集合中的每个learner都可以在一个提议被选定后通知所有其他的learner，集合learner越多，越可靠，但是通信开销越大。

实验室四篇论文被IEEE ICDCS 2018录用

姜炜祥、戴小海、李肖遥、柳　密

近日，国际学术会议The 38th IEEE International Conference on Distributed Computing Systems (ICDCS 2018) 录用结果揭晓，实验室有4篇论文被录用，分别是：博士生姜炜祥的论文“Non-IT Energy Accounting in Virtualized Datacenter”、戴小海的论文“Towards A Novel Architecture for Enabling Interoperability Amongst Multiple Blockchains”和硕士生李肖瑶的论文“DHL: Enabling Flexible Software Network Functions with FPGA Acceleration”、柳密的论文“TurboStream: Towards Low-Latency Data Stream Processing”。

姜炜祥的论文“Non-IT Energy Accounting in Virtualized Datacenter”主要研究了数据中心中非IT设施能耗在虚拟机之间的分摊与计量问题。能耗管理是数据中心实现节能减排和降低成本的重要技术基石。在整个数据中心中，制冷、电力系统等非IT设施能耗占比高达30~50%。然而，由于这些非IT设施由大量承载各种应用负载的IT设备共享和使用，并且IT能耗和非IT能耗之间存在非线性增长的关系，因此难以进行细粒度的非IT能耗测量及有效的能耗管理。论文巧妙地运用博弈论，将数据中心中非IT设施能耗在虚拟机层面进行细粒度划分的复杂工程问题提炼转化为简洁的成本分配问题，从而运用经济学中著名的夏普利值（Shapley Value）方法为数据中心中大量的动态虚拟机的非IT能耗实现公平高效的计量。针对夏普利值计算在大规模数据中心中复杂度极高O（2^N）的瓶颈，论文通过对IT能耗和非IT能耗之间的行为模式进行实测与分析，设计了高效的降维方法，将计算复杂度大幅降低至O（N）。基于真实数据中心能耗数据驱动的实验验证，论文所提出的降维方法与理论最优的夏普利值相比，最大误差只有6.97%，能够有效应用于实际数据中心系统。

戴小海的论文“Towards A Novel Architecture for Enabling Interoperability Amongst Multiple Blockchains”主要研究区块链的跨链交互问题。随着区块链技术的飞速发展，区块链系统的应用场景和体系结构趋于多样化。为促进不同区块链之间的信息流通、打破链间通信壁垒，如何设计区别于传统区块链结构，具备灵活高效且安全可扩展特征的跨链交互架构成为亟需解决的关键问题。鉴于“跨链交互”目前在学术界和工业界尚缺乏正式定义，论文首先对该问题进行了严格定义，并对现有跨链交互方案进行了分析。针对现有方案中，被动模式易于实现、但轮询开销较大，而主动模式的链结构为满足特定功能要求而被限制其适用范围等问题，提出了一种新型分层区块链架构，设计了一种MMR数据传输方案，用以降低被动式跨链交互中原有PBR方案的轮询开销。模拟实验结果表明在8条目标链读取同一条源链数据时，MMR方案可同时将源链的数据传输大小和CPU使用率降低为PBR方案的1/8。

李肖瑶的论文“DHL: Enabling Flexible Software Network Functions with FPGA Acceleration”主要设计和实现了面向软件网络功能（NFV）加速的FPGA-CPU协同设计框架

（包括一体化的硬件平台和软件库），为广大研发人员提供了通用且易编程的最新实用工具。网络功能虚拟化技术旨在将多种多样的网络功能从昂贵固化的专用网元设备解耦到通用服务器上，以软件方式灵活部署与运行。然而，当前软件网络功能在进行深度包处理时，需要消耗大量的CPU内核资源才能达到线速度，而若将整个网络功能部署到FPGA上则会造成不切实际的资源浪费。此外，网络功能的更改需要耗费数小时生成新的FPGA加速程序，阻碍网络功能的快速部署。针对上述问题，论文首次提出和实现了基于动态硬件库（DHL）的FPGA-CPU协同设计框架：（1）将FPGA中的深度包处理加速模块抽象成硬件函数并提供DHL编程API，使同类的网络功能复杂逻辑统一部署在FPGA中加速，而简单逻辑仍协同运行在CPU中，实现灵活、快速、高性价比的部署，具有通用性和易编程的优势；（2）通过有机结合无锁通信队列、用户态I/O、NUMA感知的内存分配、批处理和轮询等实用优化技术，最大化DHL整体框架的网络性能。实验验证了DHL 框架相比领域内基于纯CPU和纯FPGA加速系统的优势。

柳密的论文“TurboStream: Towards Low-Latency Data Stream Processing”主要设计和实现了面向低延时的流数据处理（DSP）系统TurboStream。在实践中发现，操作符间的延时占到DSP应用总处理延时的86%以上，TurboStream专为解决操作符间通信所带来的高延时的问题而设计，引入了两个功能互补的组件：（1）改进的IPC框架。DSP系统广泛采用的IPC框架Netty因过多的内存拷贝操作而不适用于本地进程间的数据传输，尤其是在低延时的DSP系统中。改进后的IPC框架在内部集成了一个面向DSP的堆外流式环形字节码缓冲区OSRBuffer。当在本地进程的操作符之间传输消息时，它可以将内存拷贝的次数降到最低，同时减少单个消息在缓冲区中的等待时间。（2）粗粒度调度器。OSRBuffer的效果受限于本地进程间的通信量占总通信量的比例。为了突破这种局限，进一步提出了粗粒度调度器。它在调度之前会根据操作符实例间的数据依赖关系和运行时的通信量信息合并操作符实例，再将合并后的操作符实例分配到节点，以减少节点间的IPC通信量。鉴于阿里巴巴JStorm在工业界的广泛应用以及其在低延时方面的优异表现，TurboStream的原型基于JStorm实现。实验证明，与JStorm相比，TurboStream将DSP的平均处理延时降低了83.23%。

ICDCS是分布式计算与系统领域享有盛誉和重要影响力的顶级国际学术会议，本届ICDCS在全球378篇投稿中录用78篇论文，录用率仅约20%。

姜炜祥

2014级博士研究生

研究方向：数据中心能耗管理

E-mail：wxjiang@hust.edu.cn

戴小海

2017级博士研究生

研究方向：云计算和分布式计算

E-mail：seafooler@hust.edu.cn

李肖瑶

2015级硕士研究生

研究方向：网络功能加速

E-mail：calmisi975@gmail.com

柳　密

2015级硕士研究生

研究方向：云计算与分布式系统

E-mail：2270566005@qq.com

Tigr: Transforming Irregular Graphs for GPU-Friendly Graph Processing

桂创意　推荐

“Tigr: Transforming Irregular Graphs for GPU-Friendly Graph Processing”是体系结构领域的国际顶级会议ASPLOS 2018录用的一篇文章，该会议于2018年3月在美国弗吉尼亚州东部的威廉斯堡举行。本文提出了一种用规则图替代不规则图进行图计算处理的轻量级虚拟转换方法，有效缓解了GPU在处理图任务时的负载不均衡问题，极大的提升了处理效率。

真实世界中的图结构呈现幂律分布的特征，小部分的节点通常连接了相当大数量的边，度数分布极其不规则，研究表明，真实图数据如LiveJournal有超过90%的点度数低于20，但是低于2%的点度数在1000到14000之间。GPU架构采用SIMD并行执行模式，适合于规则的数据处理任务，在处理真实图的时候存在严重的负载不均衡问题，从而无法充分发挥GPU的处理能力，导致性能减损。现有的解决方案集中于对图计算编程模型进行优化或者对GPU 底层处理单元进行复杂的调度来提升GPU 的性能，前者在设计上往往存在很大的编程难度，后者需要结合底层硬件架构来设计，难以适应处理器架构迅速的变化。

不同于以上两个角度，这篇工作从图数据结构本身出发，提出将不规则图转换成规则图来处理的思想，在提升GPU处理效率的同时保证结果的正确性。为了减少图结构的不规则性，Tigr提出了一种基于节点切割的转换方法（split transformations），迭代的将高度数节点进行切割直到度数达到预设值，并提出了一种能够同时兼顾规则度和收敛速度且保证正确性的拓扑连接方式——度数一致性树结构转换（uniform- degree tree transformation, UDT）。对真实物理拓扑进行转换往往会产生较多的预处理时间以及存储空间开销，因此作者还提出了一种虚拟拓扑转换技术（virtual split transformation），在不改变原来不规则图数据物理结构存储的情况下提供一个虚拟的规则图拓扑结构给GPU进行处理。

节点切割转换（split transformations）将高度数的节点切割成低度数节点的组合，高度数节点的出边按照预设的度数上限K被均匀的分配到不同的子节点上。子节点间的连接一般有三种常用的拓扑，集团连接、环状连接和中介点连接，如图1所示。集团连接传播速度快但是会导致较高的存储空间开销，规则性较低；环状连接空间开销小，规则性较好，但是传播速度慢；中介点连接在传播速度和空间开销上有优势，但是规则性较差。

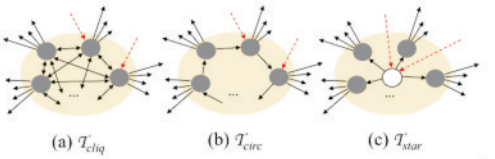


图 1　三种子节点连接拓扑

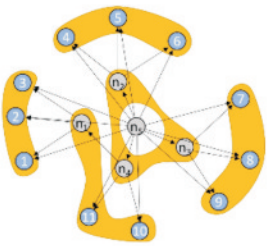


图 2　UDT转换示意图

Tigr针对上述连接方式存在的问题进行优

化，提出了UDT结构，如图2所示。在划分的过程中中介点连接直接创建新的中介点来连接子节点，而UDT采用递归的划分方式，产生一个新子节点并在分配了邻边之后将其加入待分配邻居队列作为下一个子节点的邻居，最终形成树状的连接方式。这样创建连接的好处是保证了各节点度数的一致性，在原来的拓扑基础上两点之间只会增加少数的传播跳转同时保证唯一的路径，满足了高效性和正确性。

基于物理存储的图数据进行拓扑转换需要较多的预处理时间并且消耗更多的物理存储空间。虚拟转换技术（virtual split transformation）的提出能够在保证高效、正确性的同时避免物理转换的开销。该技术在原始的物理层图数据上维持一个虚拟图拓扑层的映射，原始图数据的存储保持不变，用来进行值更新和传播，虚拟拓扑层被提供给以点为中心的编程框架来进行调度执行。具体的实现方式通过维护一个虚拟节点队列来实现，如图3所示该队列存放原始节点和切割之后节点之间的映射，当应用在CSR格式存储的图数据时，原始的节点队列被新的虚拟节点队列所替代，根据设置的度数上限K进行节点的划分和边的划分，这里K设为3，边按照CSR存储的顺序进行分配。其中不需要考虑子节点之间的连接关系，其执行的正确性可以通过原子操作等保证。虚拟转换可能会降低数据局部性，因为边表被划分为更多块分配给了不同的线程，因此作者还采用了一个边表数组联结的手段优化访存。

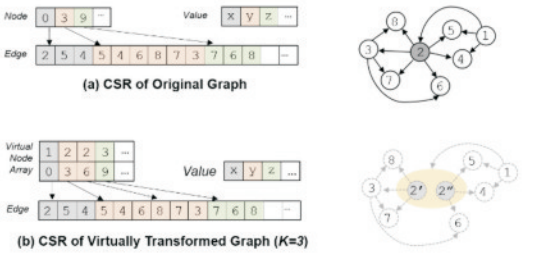


图 3 虚拟节点队列映射

文章基于以上提出的方法设计了一个轻量级的GPU图处理框架，并同现有的工作进行了对比，如图4所示。作者将Tigr同Cusha和Gunrock进行对比，在不同数据集上测试了大量的图算法，相比于Grunrock在大多数情况下取得了1.04到2.93倍的性能提升。其中PageRank算法需要在每次迭代中访问所有的点数据，因此Gunrock基于pull模式的执行比Tigr所采用的push模式要更适合该算法。

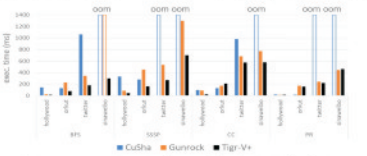


图 4 Tigr和现有工作的性能对比

Tigr还对物理UDT转换、虚拟转换以及边表访问优化手段达到的效果进行了对比，如图5所示。其中基于UDT的实现在执行过程中由于两点之间值的传播相比虚拟转换方式会有更多跳转从而增加收敛时间。对访存的优化可以较好的提升图计算的性能。此外作者还分析了UDT和虚拟转换在转换时间和存储空间上的开销，结果表明虚拟转换能够大幅度降低转换时间并支持更灵活的度数粒度。

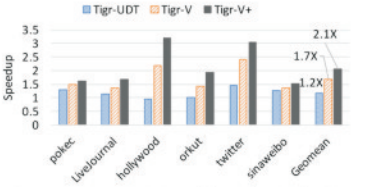


图 5 Tigr不同优化手段的性能对比

这篇文章相比于针对编程模型和底层硬件调度优化的工作能够从一个新的角度出发，将不规则图拓扑转换称为规则的图拓扑进行图计算，在减少开发难度的同时取得了较好的性能提升。虚拟化图数据的思想也能够给我们一些启示，从图数据上探索更多的优化空间。

桂创意
2017级博士研究生
研究方向：图计算
Email：chygui@hust.edu.cn

Finding Clues for Your Secrets: Semantics-Driven, Learning-Based Privacy Discovery in Mobile Apps

吴月明　推荐

“Finding Clues for Your Secrets: Semantics-Driven, Learning-Based Privacy Discovery in Mobile Apps”是国际顶级会议NDSS'18收录的一篇论文。本文通过分析手机APP的反编译代码，挖掘APP中的敏感数据，并采用静态的污点追踪技术检测隐私泄露行为。

在手机APP中，一段程序的本质可以看作是一个语义丰富的文档。这个文档包含了一些有特定意义的元素，比如：方法名称、变量、常量等。而这些具有特定意义的元素，可能直接涉及到了敏感数据。所以本文首先采用了自然语言处理（NLP）的技术来定位敏感数据，然后利用结构分析挖掘出真正的敏感数据，最后使用静态污点追踪来检测敏感数据是否被泄露。系统的整体流程如图1所示。

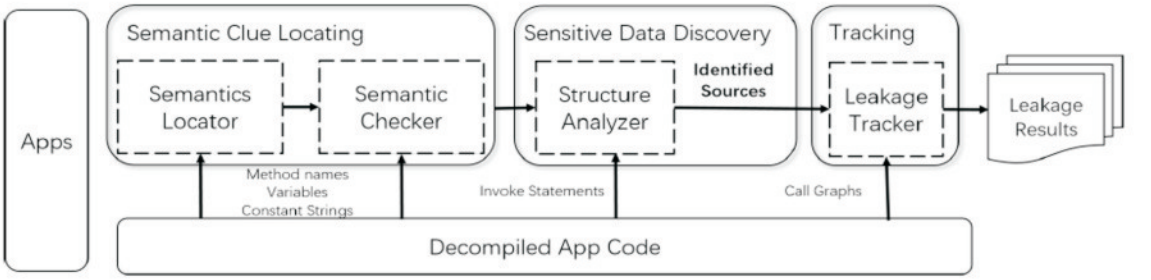


图1 ClueFinder系统框架图

表1 关键字列表

Category	Sample Keywords
User Attributes	first name, last name, gender, birth date, nick name, education, app list, device os, credit card, etc.
User Identifiers	user id, account number, access token, sina id, facebook id, twitter id, etc.
Location	latitude, longitude, lat, lng, user address, zip code, city, street, etc.
Account	account name, user name, phone number, mobile no, password, passwd, pwd etc.

最开始，本文从谷歌隐私策略中定义的隐私内容提取出了35个数据项，从以前和隐私相关的5篇研究文章中提取出了17个数据项。为了可以找到更多和敏感数据相关的关键字，文章采用word2vec搜索同义词，最后一共找到了121个关键字，可以分为以下4类，如表1所示。

然后，本文利用NLP来定位敏感数据。第一步是词干提取。将反编译得到APP代码，使用常用分隔符(e.g., user_addr)和大小写字母(e.g., getUserFbProfile)将所有的元素进行分词得到tokens。然后根据表1定义的4类，共121个和隐私相关的关键字，进行token搜索。最后将搜索到包含隐私相关token的元素标记为隐私相关的数据，以进行更深入的分析。第二步是词性标

注。敏感数据一般都是名称词性，比如在“address this problem”中的“address”就是一个动词，和隐私并不相关。第三步是语法分析。比如直接宾语：getAddressFromServer；

名词性主语：business phone number selected；否定修饰符：Do not input your password here 等。

接着，本文利用结构分析来挖掘敏感数据。比如图2中的第5行，包含敏感数据home_addr，但是整个句子的意思却只是检查是否包含home_addr，并不存在泄露的操作；第10行，包含隐私数据Profile，但是整个句子只是一个异常输出。为了了解整个句子的含义，本文从100个APP中随机选取了4326条语句，并人工标签是否包含隐私数据。然后选取了5类特征：1）方法名；2）参数类型；3）返回值类型；4）基准数类型；5）常变量模式。接着将这4326条语句作为训练集，构建一个支持向量机（SVM）来分类语句是否包含隐私数据。如果分类的结果显示包含了隐私数据，则该语句被定义为敏感源。

```
1 ## In co.snaptee.android.utils.FacebookFunctions
2 Json getUserFbProfile(HashMap userBasicInfo) {
3   JsonObject userJson = UserBasicInfo.toJson();
4   ## Gather other user information
5   If(userJson .contains("home_addr")){
6     jsonObject.put("home_addr", this.homeAddr);
7   }
8   this.uri = jsonObject.get("userProfile_uri");
9   if(this.uri == null) {
10    throwNullPointerException("Profile URI is
          null", exception);
11  }
12  return jsonObject;
13 }
```

图2 代码例子

最后，本文利用静态污点追踪来检测隐私泄露。前面确定了语句是敏感源之后，再从其参数或返回值中提取语句内的数据类型对象，针对这些对象进行数据流污点追踪，检测数据是否流转至sink。最理想的情况，是检测数据流是否被不信任的第三方库通过APIs发送至网络。如果被发送了，才为隐私泄露。但是，本文认为，只要隐私数据流转至不信任的库，则隐私数据就不再安全，其内容可以通过各种渠道向第三方披露。即只要检测到隐私数据被流转至不信任的库，就判定为泄露，不需要检测不信任的库是否将数据发送至网络。

为了验证系统的有效性，本文将ClueFinder和BibText进行了一个实验比较，一共选取了100个APP，实验结果如图3所示。实验结果显示，ClueFinder在更短的时间内找到了比BibText 更多的敏感数据，并且误报更少。

	BidText	ClueFinder
Detected sensitive data	131	177
Num. of false positives	19	15
Avg. Analysis Time (Sec)	97	55
Precision	83.5%	91.5%

图3 和BibText的实验比较

同样的，本文还选取了另一个工具SUSI，SUSI是NDSS‘14的一篇文章，但是SUSI仅仅只考虑了API是否含有敏感数据，却没有考虑语句的结构信息。在实验过程中，本文首先随机选择了15个流行的API，然后从随机选取的100个APP中提取了10116条包含这些API的语句，最后将这10116条语句作为实验对象。实验结果表明，SUSI的检测结果中有77.6%的误报率，而ClueFinder的实验检测结果为92.7的精准率和97.2%的查全率。并且ClueFinder一共找到了2266个敏感数据源。

本文提出了ClueFinder，一项新的手机隐私数据挖掘的技术。通过分析程序语句的语义信息，结合句子结构的语法分析，ClueFinder可以挖掘出更多隐藏的真实的敏感数据，最后利用静态污点追踪技术，可以检测更多的隐私泄露行为。

吴月明
2016级博士
研究方向：漏洞检测，隐私取证分析
Email：wuyueming21@163.com

Blockchain: Construction, Application and Research

报告人：穆　怡　教授
澳大利亚卧龙岗大学
时　间：2018年3月28日
整　理：陈天阳

穆怡教授对区块链进行了完整的介绍。穆教授首先介绍了区块链的基本构造，包括区块链的构成、交易的概念、账本的特性以及包括哈希、签名等在内的区块链基础组件；随后，从共识机制、挖矿、主链选择与交易执行等入手，介绍了区块链中关键操作流程及其意义；之后，介绍了典型的区块链项目，包括比特币、门罗币、以太坊等，并介绍了区块链的具体应用，如商品防伪、医药制造、IoT、选举投票等；最后，阐述了区块链的隐私保护等问题。

穆教授的报告内容十分丰富，生动形象地为我们阐明了区块链的来龙去脉及其重要的实用性意义，在场的老师与同学收获很大。在讨论环节，穆教授介绍了自己的研究方向与研究兴趣，并与我们共同探讨了密码学与区块链结合时产生的具体问题，给了我们很大启发。

穆怡教授是澳大利亚卧龙岗大学的全职教授，现任卧龙岗大学计算机安全研究所所长。穆教授研究兴趣包括密码学、网络安全、访问控制和计算机安全，已经发表了400余篇学术论文，其中许多文章发表在EUROCRYPT、ASIACRYPT、PKC、CT-RSA等顶级会议。穆教授是国际期刊《International Journal of Applied Cryptography》的主编，也是许多国际期刊的客座编辑，同时也曾担任过包括ACM CCS、ACM AsiaCCS、ESORICS、ACISP、ICICS等国际安全会议的程序委员会委员。

System Security Research: From Discovery to Innovation

报告人：王晓峰　教授
美国印第安纳大学
时　间：2018年3月26日
整　理：赵宏熠

王晓峰教授结合自身的研究成果，介绍了如何从发现系统工作的薄弱环节出发、找出系统设计存在的缺陷，并重塑系统安全设计。王教授先以他们在移动设备和物联网设备方面的研究为例，通过视频展示了即使在没有任何权限的情况下，手机应用也能通过系统底层公开的信息向外界泄露机主的地理位置、身份等隐私信息。随后，王教授介绍了他们当前正在研究的、关于自动发现系统漏洞方面的工作，通过收集CVE数据库中漏洞的函数和补丁信息，在补丁中找到修改的变量，然后对其进行模糊测试，从而发现系统中存在的已知漏洞。最后，王教授介绍了当前热门研究领域与安全相结合的机遇与挑战。

王晓峰教授的报告从身边的实际应用出发，生动而精彩的呈现了从发现问题、思考问题到解决问题、并给产品做出改进的科研过程。报告结束后，同学们结合自身工作，在模糊测试的实现、漏洞的触发方法等方面与王教授进行了深入的交流和讨论。

王晓峰教授在卡内基梅隆大学获得了电气与计算机工程博士学位，现任美国印第安纳大学计算机科学与工程系James H. Rudy荣誉教授、信息计算工程安全隐私中心的联合主任、ACM SIGSAC副主席、ACM CCS 2018 共同主席。王晓峰教授是人类基因组隐私的先驱研究员，也是iDASH基因组隐私竞赛的联合创始人。研究领域包括系统安全分析、生物医疗数据隐私、密码学、网络犯罪和物联网安全。

Privacy-Preserving Techniques to Blockchain: The Ring Signature Approach

报告人：廖启瑞　博士
澳大利亚莫纳什大学
时　间：2018年4月19日
整　理：袁　斌

首先，廖启瑞博士从区块链的基本介绍入题，介绍了区块链技术的发展现状、主要应用和前景。然后，廖博士介绍了区块链的几个核心研究问题，如智能合约、共识机制、数据一致性等。最后，廖博士详细介绍了其在环签名方面的工作，并介绍了环签名在区块链技术中的重要意义和实际应用情况。

廖启瑞博士的报告内容丰富，深入浅出地介绍了区块链和环签名相关的研究，拓展了同学们的视野。在提问环节，实验室的老师和同学与廖博士就秘密共享、区块链技术发展、区块链应用等方面的问题进行了深入探讨，给同学们诸多启发，打开了新的研究思路。

廖启瑞博士，澳大利亚莫纳什大学信息技术学院高级讲师，研究领域包括网络空间安全、区块链技术、应用密码学和隐私保护等方向。廖博士在许多国际顶级会议/期刊上发表了多篇极具影响力的文章，曾获得ESORICS 2014、ESORICS 2015、ACISP 2017等会议的最佳论文奖。此外，廖博士的研究工作有非常好的实际应用，其提出的关联环签名是著名电子货币Monero的理论基础。

Returning data control to users

报告人：高国隆（Ryan Ko）副教授
新西兰怀卡托大学
时　间：2018年4月28日
整　理：汤媛媛

高国隆教授在报告中主要介绍了如何让用户重掌数据控制权。首先，通过引述用户聊天记录被谷歌内部员工偷看和iCloud上的名人图片遭泄露等事件，高教授指出互联网给人们带来便利的同时，也让人们失去了数据控制权，因此让用户重掌数据控制权非常重要。然后，高教授介绍了可追踪文件数据活动的Progger（Provenance logger，内核空间记录器）。最后，高教授详细介绍了基于同态加密的云上移动电子投票方案，该方案在实现投票的同时保护了选民隐私。

高国隆教授的报告内容丰富，深入浅出，通过具体的应用场景，生动形象介绍了如何让用户重掌数据控制权。在报告的提问环节，高国隆教授和与会者就同态加密实用性、信任关系等问题进行了深入交流和讨论，为同学们提供了新的研究思路。

高国隆，新西兰怀卡托大学副教授，网络安全实验室主任，也是新西兰安全与犯罪科学学院的创办负责人。他的研究着眼于令用户重掌数据控制权、云计算安全与隐私所面临的挑战、数据溯源和同态加密。他同时有志于研究勒索软件传播方面的责任归属与安全隐患检测。高教授是爱斯威尔（Elsevier）安全系列书籍的编纂人，曾获得三个最佳论文奖（ICCCRI 2015，USEWOD 2011，WWW 2011），和2015年信息与软件安全认证会员制组织（ISC）2亚太信息安全领袖奖。

感恩日常

李辉楚吴

借准备这篇文章的机会，我从这规律且紧凑的生活节奏中暂时跳出来，让自己好好回顾一下这两年的生活。我觉得这两年的生活经历，是对“我们所度过的每个平凡的日常，也许就是连续发生的奇迹”（出自《日常》，由京都动画制作，于2011年4月播出）最好的诠释。

两年前的某天上午，我继续着日常的工作，一切都是那么的规律。突然，脑中凭空就出现了一个念头：十年之后的我会是怎样一个状态？答案是，看得见的稳定——这使我感到很惊恐。于是乎，我决定逃出自己的舒适圈，趁着年轻看看自己的潜力。回顾当时的情况，我为自己的决定而庆幸。离职前夕，不管是出于有意或者无意，同事们都在向我传递一个信息：你已经工作3年了，即使考上研究生，未来也不一定有现在这么稳定。神奇的是，我听了他们的话之后，却悟出了一个道理：每个人都不可能独立于世，当我们决定跳出自己的舒适圈时，可能会让我们所处的环境感到不适，从而自发的希望我们退回去，因此，改变自己的态度一定要坚决果断。假如当初态度不够坚决，我很可能会安于之前的工作，而不是在这和各位老师同学们交流自己的心得感受。

实验室的生活对我来说即熟悉又陌生，熟悉的是青春而单纯的校园生活，陌生的是对学术的热情。研一上学期是个缓冲的阶段，日常活动中除了科研还有修课的任务。听金老师讲授《并行处理》这门课的感觉就是“痛快”。最让我感到佩服的是，金老师日程安排如此忙碌的同时，还能够不断地更新授课的素材。比如2017年10月8日“鹿晗&关晓彤公布恋情搞炸微博服务器”事件发生的第二天，当时很多学生还没听说这件事的时候，金老师就在课堂上从技术的角度对这件事情进行了分析。这只有在平凡的日常中不断地重复着对科研、对教育的极致追求，才能做到课堂上的举重若轻吧。在工作生活中，另一位让我佩服的人就是我的导师——肖江。为了让我这个科研小白快速的进入状态，肖老师专门抽出时间，悉心的向我传授阅读论文的技巧。若不是肖老师倾囊相授，我可能要读过成吨的论文之后才会领悟一些阅读论文的技巧。比如，读完论文题目之后不要立马接着看摘要部分，而是先想一想“我会怎么做？”，然后带着问题有针对性的阅读。平时交流时，肖老师也会分享一些学习的方法。举个例子，当学习一个新的概念时，我之前的习惯是“弄懂为止”，但是一段时间之后记忆就模糊了。经肖老师提点，我意识到所有知识的产生都是由问题驱动的，因此在学习新知识的同时我们应该挖掘一下这个知识产生的背景，只有这样的知识才是完整的，不易遗忘的。除了阅读论文、做实验以外，我们这个小组还有一个日常工作就是准备周会的PPT。可能有些同学认为平均每周制作一个PPT太浪费时间，会耽误了课题进度。三年的工作经验却告诉我，肖老师的这个安排很合理，而且会让我们的今后的职业生涯受益匪浅。

在CGCL生活了近一年，认识了很多老师和同学，虽然性格、角色不同但是都很可爱。这个大家庭里的每一个人都很认真的经营着自己的日常。也许，咱们这个家庭的下一个奇迹就在我们现在这样努力的日常中出现吧。

李辉楚吴
2017级硕士研究生
研究方向：智能感知
Email：credolee@hust.edu.cn