



中国计算机学会
China Computer Federation

CCFSys 2023

2023年CCF计算机系统大会

智能时代的计算机系统：从研究到产业化

会议手册

主办单位 中国计算机学会

承办专委 体系结构专业委员会、分布式计算与系统专业委员会

承办单位 南昌大学

协办单位 华东交通大学、CCF南昌分部

钻石赞助  蚂蚁技术研究院  中科曙光

赞助  龙芯中科  HUAWEI  阿里云  RICORE  奇安信  魏动科技  AVNET

2023/8/4-5

江西·南昌

Version 08.02



最新、电子版会议手册下载
信息更新详见最新电子版手册

目录 CATALOGUE

一、会议须知	02
二、大会日程	07
三、大会组织机构	08
四、大会简介	13
五、中国计算机学会(CCF)简介	13
六、CCF体系结构专委会简介	14
七、CCF分布式计算与系统专委会简介	14
八、南昌大学简介	15
九、大会特邀报告	17
大会特邀报告(8月4日上午)	17
大会特邀报告(8月5日上午)	20
十、分论坛日程	23
十一、赞助企业	77
蚂蚁集团	78
中科曙光	79
龙芯中科	80
苏州睿芯集成电路科技有限公司	81
奇安信集团	82
趋动科技	83

会议须知

1.会议名称: CCF 计算机系统大会(CCF Sys 2023)

2.会议时间: 2023年8月3日:

上午、下午:报到

下午(13:30 – 17:30):前导论坛

2023年8月4 – 5日:

上午(08:00 – 12:00):特邀报告

下午(13:30 – 17:30):技术论坛

3.会议地点: 南昌国际博览城绿地铂瑞酒店

地址: 南昌市红谷滩新区上饶大街1100号

(乘坐地铁2号线至国博站, 出4号口即到)

4.官方网站: <https://conf.ccf.org.cn/CCFSys2023>



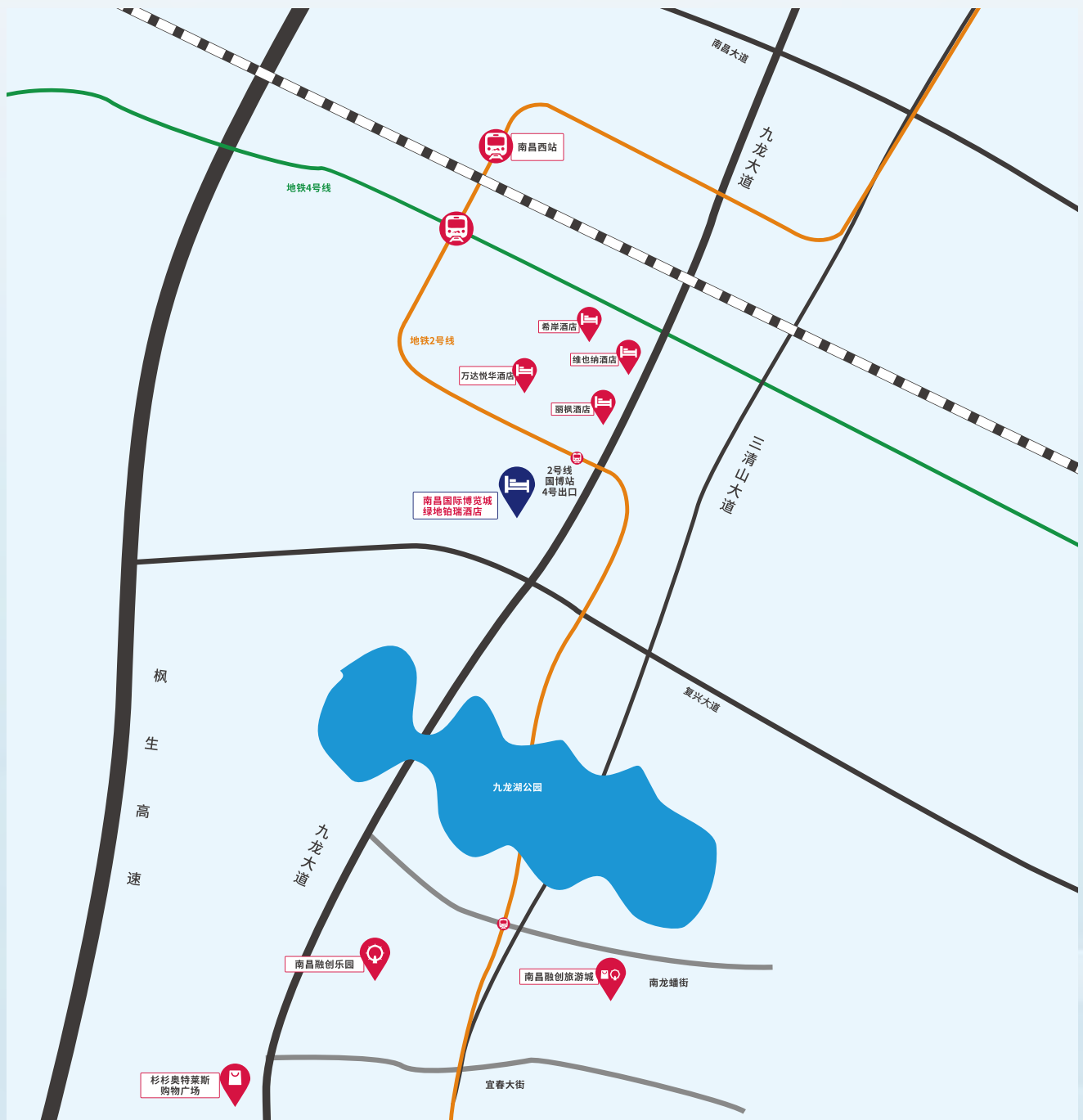
酒店信息

南昌国际博览城绿地铂瑞酒店

Primus Hotel Nanchang International Expo City

电话：0791-86198888

地址：江西省南昌市红谷滩区上饶大街1100号





酒店交通路线

昌北机场→绿地铂瑞酒店

昌北机场巴士转乘地铁2号线 - 约1小时20分钟 - 约20元



打车 - 约43分钟 - 约90~110元

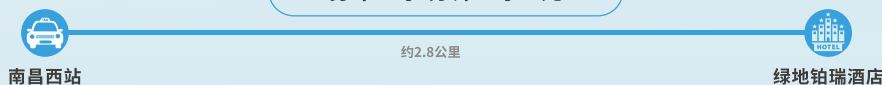


南昌西站(高铁站)→绿地铂瑞酒店

地铁2号线 - 约5分钟 - 2元



打车 - 约8分钟 - 约11元



南昌火车站→绿地铂瑞酒店

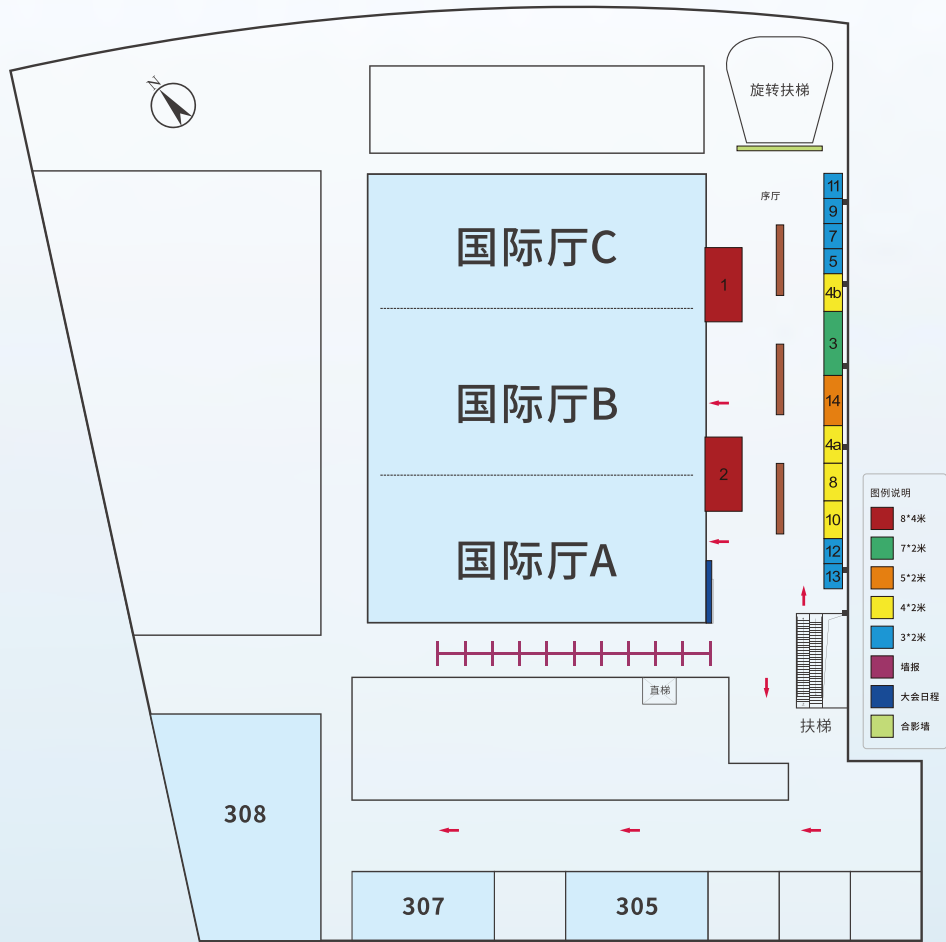
地铁2号线 - 约60分钟 - 5元



打车 - 约31分钟 - 约60元



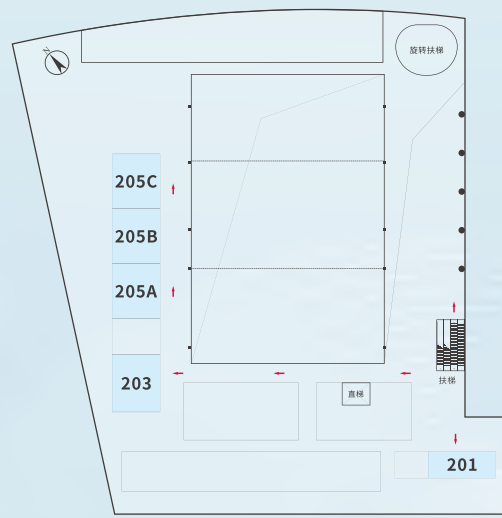
会场分布



三楼平面图



一层平面图



二层平面图

展位地图

三层会场东侧



大会日程

2023年8月3日					
全天	09:00 - 18:00	会议注册 (铂瑞酒店大堂)			
下午	13:30 - 18:00	1-体系结构优秀青年学者与博士生发展论坛	203	李超、侯小凤	P23
晚上	18:00 - 20:00	自助晚餐 (三层国际厅C)			
	19:30 - 20:00	分布式计算与系统专委会年度工作会议	308	叶保留	
	20:00 - 22:30	分布式计算与系统专委会换届工作会议	308	陈贵海	
2023年8月4日					
上午 主会场	08:30 - 08:45	大会主席致辞	三层国际厅AB	钱德沛 院士	主持人: 武成岗 研究员
	08:45 - 08:50	本地组织单位致辞		罗嗣海 南昌大学党委书记	
	08:50 - 09:00	学会领导致辞		金海 教授	
	09:00 - 09:50	工业软件的发展之路		何积丰 院士	
	09:50 - 10:10	茶歇			
	10:10 - 11:00	新形势下高性能计算的可持续发展	三层国际厅AB	钱德沛, 院士	主持人: 李涛 教授
	11:00 - 11:30	System Research is Dead, Long Live System Research!		何征宇, 蚂蚁集团平台技术事业群总裁	
	11:30 - 12:00	CCF-蚂蚁软硬协同基金发布		钱德沛、何征宇	
	12:00 - 13:30	自助午餐 (一层铂瑞厅ABC)			
下午 分论坛	13:30 - 18:00	2-概率型数据结构设计及其应用	305	杨全、罗来龙	P26
		3-智能边缘计算体系结构与软硬件系统	205A	周知、吴松	P29
		4-算力网技术论坛	三层国际厅A	郭得科、彭晓晖	P32
		5-区块链分布式账本技术论坛	205C	黄华威、郑子彬	P32
		6-新黄金时期的处理器安全	205B	吕勇强、胡伟	P39
		7-高性能图计算体系结构和系统软件	307	张宇、赵进	P43
		8-高性能互连论坛	203	董德尊、雷斐	P46
		9-量子计算系统	201	付祥、劳玲玲	P49
		18-面向下一代异构计算的软件栈建设	308	翟季冬、李斌	线下研讨会。 不允许拍照、录音、录像
		晚上	18:30 - 21:00	晚宴 + 颁奖典礼 (一层铂瑞厅ABC)	
2023年8月5日					
上午 主会场	09:00 - 09:50	人工智能应用系统内生安全问题与检测技术研究	三层国际厅AB	邬江兴, 院士	主持人: 廖小飞 教授
	09:50 - 10:40	大力开展基于国产AI芯片的大模型系统的研究		郑纬民, 院士	
	10:40 - 11:10	茶歇			
	11:10 - 12:00	Towards Data-Centric Computer Architectures	三层国际厅AB	Per Stenström	主持人: 李超 教授
	12:00 - 13:30	自助午餐 (一层铂瑞厅AB)			
下午 分论坛	13:30 - 18:00	10-Tutorial: 量子计算	201	刘英文、刘雍	P52
		11-新一代数据中心网络论坛	205A	黄家玮、崔勇	P55
		12-AI大模型应用与基础系统创新	308	金雪峰、李诚	P58
		13-安全计算系统论坛	205B	王喆、申文博	P61
		14-分布式工业网络与计算系统	307	姜楠、邱铁	P65
		15-教学论坛	305	沈立、李涛	P67
		16-RISC-V生态关键技术	203	夏虞斌、包云岗	P70
		17-隐私计算软硬协同加速技术	205C	闫守孟、路航	P73
晚上	18:00 - 20:00	自助晚餐 (一层铂瑞厅A)			
	19:00 - 22:00	体系结构专业委员会2023年度工作会	308	沈立	



大会组织机构

大会指导委员会



郑伟民院士



庄松林院士



陈左宁院士



陈国良院士



邱江兴院士



何积丰院士



张尧学院士



高孟飞院士



顾敏院士



王怀民院士



孙凝晖院士



钱德沛院士



金海教授



武成岗研究员



廖小飞教授

大会组织机构

大会共同主席



钱德沛 院士
北京航空航天大学



金海 教授
华中科技大学

大会执行主席



武成岗 研究员
中国科学院计算
技术研究所



廖小飞 教授
华中科技大学

大会组织委员会主席



李涛
南开大学



叶保留
南京大学

大会组织委员会副主席



王蕾
国防科技大学



陈旭
中山大学

大会论坛共同主席



李东升
国防科技大学



郭得科
国防科技大学

大会论坛共同副主席



王喆
中国科学院计算
技术研究所



翟季冬
清华大学

大会程序委员会主席



李超
上海交通大学



吴帆
上海交通大学

大会本地主席



徐子晨
南昌大学



姜楠
华东交通大学

大会宣传主席



李诚
中国科技大学



于东晓
山东大学



大会组织机构

大会宣传副主席



付祥
国防科技大学



魏崑
西安理工大学

大会网站主席



谢学说
先进计算与关键
软件海河实验室



詹玉峰
北京理工大学

大会注册主席



董德尊
国防科技大学



李鑫
南京航空航天大学

大会企业合作主席



裴颂文
上海理工大学



田臣
南京大学

大会专刊主席



李超
上海交通大学

大会投稿主席

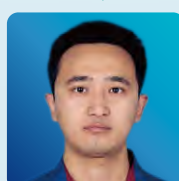


宫晓利
南开大学

大会学生资助主席

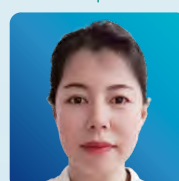


王超
中国科技大学

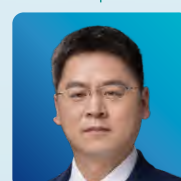


徐子川
大连理工大学

大会财务主席



赵云华
中国科学院计算
技术研究所



曾德泽
中国地质大学

大会奖励主席

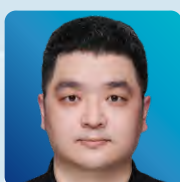


王毅
深圳大学



陆桑璐
南京大学

大会竞赛主席



陆佳华
Xilinx公司



张福新
中国科学院计算
技术研究所



何强
华中科技大学

大会组织机构

国际先进并行处理技术会议(APPT)程序委员会

包云岗	中国科学院计算技术研究所	孙广宇	北京大学
常 瑞	浙江大学	王 毅	深圳大学
陈 全	上海交通大学	王 颖	中国科学院计算技术研究所
邓 宇	飞腾信息技术有限公司	王 蕾	军事科学院
翟季冬	清华大学	王 喆	中国科学院计算技术研究所
董德尊	国防科技大学	王海霞	清华大学
杜子东	中国科学院计算技术研究所	王耀华	国防科技大学
方 娟	北京科技大学	王则可	浙江大学
宫晓利	南开大学	魏继增	天津大学
侯 锐	中国科学院信息工程研究所	翁楚良	华东师范大学
侯小风	上海交通大学	吴晨涛	上海交通大学
胡 杨	清华大学	徐子晨	南昌大学
华 宇	华中科技大学	徐子川	大连理工大学
计卫星	北京理工大学	杨海龙	北京航空航天大学
李 超	上海交通大学	叶可江	中国科学院深圳先进技术研究院
李 诚	中国科学技术大学	殷 数	上海科技大学
李东升	国防科技大学	喻之斌	华为技术有限公司
梁 云	北京大学	张 宇	华中科技大学
梁晓晓	上海交通大学	张为华	复旦大学
刘 昊	军事科学院	张一鸣	厦门大学
刘 彦	湖南大学	张殷乾	南方科技大学
刘海坤	华中科技大学	赵倩雅	浪潮集团有限公司
陆游游	清华大学	郑 龙	华中科技大学
罗英伟	北京大学	Huanle Xu, University of Macau	
吕勇强	清华大学	Peng Li, The University of Aizu	
马 胜	国防科技大学	Wei Li, The University of Sydney	
沈 立	国防科技大学	Bingsheng He, National University of Singapore	
侍 展	华为技术有限公司		



大会组织机构

全国开放式分布与并行计算学术年会(DPCS)程序委员会

曾德泽	中国地质大学(武汉)	彭晓晖	中国科学院计算技术研究所
崔禾磊	西北工业大学	钱柱中	南京大学
崔来中	深圳大学	秦建斌	深圳大学
戴海鹏	南京大学	申兆岩	山东大学
东方	东南大学	沈蒙	北京理工大学
宫晓利	南开大学	田晓华	上海交通大学
谷峪	东北大学	童咏昕	北京航空航天大学
顾荣	南京大学	王雷	大连理工大学
韩光洁	大连理工大学	王晓飞	天津大学
何源	清华大学	吴帆	上海交通大学
贺诗波	浙江大学	吴远	澳门大学
黄倩怡	中山大学	吴巍炜	香港城市大学
焦磊	俄勒冈大学	惠维	西安交通大学
金梦	上海交通大学	肖江	华中科技大学
金熠波	华为技术有限公司	徐梦炜	北京邮电大学
李诚	中国科学技术大学	张灏	华为技术有限公司
李鑫	南京航空航天大学	张胜	南京大学
李福亮	东北大学	张圣林	南开大学
李士刚	北京邮电大学	张岩峰	东北大学
李雨森	南开大学	张志威	北京理工大学
李振华	清华大学	赵东	北京邮电大学
凌振	东南大学	赵乙	清华大学
刘佳	南京大学	赵来平	天津大学
刘凯	重庆大学	郑臻哲	上海交通大学
刘思聪	西北工业大学	周知	中山大学
卢立	浙江大学	周睿婷	东南大学

大会简介

为促进计算机系统研究的学术交流,了解计算机系统国际最新学术动态,由中国计算机学会(CCF)主办,CCF 体系结构专委、分布式计算与系统专委、南昌大学联合承办的“CCF计算机系统大会(CCF Sys)”,将于2023年8月4-5日在南昌召开。

本次大会聘请到郑纬民、庄松林、陈左宁、陈国良、邬江兴、何积丰、张尧学、杨孟飞、顾敏、王怀民、孙凝晖、钱德沛等12位中国科学院、中国工程院(外籍)院士,金海、武成岗、廖小飞等专家学者组成指导委员会,并邀请到郑纬民院士、邬江兴院士、何积丰院士、钱德沛院士、ACM/IEEE会士 Per Stenström等重磅嘉宾做大会特邀报告。大会集结全国计算机、人工智能、电子信息等学科领域从事计算机系统相关的研究与技术开发的专家学者、研究人员,分享技术研发经验与合作需求,促进产学研合作,为国内外科研机构、高校、企业搭建最广阔、最深入的学术交流平台。



中国计算机学会(CCF)简介

中国计算机学会(CCF)成立于1962年,全国一级学会,独立社团法人,中国科学技术协会成员。

CCF是中国计算机及相关领域的学术团体,宗旨是为本领域专业人士的学术和职业发展提供服务;推动学术进步和技术成果的应用;进行学术评价,引领学术方向;促进技术和产业应用一线的交流 and 互动;对在学术和技术方面有突出成就的个人、企业和单位给予认可和表彰。

秉承会员构成、会员治理和服务会员的理念,CCF一切工作的出发点和落脚点是服务会员和业界专业人士,CCF每年举办几百场/次各种规模、层次的学术会议、产业与技术论坛和培训,开展评奖、竞赛、计算机职业资格认证、科学普及、计算机术语审定等系列活动。

CCF以承担社会责任为己任,“CCF走进高校”,每年举办百场公益免费演讲,帮助大学生进行职业规划,奠定专业成长基础;“CCF吕梁教育扶贫”,把现代教育理念带进贫困闭塞的山区,促进城乡教育资源流动和共享;CCF每年资助上百位缺乏资金的高校师生参加中国计算机大会(CNCC),让他们和顶尖团体一起同步学习、开拓视野;CCF积极承担社会责任,关注教育、人才、学术评价、产业政策等公共话题,就重大政策及相关问题向政府建言或向公众发声。

CCF下设16个工作委员会,40个专业委员会,38个地方会员活动中心。依托立体多维的治理架构,通过上述各种形式的品牌活动,服务于全国乃至全球的计算机专业人士、企业、科研机构和组织。CCF编辑出版的旗舰刊物有《中国计算机学会通讯》(学术性月刊),与其他单位合作编辑出版的会刊13种。CCF与IEEE-CS、ACM等国际学术组织有密切的联系和合作。CCF实行会员制,分为个人会员和公司会员,任何从事计算机及相关领域专业工作的人士和单位均可申请成为CCF会员。欢迎加入中国计算机学会。

CCF体系结构专委会简介

中国计算机学会体系结构专业委员会成立于1989年。专委会致力于团结、联合、组织计算机体系结构领域的专业人士，开展学术交流等相关活动，提高我国计算机体系结构领域的科研、教学、应用水平，促进研究成果的应用和向产品的转化，提升在国家科技活动和国际学术方面的影响力。

专委会的日常学术交流活动包括举办全国计算机体系结构学术年会、高级并行处理国际会议等学术会议；举办体系结构前沿技术和相关产业研讨会；组织学会的专家、教授参与国家相关产业政策制定，咨询和规划等。

全国计算机体系结构学术年会紧扣体系结构发展热点，体现了当前我国计算机体系结构与相关技术及理论研究水平。高级并行处理国际会议（APPT）由我专委会发起，中、德双方联合组织，自1995年起每两年一次，轮流在中国和瑞典、西班牙等欧洲国家举办。APPT会议对于加强我国体系结构领域的国际学术合作和学术交流起到了重要作用，一直受到中、欧双方高度重视。2001年，专委会还成功主办第四届亚太地区WEB技术国际会议。

今后，专委会将在中国计算机学会的指导下，继续加强与国家自然科学基金委员会、ACM SIGARCH、YOCSEF以及其他国内外学术团体的联系，并保持与全体委员的经常性沟通和联系；更多地向年轻人才倾斜，为他们提供展示才华的舞台；积极与产业界联合，举办前沿技术和相关产业发展研讨会，推动我国信息技术产业发展。同时，争取更多的公司加入到学术活动中来，促进国内外同行的学术交流和合作，使我们的学术活动上升到一个新的台阶。

CCF分布式计算与系统专委会简介

中国计算机学会（CCF）分布式计算与系统专业委员会，前身为1981年8月3日成立的“中国电子学会计算机学会微程序设计及分布式系统专业学组”，1988年10月更名为“CCF分布计算系统与固件工程专业委员会”，1994年9月扩建成“CCF开放系统专委会”，2015年8月根据学科发展更名为“CCF分布式计算与系统专业委员会”。作为CCF较早成立的专委会之一，分布式计算与系统专业委员会为推动专业委员会建设、推进计算机类专业教育改革等作出重要贡献，孙钟秀院士、陈国良院士、陈道蓄教授等资深委员获CCF创建60周年杰出贡献奖。

分布式和并行是提高计算系统处理能力的两个重要途径。在中国计算机学会的领导与支持下，专委会组织和团结国内分布式计算技术、分布式并行处理系统、分布式泛在网络等专业领域人才，结合新型开放网络环境下的分布式技术与并行处理研究趋向及产业化需求，加强学术交流、提升学术影响、促进产学研合作，推进我国新一代分布式计算与系统技术研究，服务相关产业的发展。

在孙钟秀院士、陈国良院士、金士尧教授、鞠九滨教授等老一辈科学家的带领与影响下，专委会队伍不断壮大，汇集了国内分布式计算与并行处理系统领域最优秀的专业人才，一批才俊入选国家级高层次人才及青年人才，引领分布式计算与系统领域发展，取得令人瞩目的创新成果，展现出良好的发展潜力。

专委会每年均会组织丰富多彩的学术活动，如一年一度的全国开放式分布与并行计算学术年会、各类前沿技术和产业研讨会、以及不定期举办的走进高校/走进企业等，旨在为分布与并行计算的相关从业人员构建更好的平台，提供更好的服务，共同推进分布式计算与并行处理系统领域的发展。

南昌大学简介

南昌大学是国家“双一流”建设高校、教育部与江西省部省合建高校、江西省一流大学整体建设高校。学校地处“英雄城”南昌市，拥有前湖、青山湖、东湖3个校区，其中前湖主校区占地面积4321亩，校舍建筑面积150万平方米。

学校前身江西医学院、江西大学和江西工业大学。党和国家领导人、教育部和江西省委、省政府高度重视学校的建设与发展。2008年10月，习近平同志视察学校时深情寄语“南昌大学前景无限”。2016年2月，习近平总书记再次亲临学校视察，就高校科研创新和人才培养发表重要讲话，并勉励学校走创新发展之路，希望当代大学生珍惜韶华，用青春铺路，让理想延伸。2017年学校进入国家“双一流”世界一流学科建设高校行列。2018年学校成为教育部与江西省部省合建高校。2021年2月，《江西省国民经济和社会发展第十四个五年规划和二〇三五年远景目标纲要》中明确提出“举全省之力办好南昌大学，争取进入世界一流大学建设高校行列”。



学校现设有人文学部、社科学部、理工一部、理工二部和医学部，拥有12个学科门类的132个本科专业，3个国家重点(培育)学科，22个博士学位授权一级学科，3个博士专业学位授权类别，11个博士后科研流动站；51个硕士学位授权一级学科，31个硕士专业学位授权类别。学校设有5所直属附属医院，共有9个国家临床重点专科。材料科学与工程学科入选国家“双一流”建设学科，6个学科入选省一流学科。包括计算机科学与技术在内12个学科进入ESI全球排名前1%。

进入新时代，南昌大学坚持以立德树人为根本，积极落实“人才强校、特色创新、产教融合”三大战略，遵循“育人为本、创新引领、合建驱动、改革攻坚、实干兴校、力争一流”的发展思路，求真务实，开拓创新，朝着创建具有“江西底色、中国特色”的世界一流大学的宏伟目标奋勇前行！



大会特邀报告

(8月4日上午)



何积丰

工业软件的发展之路

何积丰, 中国科学院院士, 华科智谷人工智能研究院院长, 著名计算机软件科学家, 现任中国信息技术标准化技术委员会软件与系统工程分技术委员会主任委员、国家可信嵌入式软件工程技术研究中心首席科学家、教育部可信软件国际合作联合实验室主任、上海市人工智能战略咨询专家委员会委员。

何院士是程序统一理论创立者, 数据精化完备理论奠定者, 基于模型的可信软件设计技术的开拓者。何院士在世界上首次提出“可信人工智能”, 自此世界各地开始广泛关注人工智能的安全可信问题。曾担任国家自然科学基金委重大研究计划, 科技部973计划、863计划主题项目首席科学家, 领衔国家自然科学基金委创新研究群体, 出版英文专著2部, 在国际期刊和会议上发表论文160余篇, 他引4000余次。以唯一完成人荣获国家自然科学基金二等奖和上海市科技进步奖一等奖各1项; 以项目第一完成人荣获上海市科技进步奖特等奖; 以第一完成人获省部级科技进步奖与科技成果奖一等奖4项, 两次荣获英国女王先进技术奖, 以及获得法国骑士勋章。另外授予全国五一劳动奖章, 上海市教育功臣和上海市科技功臣等称号。

报告摘要:

工业软件的特征; 技术发展路线; 未来展望。



钱德沛

新形势下高性能计算的可持续发展

钱德沛,男,北京航空航天大学教授、中国计算机学会会士。长期担任国家863计划计算机主题专家组和信息领域专家组专家,曾任863计划和国家重点研发计算高性能计算方向4个重大项目、重点专项的总体专家组组长。曾主持或参与国家自然科学基金项目、863计划课题、973子项目和其他科技项目二十余项。主要研究方向为高性能计算机体系结构、分布式系统、网络计算、多核/众核编程支持等。发表论文400余篇。

报告摘要:

本报告首先简要回顾过去20余年我国高性能计算发展的历程、成绩与历史经验,然后引入了高性能计算在问题求解模式、算力形式的多样化、外部环境和国家对算力的需求等方面出现的新变化,分析了这些新变化带来的新的技术挑战。为应对新挑战,提出了实现高性能计算可持续发展所需要关注的若干研究问题,最后提出建立国家超算互联网的愿景。



何征宇

System Research is Dead, Long Live System Research!

何征宇, 蚂蚁集团平台技术事业群总裁, 佐治亚理工学院博士, 杭州市第十四届人民代表大会代表, 2019年入选“国家海外高层次人才引进计划”(创新长期)专家, 2021年入选世界经济论坛年度全球青年领袖, 先后就职于谷歌、蚂蚁集团。2012年加入谷歌, 创立开源项目 Google gVisor, 实现了容器更安全的隔离。2018年加入蚂蚁集团, 负责公司技术基础设施建设, 主导了绿色计算实践、机密计算创新、金融级云原生操作系统研究创新等重大项目。参与国家专项“2020年关键软件适配验证中心项目-数据库适配验证中心”, 其负责的“金融级云原生操作系统研究创新团队”获杭州市领军型创新创业团队。

报告摘要:

摘要: 在AI大发展的时代, 有观点认为系统研究已经无足轻重了。真的如此? 系统研究的未来将走向何方? 本报告将分享蚂蚁集团在系统研究方面的一些观察、思考、和展望。

大会特邀报告

(8月5日上午)



邬江兴

人工智能应用系统内生安全问题与检测技术研究

邬江兴, 中国工程院院士, 国家数字交换系统工程技术研究中心主任、教授、博士生导师, 国际知名网络交换专家和我国信息通信领域产业跨越式发展的引领者。先后担任国家“八五”“九五”“十五”“十一五”863计划通信技术主题专家、副组长、信息领域专家组副组长, 国家移动通信重大专项论证委员会主任, 国家“三网融合”专家组副组长, 国家下一代广播电视网专家委员会主任, 原国家奖励委员会委员。

获国家科技进步奖创新团队奖1项, 国家科技进步一等奖3项、二等奖4项, 国家教学成果一等奖1项, 何梁何利基金科学与技术成就奖, 王丹萍科学基金奖, 中国APEC工商咨询理事会代表工作十年杰出贡献奖。获评全国优秀科技工作者, 国家863计划突出贡献者, 国家科技攻关计划突出贡献者, 国家有突出贡献中青年专家。2013年入选改革开放35周年重要活动“改革时代人物志”, 2008年、2018年, 其主持研制的HJD04机先后入选《改革开放30年科技成就撷英》和《改革开放40年科技成就撷英》并作为永久展品入选中国共产党历史展览馆。

报告摘要:

当前, 人工智能(Artificial Intelligence, AI)被认为是引领新一轮产业变革的核心技术引擎。AI应用系统作为国家“新基建”发展战略的重要组成部分, 如果其可用、可靠、可信性问题不能有效处置, 将会成为数字经济时代的社会性软肋、麻烦策源地或新型社会公害! 因此, AI应用系统的安全性问题即不能再被选择性忽视, 更不能再走先发展后治理的老路!

本汇报首先从AI应用系统之内生安全问题展开分析, 以新的视角讨论并得出“AI应用系统安全主要隐患源于内生安全个性问题和共性问题”的结论, 并介绍了内生安全之广义功能构造及其安全构造效应, 阐述了利用“构造决定安全”的系统工程理论破解AI应用系统内生安全问题之理论可行性; 最后, 借助广义功能安全构造(DHR构造)的未知威胁与破坏感知效应, 研究AI应用系统的内生安全检测技术, 通过实验案例验证了其工程实践上的有效性。



郑伟民

大力开展基于国产AI芯片的大模型系统的研究

郑伟民，清华大学计算机系教授，中国工程院院士。长期从事高性能计算机体系结构、并行算法和系统研究。提出可扩展的存储系统结构及轻量并行的扩展机制，发展了存储系统扩展性理论与方法。在国内率先研制并成功应用集群架构高性能计算机。在国产神威太湖之光上研制的极大规模天气预报应用获得ACM Gordon Bell奖。曾获国家科技进步一等奖1项、二等奖2项、国家技术发明二等奖1项，何梁何利科技进步奖，获得首届中国存储终身成就奖。

报告摘要：

训练AI大模型对于算力有非常强的需求，但当前面临的情况是价格非常高的国外GPU卡一卡难求，而国产的AI卡不太喜欢用。怎么办？

一是要改变国产卡的生态系统不好的局面，需要做好国产软件生态的几件事，包括编程框架，并行加速，通信库，算子库，AI编译器，编程语言，调度器、内存分配、容错系统、存储系统等支撑软件。二是要做好软硬件协同设计的整体系统工程化。这两件事情做好了，基于国产AI芯片的大模型系统一定会做好。



Per Stenström

Towards Data-Centric Computer Architectures

There are two trends that will have a significant impact on how to sustain an exponential computational performance growth at a reduced power consumption in the future. One trend is that applications have shifted from being compute centric to data centric and the other trend is that all technology scaling laws (Dennard scaling and Moore's Law) have or will soon come to an end. In this talk, I will elaborate on the implications of these trends on the design of computer systems in the future and give a few glimpses on work in my research lab being underway towards data-centric computer architectures.

报告摘要:

Per Stenström is a professor of computer engineering at Chalmers University of Technology since 1995. Before coming to Chalmers he was on the faculty of Lund University where he received his MSc in Electrical Engineering and Ph D in Computer Engineering. His research program centers on computer architecture with a current research emphasis on design principles for chip multiprocessors. His publications in the past include more than a hundred published papers and three textbooks. Per Stenström has been a visiting scientist at Carnegie Mellon University (1987-1988), Stanford University (1991), and University of Southern California (1993) and in 2003, he spent a sabbatical leave at Sun Microsystems. He was general chair of the 28th Annual International Symposium on Computer Architecture (ISCA) in 2001 and program chair of the same symposium in 2004. In 2008 he was program co-chair of the IEEE International Symposium on High-Performance Computer Architecture (HPCA) held in Salt Lake City and in 2009 I was the program chair for the IEEE IPDPS held in Rome. He is a co-founder of the International Conference on High-Performance and Embedded Architectures and Compilers and acted as its general co-chair in 2008 and program co-chair for the same in 2007 and is now Steering Committee chair for this conference series. He is also spending a considerable amount of time as a co-founding partner of the HiPEAC network. Per Stenström is a Fellow of the IEEE and the ACM. Since 2009 he is a member of the Royal Swedish Academy of Engineering Sciences and since 2010 of the Academia Europaea.

分论坛日程

论坛1:体系结构优秀青年学者与博士生发展论坛

时间:2023年8月3日13:30 - 17:30

如今被誉为体系结构研究的黄金时代。在计算机体系结构领域,国内近年来先后涌现出一批优秀的青年学者和具有潜力的博士生。体系结构专委会近年来连续多次在ACA和APPT上邀请一部分科研新星做学术报告,分享前沿成果,交流科研心得,促进青年学者和博士生的进一步成长。

具体日程安排(论坛1)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	Domain-Specific System and Architecture	钱学海/Purdue University
14:00 - 14:20	芯片安全计算架构	王文浩/中国科学院信工所
14:20 - 14:40	一种加速高精度计算的专用计算架构设计	郝一帆/中国科学院计算技术研究所
14:40 - 15:00	面向图计算的领域通用加速器	姚鹏程/华科大&之江实验室
15:00 - 15:20	茶歇	
15:20 - 15:40	大语言模型量化在体系结构中的挑战和机遇	郭聪/上海交大
15:40 - 16:00	通过算法和架构的协同设计加速智能合约执行	潘锐/湖南大学
16:00 - 16:45	Panel 讨论:青年人如何做原创性研究?	
16:45 - 17:30	体系结构设计竞赛决赛答辩(45分钟)	

论坛主席:



李超

上海交通大学

李超,上海交通大学教授,博士生导师。致力于研发高性能可扩展的计算机系统,主持国家自然科学基金委优秀青年科学基金,科技创新2030“新一代人工智能”重大项目课题,入选上海市青年科技启明星计划;担任A类国际期刊IEEE Transactions on Computers的副主编和领域首席编委,B类国际会议IEEE SRDS的第39届程序委员会共同主席。中国计算机学会体系结构专委会副主任,中国计算机学会杰出演讲者,中国计算机学会青年计算机科技论坛2019-2020届上海主席;发表论文100余篇,获最佳/优秀/焦点论文奖6次;授权国内外发明专利20余项。上海交通大学教学能手,入选全国高校计算机专业优秀教师奖励计划,主编/参编教材2部。本科就读于浙江大学竺可桢学院,博士毕业于美国佛罗里达大学,先后获得佛罗里达大学工学院“盖特工程师特质奖”、Intel青年学者提升计划、中国计算机学会青年人才托举计划、微软亚洲研究院铸星计划、IEEE TCSC可扩展计算技术委员会早期职业成就奖、上海交通大学孙永强青年学者奖等荣誉。



侯小凤

上海交通大学

侯小凤,上海交通大学助理研究员,主要面向新一代云边融合智能应用研发高性能高能效的智算中心。博士毕业于上海交通大学,曾赴美国加州大学河滨分校(UCR)访问学习,并担任香港科技大学智能晶片研究与制造中心(Access)博士后研究员。作为骨干先后参与国家自然科学基金、国家重点研发等项目,以及华为、阿里等技术合作项目。先后发表包括ISCA、SC、AAAI、IPDPS、ICPP、ICCD、Euro-Par、TC等在内的十余篇国际会议/期刊论文,2018年获得ICCD最佳论文奖,2023年入选Euro-Par最佳论文奖候选,作为主要发明人已授权4项国家发明专利和1项美国发明专利。2015年度大连市优秀毕业生,2020年度上海交通大学优秀博士毕业生。曾入选CCF体系结构专委会和CCF分布式计算与系统专委会的优秀博士生论坛担任特邀讲者。

讲者报告详细内容

论坛讲者:



钱学海
Purdue University

主旨报告

钱学海教授主要从事计算机体系结构研究。博士毕业于伊利诺伊大学厄巴纳 - 香槟分校, 并于加州大学伯克利分校担任博士后研究员, 曾任美国南加州大学助理教授。钱学海教授的研究获得得美国自然科学基金职业成就奖(CAREER Award)等多个项目支持, 总资助超过 220 万美元; 是美国首届华人计算机新星奖的获得者; 在体系结构国际顶级会议和期刊上发表多篇高水平论文。进入所有四个计算机体系结构顶级会议“名人堂”(Hall of Fame): 在 ASPLOS 发表 18 篇、ISCA 发表 9 篇、MICRO 发表 8 篇、HPCA 发表 12 篇论文。

报告题目: Domain-Specific System and Architecture

报告摘要:

主要讨论如何提升通用多核处理器的易编程性、可调试性, 以及优化领域定制加速体系结构的设计, 报告重点介绍如何有效地将图计算和人工智能算法、软件系统和硬件设备有机结合起来开展研究。报告也将分享个人研究经历和心得体会。



王文浩
中国科学院
信工所

论坛讲者2

王文浩, 中国科学院信息工程研究所副研究员, 硕士生导师, 长期从事芯片安全计算架构研究, 尤其是可信执行环境(TEE)安全性分析以及自主可控可信执行环境架构设计等研究。曾发现 SGX 多个侧信道泄露途径和泄露方式, 从软件和硬件角度提出了 TEE 侧信道防御的方案, 设计了支持异构硬件和跨平台的 TEE 架构, 研究成果发表于 CCS、S&P、EUROCRYPT、USENIX ATC、IEEE TDSC、IEEE TIFS、IEEE TC、ACSAC、CHES、DSN 等, 主持国家自然科学基金青年项目、面上项目, 担任国家自然科学基金重点项目课题负责人、国家重点研发项目课题骨干, 承担中国计算机学会 - 华为胡杨林基金可信计算专项、蚂蚁集团校企合作项目等课题。担任 Inscrypt 2022 大会主席, 以及 CCS、ACNS、ICICS 等会议程序委员会委员, 获得 ACM 中国新星奖提名奖和 ACM 中国 SIGSAC 分会新星奖。

报告题目: 芯片安全计算架构

报告摘要:

系统安全的研究目标之一是通过软硬件协作的隔离机制(例如进程隔离、虚拟机隔离等)实现安全计算环境。隔离机制影响安全计算环境的可信计算基和攻击面, 决定了与其它组件的交互和协作方式, 进而影响安全计算环境的计算和通信效率。本报告总结安全计算环境的资源共享导致的隔离不充分的问题, 并以可信执行环境这一新型安全计算环境为目标, 介绍报告人在侧信道攻击和防御以及新型可信执行环境架构方面的工作进展。



郝一帆
中国科学院计算
技术研究所

论坛讲者3

郝一帆, 中国科学院计算技术研究所工程师。主要研究方向是智能芯片架构与系统。研究成果曾发表于 MICRO、ICLR、SCIS 等国际会议和期刊。获 MICRO 2022 最佳论文提名。主持国家部委项目 1 项, 参与多款国产商业化智能芯片的研发。

报告题目: 一种加速高精度计算的专用计算架构设计

报告摘要:

本报告主要介绍一种加速高精度计算的专用计算架构设计, 该架构通过串行比特流处理实现了高效、灵活的长位宽标量计算。基于此, 本报告将讨论以加速低位宽张量为主的深度学习处理器对高精度计算的适配, 期待能为“HPC+AI”算力融合趋势下的硬件设计提供新思路。



姚鹏程
华中科技大学&
之江实验室

论坛讲者4

姚鹏程, 博士毕业于华中科技大学, 师从金海教授, 现为之江实验室图计算中心博士后。姚鹏程长期致力于图计算和体系结构领域的研究, 在 MICRO、HPCA、FPGA、PACT、TC 等知名国际会议和期刊上发表论文 18 篇, 并曾获“互联网+”大学生创新创业大赛全国金奖, 图计算挑战赛 Graphchallenge 冠军, 图计算超算榜单 GreenGraph500 第四名等奖项。

报告题目: 面向图计算的领域通用加速器

报告摘要:

图作为一种灵活表达对象之间复杂关系的数据结构, 广泛地应用于社交网络、知识图谱和生物制药等重要现实场景。随着大数据时代的到来, 图数据的规模呈爆炸式增长, 现实应用对计算能力的需求急剧上升。然而, 由于图数据的不规则性, 通用处理器通常难以高效地执行图计算应用。更为严峻的是, 由于制造工艺逼近物理极限, 登纳德缩放定律和摩尔定律正在逐渐失效。近年来, 随着性能提升速度放缓, 通用处理器的性能逐渐无法满足图计算应用的需求。为此, 我们围绕图计算专用加速器开展了一系列研究, 并通过采用定制化的硬件架构取得通用处理器数十倍的性能。



郭聪
上海交通大学

论坛讲者5

郭聪, 上海交通大学计算机系博士生。他目前的研究方向是轻量化人工智能的软硬件协同优化加速, 主要针对神经网络的稀疏化和量化设计新型体系结构和系统。近三年, 郭聪共发表 10 篇论文, 第一作者文章 6 篇, 其中包括 ISCA, Micro, SC 等 CCF-A 类会议论文。他的工作受到多个硬件厂商的关注, 获得了 2022 年华为火花奖。在博士期间, 郭聪获得了包括博士研究生国家奖学金在内的多项奖学金, 获得了 2022 年 IEEE Micro 年度最佳论文(Tops Picks)提名奖。2023 年, 郭聪荣获上海交通大学优秀毕业生荣誉称号。

报告题目: 大语言模型量化在体系结构中的挑战和机遇

报告摘要:

LLM(大语言模型)随着模型规模的增长取得了巨大成功, 但越来越大的模型使得模型推理成本越来越高, 且远远超过了计算机硬件的发展速度。量化是加速并压缩大模型最有效的方法之一, 然而目前 LLM 中存在的新特性使得现有量化方法的效果变差。

在此次报告中, 首先对现有大规模语言模型量化中存在的挑战进行分析和介绍。针对这些挑战, 报告提出了一种硬件友好性的量化方法。首先, 我们对 LLM 中张量内和张量间不同的分布, 提出了一种分布感知的自适应数值类型。其次, 我们面向 LLM 中存在的 Outlier, 提出了 Outlier 感知的量化架构。我们基于这两种架构提出了一种硬件友好的量化方法以及高效的硬件设计与实现, 可以对 LLM 进行无需训练的快速量化(PTQ), 在加速 LLM 的推理速度、减少运算开销的同时, 保证 LLM 的准确率与性能表现。



潘锐
湖南大学

论坛讲者6

潘锐, 湖南大学 2020 级博士生, 导师: 李克勤老师、刘楚波老师。研究方向是区块链和计算机体系结构。目前正在研究区块链中的事务执行, 探索如何通过架构设计来加速智能合约的执行。已有一篇相关论文被 ISCA-2023 录用, 被审稿人评价为: “it is one of the first that explores hardware acceleration in this blockchain contract acceleration”。在校期间, 曾获得国家奖学金, 优秀毕业生称号, 数学建模国家一等奖。

报告题目: 通过算法和架构的协同设计加速智能合约执行

报告摘要:

相比于集中式的数据库系统, 吞吐受限的区块链系统无法被广泛应用。虽然高效的共识算法能够缓解这一问题, 但事务执行始终处于区块链吞吐的关键路径上, 其低效的执行方式成为了吞吐提升的新瓶颈。现有的方法主要从软件角度发掘事务之间粗粒度的并行性, 实际的加速效益受限于事务对智能合约的调用执行。本次报告揭示了加速事务执行的新契机, 充分挖掘事务在时空分布上的特点, 进行了算法和架构的协同设计。

分论坛日程

论坛2:概率型数据结构设计及其应用

时间:2023年8月4日13:30 - 17:30

论坛简介:

网络大数据、物联网大数据、社会大数据和视频大数据的集中全面爆发使数据价值密度日益稀疏,而摩尔定律的逐渐失效则令算力的持续高速增长潜力达到极限,依赖于算力纵向提升的数据挖掘策略面临着前所未有的性能瓶颈。为缓解大数据4V特性为数据的计算、存储和传输带来的巨大挑战,多种数据概率型数据结构被广泛用于提取海量数据中的关键信息并支撑上层数据挖掘业务。本论坛围绕“概率型数据结构设计及其应用”这一主题,邀请业界顶级研究人员进行专题报告并开展讨论,促进学术交流、观点交锋和思想碰撞,挖掘领域未来的研究方向。

具体日程安排(论坛2)

时间	报告内容	报告嘉宾/单位
13:40 - 14:20	稀疏网络测量:一种低成本的全网测量方案	谢鲲/湖南大学
14:20 - 15:00	意图驱动网络测量	李振宇/中国科学院计算技术研究所
15:00 - 15:20	茶歇	
15:20 - 16:00	新型网络数据挖掘研究	戴海鹏/南京大学
16:00 - 16:40	近似零误差的网络测量技术	黄群/北京大学
16:40 - 17:20	轻量级高性能带内全网遥测	李福亮/东北大学
17:20 - 17:50	Panel 讨论:概率型数据结构的发展与应用	

论坛主席:



杨全
北京大学

杨全,清华计算机系毕业,现为北京大学计算机系网络所研究员,博士生导师,教育部青年长江学者。主要研究网络大数据算法与系统。近年发表一作/通信作者CCF A类论文60多篇,领域排名第一SIG论文17篇。负责1项国家自然科学基金重点支持项目、3项国家重点研发计划课题/子课题、国家自然科学基金、青年基金、华为、中心、今日头条等项目。研究成果被国家自然科学基金委网站报道2次,教育部科技发展中心1次,研究成果被基金委报道2次,教育部科技发展中心1次,多项研究成果在华为海思、华为产品线、今日头条、Redis数据库中得到了应用部署。



罗来龙
国防科技大学

罗来龙,博士,国防科技大学副研究员,硕士生导师,中国计算机学会高级会员,《计算机学会通讯》期刊编委。长期围绕大规模网络系统管理基础理论开展研究,在相关领域发表论文50余篇,其中中国计算机学会(CCF)推荐的A类期刊/会议14篇,出版专著1部,获批美国发明专利6项、中国发明专利22项。获得2021年中国电子学会自然科学二等奖(排名第二),获得湖南省优秀博士学位论文。主持湖南省优秀青年基金、国家自然科学基金等项目共8项,参与重点研发计划、基础加强计划、联合基金重点项目等军队和国家级项目10余项。

讲者报告详细内容

论坛讲者：



谢鲲
湖南大学

论坛讲者1

谢鲲为湖南大学二级教授，国家杰青，湖南省杰青，湖南省青年骨干教师，湖南省优秀硕士生导师，长沙市“巾帼建功”标兵，湖南大学岳麓学者，湖南大学科研标兵。目前担任超算与人工智能融合计算教育部重点实验室主任。谢鲲教授研究方向为：计算机网络，网络安全，大数据和人工智能。先后主持国家自然科学基金4项，华为技术开发项目4项等。授权发明专利30余项。在国际主流会议和期刊发表论文170余篇，其中CCF A类期刊和会议论文40余篇。坚持科研一线工作，第一作者发表CCF A类期刊和会议论文24篇。成果应用于多个实际平台。两次获得湖南省科学技术进步奖二等奖，一次获得吴文俊人工智能科技进步一等奖。

报告题目：稀疏网络测量：一种低成本的全网测量方案

报告摘要：

全网测量对于许多网络功能非常重要，包括网络安全、网络预测、网络规划、网络服务SLA跟踪等。然而，网络测量存在一个困境：如何在获取细粒度、准确的全网性能数据的同时，减少测量开销以及测量对网络带来的影响？受信号处理中稀疏表征技术的启发，最近出现稀疏网络测量的方案。该方案通过精心选择一些节点对进行测量，并通过稀疏重构算法推断未测量的数据。稀疏网络测量可以获得完整的全网性能数据并大大减少测量代价。本次报告将阐述稀疏网络测量的基本思想、挑战问题和我们最近的进展。



戴海鹏
南京大学

论坛讲者2

戴海鹏，南京大学计算机科学与技术系副教授，博导，国家级青年人才计划入选者。获ACM中国新星奖、IEEE可扩展计算技术委员会职业中期卓越研究成就奖、中国电子学会优秀科技工作者等荣誉。研究方向为物联网、数据挖掘、移动计算等。发表国际著名会议期刊论文200余篇，含CCF A类80余篇，包括SIGMOD、VLDB、ICDE、WWW、ATC、INFOCOM等国际一流会议。曾获CCF A类会议INFOCOM最佳论文提名奖；CCF B类会议ICNP最佳论文奖及SECON最佳论文奖亚军。主持和承担国自科面上、联合基金重点、国家重点研发等项目十余项。担任ACM SIGCOMM China秘书长等职务。担任ISPA、HPCC等十余次会议主席职务。担任中文CCF A类期刊《电子学报》青年编委等职务。

报告题目：新型网络数据挖掘研究

报告摘要：

随着移动互联网、物联网等技术不断发展，全球网络数据量呈爆炸性增长趋势，数据形态亦由传统静态数据向海量动态数据转变，呈现的模式更为复杂，获取的信息更为多样。网络数据挖掘目前已成为学术界研究热点，其关键在于设计概率型数据结构对数据进行抽象存储，并设计场景适配的查询及计算方法。研究的挑战是如何协调海量数据和有限存储空间之间的矛盾，即如何对数据进行压缩、抵御数据处理过程中噪声、及保证处理算法的理论性能等。本次报告主要介绍作者近几年在持续项元素挖掘、负载感知的数据结构设计等方面研究，重点介绍在新型数据模式和信息场景下，如何应对上述挑战。



黄群
北京大学

论坛讲者3

黄群，北京大学博士生导师，助理教授，主要研究方向为分布式流处理和网络测量，在网络与系统领域顶级会议(包括SIGCOMM、INFOCOM、VLDB、USENIX ATC)发表论文多篇，主持国家重点研发计划子课题、国家自然科学基金等项目。

报告题目：近似零误差的网络测量技术

报告摘要：

网络测量技术在网络管理中的重要性日益凸显。然而，随着网络用户数量的增加、网络流量的增长、以及网络自身架构的日趋复杂，网络测量仍然面临着测量精度与资源开销的矛盾。本报告将从系统与算法两个角度，介绍近年来近似零误差网络测量技术的研究进展。



李振宇
中国科学院计算
技术研究所

论坛讲者4

李振宇，中国科学院计算技术研究所研究员，博士生导师，中国科学院大学岗位教授。主要研究方向为网络体系结构与网络测量，重点研发计划项目首席科学家，主持NSFC区域联合重点、863等项目。相关研究成果发表在SIGCOMM、ATC、INFOCOM等领域著名学术刊物和会议上，部分成果已成功应用于多个互联网系统，服务上亿用户。担任INFOCOM (2019-2024)、ICNP (2019-2023)等国际会议的TPC成员、《计算机研究与发展》青年编委。2017年度入选中国科学院青促会优秀会员，2020年获得北京市杰出青年基金项目资助。

报告题目：意图驱动网络测量

报告摘要：

传统被动网络测量捕获流经的网络流量，通过离线方式分析网络性能、访问行为，发现网络攻击等。由于采用离线分析方式，网络攻击发现等存在滞后问题。此外，随着网络带宽的增长，这种“自底向上”的测量方式面临严重的性能挑战。事实上，网络管理人员通常仅关注网络的某一部分流量是否存在网络攻击行为。为了灵活实现网络管理人员的测量意图，并完成尽可能多的实时攻击检测等测量任务，我们设计了意图驱动的网络测量系统，测量意图通过原语的组合实现，采用“自顶向下”的测量方式。我们在可编程交换机上实现了测量数据平面，采用流摘要等技术实现在网攻击检测等，并设计了测量意图到 P4 语言的转译方法。本报告将介绍意图驱动的网络测量关键技术与原型系统。



李福亮
东北大学

论坛讲者5

李福亮，博士毕业于清华大学计算机系，现任东北大学大学计算机学院副教授、博士生导师。主要研究方向为未来网络技术、网络遥测、网络智能运维等；主持国家级项目 6 项，主持华为、字节跳动、阿里巴巴合作项目 6 项；第一及通讯作者发表学术论文 30 余篇，申请发明专利 10 余项，出版学术专著 6 部；担任 CCF 互联网专委会执行委员、华为中央研究院专家顾问；入选辽宁省百千万人才工程、沈阳市高层次人才计划；获沈阳市中青年科技创新人才、辽宁省通信学会优秀科技工作者等荣誉称号。

报告题目：轻量级高性能带内全网遥测

报告摘要：

带内网络遥测借助数据平面可编程性对网络状况进行收集、携带、整理、上报，是一种高效的混合式测量技术，能够满足细粒度、实时性的网络监测需求，得到学术界和工业界的广泛关注。报告分享带内网络遥测方面的一些研究工作，重点讨论如何通过多条或单条遥测路径实现轻量级、高性能的全网遥测，如何解决带内全网遥测过程中的 MTU 约束限制问题，以及探索如何实现分布式的大规模带内全网遥测系统的部署。

分论坛日程

论坛3:智能边缘计算体系结构与软硬件系统

时间:2023年8月4日13:30 - 17:30

论坛简介:

作为边缘计算和人工智能深度融合的新兴计算范式,智能边缘计算被认为是赋能 5G/6G 通信、推动智能万物互联、打通人工智能落地最后一公里的关键使能技术,同时也是国民经济、现代社会以及国家安全的重要基础设施。鉴于其重要的社会经济和学术研究价值,目前各国均对边缘智能展开政策规划和科研布局。多地发布的“新基建”行动方案均提出要推动边缘计算基础设施建设,来支撑各类新兴智能应用;此外,国家重点研发计划“先进计算与新兴软件”和“宽带通信和新型网络”等专项,智能边缘计算作为重点方向被列入。本论坛将从智能物联网操作系统、实时视频分析系统、边缘存储、边缘加速芯片、边缘智能、云边协同机器学习和大规模推理服务系统等不同角度来探讨智能边缘计算体系结构与软硬件系统方面的前沿进展。

具体日程安排(论坛3)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	智能物联网:前沿探索与系统创新实践	郭斌 (西北工业大学)
14:00 - 14:30	云边端广域协同视频感知与计算	舒元超 (浙江大学)
14:30 - 15:00	边缘存储技术与系统的初步探索	陆游游 (清华大学)
15:00 - 15:20	茶歇	
15:20 - 15:50	边缘赋能的人工智能:新机遇与新挑战	何强 (华中科技大学)
15:50 - 16:20	服务质量感知的高吞吐云边协同推理系统	陈全 (上海交通大学)
16:20 - 16:50	大语言模型时代下的边缘智能系统	徐梦炜 (北京邮电大学)
16:50 - 17:30	Panel: 大模型时代边缘智能的机遇与挑战?	

论坛主席:



周知
中山大学

周知,中山大学计算机学院副教授、博士生导师,广东省杰出青年科学基金获得者,近年来的主要研究方向包括云计算与边缘计算、大数据处理和机器学习系统等,在 PIIEEE、ToN、JSAC、TC、TPDS、TMC、INFOCOM、RTSS、MobiHoc、ICDCS 等国际学术期刊和会议发表论文 80 余篇,论文总引用次数逾 5000 次,另有 6 篇论文入选 ESI 高被引论文,3 篇论文入选 ESI 热点论文。曾获得 2021 年华为云与计算先锋教师、2019 年中国计算机学会优秀博士学位论文奖提名、2018 年 ACM 武汉优秀博士学位论文奖(唯一获奖者)、2018 年湖北省计算机学会优秀博士学位论文奖(唯一获奖者)、IEEE IoTJ2022 年度最佳论文亚军奖、IEEE UIC2018 国际学术会议最佳论文奖、IEEE WCNC2023 最佳论文提名等奖励。



吴松
华中科技大学

吴松,华中科技大学二级教授,博士生导师,现任计算机学院副院长,“服务计算技术与系统教育部重点实验室”和“集群与网格计算湖北省重点实验室”副主任,入选国家科技创新领军人才、教育部新世纪优秀人才。在 IEEE/ACM Transactions 和 HPDC、ICDCS、SC、MSST、SoCC、IPDPS 等国内外重要学术期刊和中国计算机学会推荐的重要国际学术会议上发表论文 200 余篇;获得国家自然科学二等奖 1 项、国家科技进步二等奖 1 项、教育部自然科学一等奖和科技进步一等奖各 1 项、湖北省科技进步一等奖 1 项。主持云计算与虚拟化领域的国家自然科学基金重点项目 1 项、面上项目 4 项、863 重大专项课题 1 项、国家重点研发计划课题 1 项,作为核心成员参加本领域 973 项目、863 重点项目、国家自然科学基金重点项目等 10 余项。获得国内外发明专利授权 50 余项(含美国专利 6 项)、软件著作权 10 余项。研究成果应用于包括腾讯、华为、阿里、微软、曙光、浪潮等公司的大规模生产性云计算系统和软件产品、以及全国 40 多所高校的数据中心等。担任国内外学术会议的程序委员会主席 9 次、程序委员 30 余次。中国计算机学会杰出会员、高性能计算专委会委员和体系结构专委会委员、中国电子学会青年科学家俱乐部成员、云计算与大数据专委会委员、ACM 武汉分部理事、湖北省青年科技协会理事等。

讲者报告详细内容

论坛讲者:



郭斌
西北工业大学

论坛讲者1

郭斌, 工学博士, 西北工业大学计算机学院教授、副院长, 国家杰出青年科学基金获得者, 人机物融合群智计算教育部重点实验室副主任, 工信部智能感知与计算重点实验室副主任, 西北工业大学计算与艺术交叉研究中心主任。2009 年在日本庆应大学获博士学位, 2009-2011 年在法国国立电信学院进行博士后研究。入选教育部“新世纪优秀人才”(2012)和国家“万人计划”青年拔尖人才(2017), 爱思唯尔中国高被引学者。主要从事智能物联与普适计算、群体智能、人机物融合群智计算等方面研究。在 IEEE/ACM 汇刊等国内外重要期刊和会议上发表论文 150 余篇, 曾获得教育部自然科学一等奖、陕西省自然科学一等奖以及 IEEE UIC' 17、ISI' 19、BIBM' 20 等国际会议“最佳论文奖”。担任《IEEE Transactions on Human-Machine Systems》、《IEEE Communications Magazine》、《ACM IMWUT》等国际权威期刊编委。IEEE 高级会员, 中国计算机学会 CCF 杰出会员、CCF 西安分部主席。

报告题目: 智能物联网: 前沿探索与系统创新实践

报告摘要:

人类正在进入“人机物”融合万物智能互联的时代。随着物联网和人工智能技术的飞速发展, 智能物联网(AIoT)正成长为一个极具前景的新兴领域, 它以人工智能、边缘计算与物联网等技术的深度融合为基础, 在感知、通信、计算和应用中通过 AI 技术赋能, 构建具有更高灵活性、自组织性、自适应性、持续演化的物联网系统。本次报告将分享西工大在智能物联网概念体系、前沿探索与系统创新方面的最新进展。



舒元超
浙江大学

论坛讲者2

舒元超, 浙江大学控制科学与工程学院教授, 博士生导师。2015 年博士毕业于浙江大学控制科学与工程学院, 导师孙优贤院士, 陈积明教授。2013-2015 作为国家公派联合培养博士研究生在密歇根大学安娜堡分校计算机系访问, 导师 Kang G. Shin。2015-2022 年在微软研究院及云计算部门担任研究员、高级研究员, 首席研究员。2022 年 11 月回国加入浙江大学任求是特聘教授。研究领域主要包括移动及无线系统、网络控制与优化、边缘计算。其研究成果在计算机、通信、控制等领域的国际顶级期刊及会议上发表论文 50 余篇, 授权受理中美专利 20 余项。曾获 ACM 中国优博论文奖, IEEE WCNC, ACM SenSys, ACM/IEEE SEC 最佳论文奖, IEEE INFOCOM, ACM MobiSys 最佳系统奖等。现担任 IEEE Transactions on Wireless Communications 编委、ACM Transactions on Sensor Networks 副主编, ACM MobiCom, MobiSys, SenSys, IPSN 等国际顶级会议程序委员会委员。

报告题目: 云边端广域协同视频感知与计算

报告摘要:

伴随着算力基础设施的飞速进步, 人工智能正在进入一个黄金时代。受益于机器学习算法的突破、海量的训练数据以及广泛部署的视频采集设备, 视频分析应用正从众多的 AI 应用中脱颖而出, 在科学和社会发展中展示出巨大的潜能。作为 AI 系统领域最具挑战的课题之一, 实时视频分析具有数据量大、计算密集、语义丰富等特征, 以及低延迟和高隐私保护等需求。本次报告将系统性介绍边缘视频分析系统的最新进展, 包括边缘侧实时推理加速、跨摄像头网络的协同视频分析以及边缘模型自演进学习等方面。



陆游游
清华大学

论坛讲者3

陆游游, 清华大学计算机系副教授、博士生导师。主要研究方向是计算机存储系统, 在 FAST、OSDI、SOSP 等国际顶级会议期刊上发表论文 30 余篇, 曾获得最佳论文奖和最佳论文提名奖各一项。研制高性能文件系统 SuperFS, 部署于鹏城云脑 II, 获得 2023 年世界超算存储 IO500 榜单第一名。担任 FAST、USENIX ATC、EuroSys 等国际会议程序委员会委员。曾获中国计算机学会优博奖、ACM 中国操作系统分会新星奖、省部级奖项两项等, 入选首届中国科协青年人才托举工程、国家自然科学基金优青和国家重点研发计划青年科学家项目资助。

报告题目: 云-边协同的智能感存算融合系统与技术

报告摘要

随着5G、物联网等技术的商用进程不断加快,面向万物互联的智能终端已经无处不在,据统计,2020年全球物联网设备总量已达240亿。在如此庞大的设备终端构建的复杂网络中,由云-边-端构成的智能应用在数据处理时,由于任务分工的固有限制以及传统体系架构存在的“内存墙”、“存储墙”等问题,正面临计算效率、通信延迟、扩展性、实时性、能耗等各方面的严峻挑战。为此,相关研究提出一种让“边”更强更近的近数据感存算融合技术,从而提供更强大的计算、更快/大的存储、更低的能耗以及更短的通信距离。报告将首先综述国内外围绕存算融合技术展开的相关工作;然后介绍本团队围绕云边协同的智能感存算融合技术开展的系列研究,包括1)存算融合体系架构的数据流与计算流重构,2)基于感存算融合的认知与决策,3)云-边协同的智能计算与数据共享等;最后介绍近数据智能感存算融合系统在多场景下的应用。



何强
华中科技大学

论坛讲者4

何强,华中科技大学计算机学院教授,博导,国家级青年人才,获得华中科技大学-澳大利亚斯威本科技大学双博士学位,曾在澳大利亚斯威本科技大学任讲师、高级讲师和副教授,主持和参与国家级科研项目7项,在CCF A类会议和期刊上发表高水平学术论文70余篇,在ACM/IEEE Transactions上发表论文80余篇,5次获国际会议最佳论文奖和最佳学生论文奖,3次获得校级杰出科研奖,培养的博士生3次获得优秀博士生奖,主要研究方向包括边缘计算、云计算和服务计算。

报告题目: 边缘赋能的人工智能:新机遇与新挑战

报告摘要:

作为一种新的计算架构,边缘计算可以在靠近数据源的地方进行数据的分析和处理,在一些对反应速度要求比较高的应用场景中帮助用户和终端设备进行快速决策,如自动驾驶和虚拟/增强现实等,为人工智能在5G/6G网络边缘端的应用带来了全新的机会和挑战。本次报告将分享何强团队在边缘赋能的人工智能这一新兴领域中的最新研究进展。



陈全
上海交通大学

论坛讲者5

陈全,博士,上海交通大学计算机系教授。长期从事计算机系统、计算机体系结构的研究,目前在相关领域发表学术论文一百余篇。研究获国家优青、国家自然科学基金重点项目资助。曾获中国计算机学会优秀博士学位论文奖、IEEE TCSC青年学者奖、阿里青橙奖。担任SCI期刊Parallel Computing、JCST、FCS编委及青年编委。研究成果曾获国家技术发明二等奖、省部级科学技术一等奖多项。

报告题目: 服务质量感知的高吞吐云边协同推理系统

报告摘要:

基于深度神经网络的应用现在被广泛使用于互联网服务当中。将整个或部分模型卸载到用户端移动设备的推理模式可以消除长数据传输时间。然而,由于异构的边缘设备具备不同的计算性能和网络条件,不同的DNN层应该在不同的设备上运行,使得数据中心上的请求显示出不规则的结构。如何根据不规律负载的到来顺序和任务量确定请求处理的优先级是亟需解决的问题。该报告将介绍一个支持不规则请求的协作推理服务系统ICE,包括切片器、请求管理器和延迟增强器等组件。实验结果表明,ICE可以显著提高数据中心支持的峰值负载,同时保证了所需的99%延迟。



徐梦炜
北京邮电大学

论坛讲者6

徐梦炜,北京邮电大学计算机学院特聘副研究员(人才引进),博士生导师。于北京大学获得本科与博士学位,入选中国科协青年人才托举工程,北京市科技新星,微软亚洲研究院“铸星计划”访问学者,普渡大学访问学者,ACM SIGMobile China 2021优博。主要研究领域为移动/边缘计算和系统软件,近些年专注于边缘场景下的机器学习系统构建与优化,相关成果发表于ACM MobiCom/MobiSys/UbiComp/USENIX ATC/IEEE TMC/软件学报等国内外顶级会议期刊,主持国家自然科学基金、科技部重点研发项目课题、百度松果基金等多个项目。

报告题目: 大语言模型时代下的边缘智能系统

报告摘要:

边缘智能技术是指将机器学习算法和模型部署到距离用户和数据源更近的边缘/端侧(智能手机、IoT、可穿戴设备、边缘云等),以实现更好的用户隐私保护和低延时推理。但是,以神经网络模型(DNN)为代表的机器学习算法普遍存在参数量大、计算量大等特点,难以在资源受限的边缘侧高效部署。为此,相关领域的学者从硬件设计、系统构建、算法模型等多个维度展开了深入的探索。在本次报告中,徐梦炜博士将重点介绍团队在边缘智能系统软件方向的相关成果和思考。同时,近些年在机器学习领域取得重大进展的大语言模型(large language models)为边缘智能提出了新的要求和挑战,传统的机器学习系统范式正在被颠覆,徐梦炜博士也将就这一转变讨论他的观点。

分论坛日程

论坛4:算力网技术论坛

时间:2023年8月4日13:30 - 17:30

论坛简介:

算力网已经成为近年来信息领域产学研各界关注的热门方向,中国科学院计算技术研究所、鹏城实验室、中国移动等研究机构和企业相继提出了“信息高铁”算力网、算力网络(CPN)、中国算力网(C2NET)等计划和构想。科技部也在今年4月启动了国家超算互联网部署工作。本论坛将邀请产学研各界研讨算力网的系统结构、关键技术和建设方法等内容,介绍算力网的关键技术研究和产业化方面的最新动态。

具体日程安排(论坛4)

时间	报告题目	讲者/单位
13:30 - 14:00	算力经济的发展趋势分析与展望	张云泉/中国科学院计算技术研究所
14:00 - 14:30	无线算力网络	张彦/挪威奥斯陆大学
14:30 - 15:00	基于在网计算的分布式系统加速方法	吴文斐/北京大学
15:00 - 15:20	茶歇	
15:20 - 15:50	智能算力云际互联的探索与实践	余跃/国防科大
15:50 - 16:20	全国一体化算力服务网络探索与实践	王浩/曙光智算
16:20 - 16:50	面向算网一体的未来网络技术思考与实践	刘鹏/移动研究院

论坛主席:



郭得科
国防科技大学

国防科技大学教授,博士生导师。主要研究方向为网络计算与系统、分布式计算与系统、网络空间安全等。获2020年CCF-IEEE CS青年科学家奖,以第一完成人获2019年度湖南省自然科学一等奖1项、2021年度中国电子学会自然科学二等奖1项,参与中国计算机学会自然科学二等奖1项、军队科技进步二等和三等奖各1项;发表CCF A/B类论文110余篇,获IEEE ICNP 2019最佳论文;以第一著者出版中英文学术专著3部;以第一发明人授权美国发明专利11项、中国发明专利60余项。入选国家级领军人才、国家级青年人才、军队学科拔尖人才、教育部新世纪优秀人才等,主持第3批NSFC优秀青年基金项目。



彭晓晖
中国科学院计算技术
研究所

彭晓晖，中国科学院计算技术研究所副研究员、博士生导师。分别于西北工业大学、北京大学、东京大学获得本科、硕士和博士学位，入选中国科学院海外人才引进计划。主要从事新一代分布式计算系统方面的研究，目前正在开展面向万物互联的算力网分布式系统架构和操作系统的研究与开发工作。主持了国家自然科学基金面上项目、重点项目子课题等多个国家纵向科研项目。在《计算机研究与发展》、《计算机学报》、PIEEE, IEEE IoT-J 等国内外著名刊物上发表相关论文 20 余篇。兼任 IEEE Internet Computing 编委会委员，边缘计算国际旗舰会议 ACM/IEEE SEC 2019-2023 年度 TPC 成员，CCF 普适计算、分布式计算与系统专委会委员，中国通信学会第一届算力网络委员会委员。

讲者报告详细内容

论坛讲者：



张云泉
中国科学院计算
技术研究所

论坛讲者1

张云泉，博士，中国科学院计算技术研究所研究员，博士生导师，并行软件实验室主任，全国政协委员，中国计算机学会会士，高专委秘书长，中国工业与应用数学学会副秘书长，主要研究方向为并行程序设计和性能评价、并行计算和并行编程模型等。已在国内外学术刊物上发表论文二百余篇，出版专著三部，译著七部。获2000年和2019年国家科技进步奖二等奖各一项，2017年首届CCF青竹奖获得者，2017年中国科学院科教成果一等奖，2017年中国科学院杰出科学与技术成就奖，2017年度国际艾奇奖，获2021年度ACM 戈登贝尔奖提名。中国智能计算产业联盟执行理事长，ACM中国高性能计算专家委员会主席，中国软件行业协会常务理事。中国高性能计算机TOP100排行榜创始人和发布者，国际人工智能算力排行榜AIPerf500发起人，PAC, CCF CPC和ACM IPCC超算大赛创始人。一百多次担任国际会议程序委员会委员和共同主席。中国科学院《数据与计算发展前沿》副主编。《中国图象图形学报》和《计算机科学与探索》编委。光明日报科普专家委员会委员，国家基因库高级顾问，青海省大数据与云计算咨询专家委员会委员，贵州省农业大数据专家委员会会委员，内蒙古环球智库大数据发展中心专家委员会委员，郑州市智慧城市专家委员会委员，吕梁市大数据专家咨询委员会委员和政府转型专家顾问，江西抚州市算力专家咨询委员会执行主任。九三学社中央科技专委会委员，中央科普工委委员，中央促进技术创新工作委员会委员，九三北京科技委副主任。

报告题目：算力经济的发展趋势分析与展望

报告摘要：

本报告根据超算世界TOP500和中国TOP100排行榜的数据，总结和分析了超级计算的若干发展趋势，特别是与人工智能的融合创新，指出算力经济时代已经到来，应该通过尽快建立新的算力产业商业模式和评价体系，指导算力经济健康发展。



张彦
挪威奥斯陆大学

论坛讲者2

张彦现任挪威奥斯陆大学信息工程学院教授，IEEE Fellow, IET Fellow。入选欧洲科学院院士，挪威皇家科学院院士，挪威工程院院士，2018-2022连续五年全球“高被引科学家”。任CCF杰出会员，CCF区块链专委会执行委员，2019年CCF杰出演讲者。近期主要研究方向为新一代无线通信网络和智能安全物联网。在相关领域研究成果被引用38000+次，H因子为100。现任多个IEEE Transactions/Magazine期刊编委，包括IEEE TWC, TII, TVT, TSUSC, TGCN, TNSE, NETWORK, IoT, VTM 及 Systems Journal。并任国内多个核心期刊编委，包括中国通信，计算机科学，物联网学报，网络与信息安全学报，及数字通信与网络。

报告题目：无线算力网络

报告摘要：

本报告主要介绍我们近期提出的新研究方向：无线算力网络(Wireless Computing Power Networks, WCPN)。首先介绍无线算力网络的概念和模型。然后定义并解决无线算力网络中新颖独特的科学研究问题，如算力资源的优化分配和动态共享，算力协作和集群机制。



吴文斐
北京大学

论坛讲者3

吴文斐, 2015年博士毕业于威斯康星大学麦迪逊分校, 现任北京大学计算机学院研究员和博士生导师。吴文斐博士长期从事分布式系统和计算机网络方向的研究, 共发表高水平论文51篇, 包括NSDI、ASPLOS、KDD、INFOCOM等CCF A类文章18篇。吴博士主持研发的基于在网计算的机器学习系统获得NSDI21最佳论文奖(中国首次), 主持研发的基于在网计算的分布式大数据分析系统获得ASPLOS23的杰出论文奖, 其他成果曾获得IPCCC19最佳论文提名、SoCC13最佳学生论文奖、SIGCOMM10最佳论文提名。

报告题目: 基于在网计算的分布式系统加速方法

报告摘要:

摩尔定律正在逐步失效, 但是各种信息处理场景中的算力需求正在指数增长, 设计新的计算集群体系结构实现算力的高效扩展成为产业界关注热点, 而在网计算是一种支持新体系结构的有效途径。在网计算利用交换机的处理流经的数据报文, 可以提供线速、低时延的数据处理, 且可以压缩数量, 从而提升计算和传输任务的整体效率。本讲座中, 我们将分享在网计算技术的现状和发展, 包括支持机器学习和大数据系统的在网计算协议的设计、在网计算协议与现网 RDMA 技术的集成方法、集群中在网计算多任务的运行时管理方法。相关工作发表在 NSDI21、NSDI23、ASPLOS23、INFOCOM23。



余跃
国防科技大学

论坛讲者4

余跃, 博士, 国防科技大学副研究员, AITISA联盟算力网络推进组组长, OpenI启智开源平台运营中心主任, CCF开源发展委员会常务委员, 主要从事开源软件、群体智能、云计算等相关领域的研究工作, 在CHI、CSCW、TSE、ICSE、FSE、ASE等软件领域国际重要会议和期刊发表论文50余篇。

报告题目: 智能算力天际互联的探索与实践

报告摘要:

以人工智能应用为主的智能计算基础设施是我国新基建的重点领域。中国算力网—智算网络板块通过对分布在不同地域的智能计算中心高速互联, 形成灵活调度、资源共享、统一服务的一体化智算基础设施, 正在改变现有算力的服务模式, 提升智算中心的整体运行效率和服务能力。本报告将介绍我们以鹏城云脑为核心枢纽节点、多个算力中心天际协同的探索与实践。



王浩
曙光智算

论坛讲者5

王浩, 曙光智算信息技术有限公司副总裁, 负责国内多个国家级超大规模计算中心运营管理、产业支持与生态合作工作, 运营总算力超过 1000P。主持合肥综合性国家科学中心重点项目—合肥先进计算中心“巢湖明月”的设计、建设、运营工作, 凭借丰富的产业及生态拓展的经验, 开创计算服务市场新业态、新模式。同时参与科技部国家重点研发计划等相关科研课题工作, 负责曙光智算算力网络服务平台架构设计及运营体系规划, 构建全国一体化算力网络, 积极推动计算服务产学研用一体化创新发展。演讲题目:《全国一体化算力服务网络探索与实践》。

报告题目: 无线算力网络

报告摘要:

《全国一体化算力服务网络探索与实践》, 曙光算力网络接入国内诸多先进计算中心、智能计算中心等, 拥有海量算力资源。智能调度平台基于算力网络, 对算力资源展开感知、测量、调度、运营, 实现算力供需双方匹配, 并协同数据、环境、应用提供一体化服务。通过在业务、技术和市场等方面的创新实践, 推动超算中心从算力中心时代走向算力网络时代, 使算力真正成为促进数字经济发展的核心驱动力。



刘鹏
中国移动研究院

论坛讲者6

刘鹏, 中国移动研究院基础网络技术研究所技术经理, 从事未来IP网络、算力网络/算网一体等架构设计和算力路由、确定性网络、工业互联网等前沿技术研究。在国际互联网工程任务组(IETF)推进算力路由工作组(CATS)的成立并担任工作组主席, 牵头大规模确定性网络需求立项以及多个技术方案的推进。担任中国通信标准化协会(CCSA) TC13 WG2副组长, 主导完成10余项工业互联网边缘计算、工业SDN、确定性网络等技术标准。同时也推进新型网络技术在产业的应用, 在工业互联网产业联盟(AII)网络组联执主席, 牵头多项测试床以及技术案例。

报告题目: 面向算网一体的未来网络技术思考与实践

报告摘要:

随着算力网络的发展, 网络和计算两大学科的交叉融合成为新的技术发展方向。算网一体是算力网络发展的目标阶段, 网络的计算将会深度融合, 从“网络算力化”和“算力网络化”两条路径融合至一体共生。报告面向算网一体的未来网络发展, 探讨算网度量、算网感知、算力路由、在网计算等算网一体关键技术。

分论坛日程

论坛5:区块链分布式账本技术论坛

时间:2023年8月4日13:30 - 17:30

论坛简介:

本专题论坛专注于区块链分布式账本的底层机制、区块链系统与分布式协议设计、区块链性能瓶颈突破、以及区块链原理等方面。目标是参会专家学者提供一个分享国际前沿、权威、最新颖的学术观点以及交流合作的平台, 共同探索区块链技术的最新进展和发展趋势。本技术论坛邀请6位海内外知名区块链专家学者进行主题演讲, 内容涵盖:去中心化存储网络、加密技术、隐私保护数据交互、区块链跨链技术、区块链系统优化、共识算法、可扩展性技术、区块链存储优化等等区块链关键技术与应用。欢迎各位专家学者参加本次论坛, 共同探讨区块链技术的发展趋势和实际应用价值。

具体日程安排(论坛5)

时间	报告内容	报告嘉宾/单位
13:30 - 14:15	去中心化存储网络	成秀珍/山东大学
14:15 - 15:00	同态加密:原理与应用	刘哲/之江实验室
15:00 - 15:20	茶歇	
15:20 - 15:55	Privacy-Preserving Data Exchange	肖斌/香港理工大学
15:55 - 16:30	面向多链融合的区块链跨链互操作与可扩展关键技术	伍前红/北京航空航天大学
16:30 - 17:00	区块链技术与系统	刘秀龙/天津大学
17:00 - 17:30	可扩展图式区块链存储系统的研究与实践	肖江/华中科技大学

论坛主席:



黄华威

中山大学
软件工程学院

黄华威, 中山大学副教授, 博导, IEEE Senior Member, CCF高级会员, CCF分布式与并行计算专委会与CCF区块链专委会执行委员。研究方向包括区块链底层技术、分布式系统与协议设计。负责多项科技部与广东省重点研发计划课题, 主持国家自然科学基金面上青年项目、CCF-华为胡杨林基金区块链专项等十余项科研项目。发表论文100余篇, 部分研究成果发表在顶级国际学术期刊与会议, 如 ToN, JSAC, TPDS, TDSC, TC, TMC, INFOCOM, ICDCS, SRDS等等。担任多个区块链期刊的副编辑; 曾组织多个区块链专刊、区块链国际研讨会、与CCF区块链技术论坛, 例如曾在CCF A类期刊IEEE JSAC举办区块链专刊。出版Springer英文区块链学术专著2本、中文书籍《从区块链到Web3》。



郑子彬

中山大学
软件工程学院

郑子彬, 中山大学软件工程学院副院长、IEEE Fellow、IET Fellow、全球高被引科学家、国家优秀青年科学基金获得者、国家数字家庭工程技术研究中心副主任、区块链与可信软件研究中心主任。发表论文 200 余篇, 论文谷歌学术引用超过 3 万次。主持国家重点研发计划项目、自然科学基金重点项目等多个项目; 获得教育部自然科学二等奖、吴文俊人工智能自然科学二等奖、ACM 中国新星提名奖、IEEE TCSVC Rising Star Award、CCF 服务计算专委会杰出青年奖、ACM SIGSOFT Distinguished Paper Award、ICWS 最佳学生论文奖等奖项; 担任 TOSEM、TVT、OJCS 等期刊的副编辑。

讲者报告详细内容

论坛讲者:



成秀珍

山东大学计算机
科学与技术学院

论坛讲者1

成秀珍教授现担任山东大学计算机科学与技术学院院长、数链融合教育部工程中心主任, 是 IEEE Fellow、CSEE Fellow、AAIA Fellow、国家高层次海外人才、国家重点研发计划首席科学家、山东省“泰山学者”海外特聘专家。成教授现主要研究方向为区块链理论与应用、安全与隐私保护、物联网与边缘计算等。曾担任(或正在担任)多家期刊编委及多个国际会议主席, 是中国计算机学会推荐国际学术会议 WASA 的创办者和国际期刊《High-Confidence Computing》的创刊主编。H 指数 64, Google Scholar 总引用超过 1.8 万。

报告题目: 去中心化存储网络

报告摘要:

Decentralized Storage Networks (DSNs) can gather storage resources from mutually untrusted providers and form worldwide decentralized file systems. Compared to traditional storage networks, DSNs are built on top of blockchains, which can incentivize service providers to contribute their unused storage spaces and ensure strong security. However, pioneering DSNs still face challenges in many aspects such as data privacy, proof of storage complexity, multi-version file management, and file indexing. In this talk, we will report two of our recent works, namely FileDAG and FileDES, which intend to overcome the above problems in designing DSNs. Specifically, FileDAG is built on a DAG-based blockchain to support file-level deduplication in storing multi-versioned files while relying on a two-layer DAG-based blockchain ledger to provide flexible and storage-saving file indexing; FileDES (Decentralized Encrypted Storage) extends FileDAG from multiple aspects, incorporating data security, system scalability, and proof succinctness respectively with file encryption, a lightweight proof of encrypted storage algorithm (PoES), and a succinct multi-version verification method. Open research challenges and future research directions will also be discussed.

**刘哲**
之江实验室**论坛讲者2**

刘哲，之江实验室副院长，浙江大学兼职教授/博导，国家重大人才计划A类项目专家。研究方向为网络空间安全，发表学术论文160余篇，其中10篇论文获得包括CCF-A类会议ACM SIGSOFT FSE' 2022, WASA' 2022等在内国际会议最佳论文奖。曾获教育部高校计算机专业优秀教师奖、中国密码学会密码创新奖一等奖/优秀青年奖、MIT TR China 35 under 35、阿里巴巴达摩院青橙奖和欧洲FNR基金会杰出博士毕业论文奖等多个国内外奖励。主持国家自然科学基金重点项目和科技部重点研发计划课题等20余项国家级项目/课题，受邀担任国家自然科学基金委人才类项目/重点项目会评专家。

报告题目： 同态加密:原理与应用**报告摘要：**

数字经济时代，数据凭借其成本低、规模大、流动性高、复用性强等特点成为一个非常重要资源，数据相关产业在飞速发展的同时也面临数据安全、数据确权、隐私保护等一系列问题。因此，使用安全的数据处理方法，在保护用户隐私的同时确保数据可用并创造价值，是值得思考也亟待解决的问题。而隐私计算技术正是一种实现数据处于加密状态或非透明状态下计算的技术，达到了各参与方隐私保护的的目的。本次报告将介绍同态加密技术，这是隐私计算主流技术之一。我们将讲解同态加密技术的基础知识，研究现状和应用实现，以及我们团队的研究方向和成果，旨在为同态加密技术的研究和应用提供新思路和方向，以促进其发展和推广。

具体而言，本次报告将分为五个部分。首先，我们将介绍隐私计算的背景和定义，对比现有的三类主流技术路线，分析使用同态加密技术做隐私保护的的优势。其次，我们将介绍同态加密的基础知识，包括基本原理、发展历程和运算操作。然后，我们将探讨同态加密的研究现状，包括同态加密算法库、在CPU、定制硬件和GPU方面的实现。同时，我们还将介绍同态加密在隐私保护下的数据分析和机器学习等领域的应用。随后，我们将总结同态加密的性能瓶颈和未来发展趋势，并介绍我们团队的工作，包括GPU加速实现优化、隐私保护决策树、安全矩阵乘法等方面。最后，我们将讨论同态加密面临的挑战和未来方向，包括性能瓶颈、标准化以及在新兴领域(如区块链和机器学习)中的潜在案例研究。

**肖斌**
香港理工大学**论坛讲者3**

Dr. Bin Xiao is a professor at Department of Computing, the Hong Kong Polytechnic University, Hong Kong. Prof. Xiao received the B.Sc and M.Sc degrees in Electronics Engineering from Fudan University, China, and Ph.D. degree in computer science from University of Texas at Dallas, USA. His research interests include AI and network security, data privacy, and blockchain systems. He published more than 200 technical papers in international top journals and conferences. Currently, he is the associate editor of IEEE IoTJ, IEEE TCC, and IEEE TNSE. He has been the associate editor of Elsevier JPDC from 2016 to 2021. He is the IEEE ComSoc Distinguished Lecturer and the vice chair of IEEE ComSoc CISTC committee. He has been the track co-chair of IEEE ICDCS2022, the symposium co-chair of IEEE ICC2020, ICC 2018 and Globecom 2017, and the general chair of IEEE SECON 2018. He is a senior member of IEEE, the member of ACM and CCF.

报告题目： Privacy-Preserving Data Exchange**报告摘要：**

The blockchain technology can be extensively applied in many areas, from cryptocurrency, DeFi, NFT (Non-Fungible Token) to Web3. In a Web3 platform for data exchange, we need to provide data traceability and privacy while ensuring fairness. Existing systems encrypt the data to protect data privacy, which hinders demanders from verifying the correctness of the data and evaluating its value. In this talk, we propose a novel traceable data exchange scheme based on non-fungible tokens and zero knowledge. Our scheme has several advantages that is able to (i) track all transformations of data during their lifecycle and record them on the blockchain; (ii) provide zero-knowledge proofs to securely guarantee that all complex transformations and data contents are correct and meet demanders' requirements; and (iii) warrant exchange fairness and data privacy in public storage platforms.

**伍前红**
北京航空航天大学**论坛讲者4**

伍前红，北京航空航天大学网络空间安全学院教授，博士生导师，副院长。从事应用密码学与区块链研究与教学工作，出版教材与译著5部，获北京市高等教育教学成果二等奖；获中央网信办网络安全优秀学生指导教师，北京市优秀毕业生指导教师称号；主持和承担国家自然科学基金重点项目、科技部重点研发计划等项目20余项，发表论文200余篇，授权国家发明专利40余项，转化专利近10项，参与制订多项国际标准和国家标准，获2021年度麒麟科学技术奖。

报告题目：面向多链融合的区块链跨链互操作与可扩展关键技术

报告摘要：

现有区块链系统由于各自的密码学算法、数据结构、共识机制等方面存在巨大差异，彼此之间难以实现互联互通，逐渐形成独立封闭的数据和价值孤岛。本报告针对多源、异构链链际融合应用需求，分析大规模可扩展异构区块链跨链模型，探讨区块链链间协同方法、跨链价值转移、跨域审计与监管等新型跨链架构和密码学方法，分析目前针对异构链跨链系统及其应用的研究进展。



刘秀龙
天津大学

论坛讲者5

刘秀龙，天津大学智能与计算学部教授，博导，入选国家级青年人才。近年来主要在区块链、物联网领域开展研究工作，累计发表论文70余篇，成果发表在MobiCom、INFOCOM、TON、TMC、TC、TPDS、IMWUT等多个国内外知名期刊/会议。曾获CCF优秀博士学位论文奖；IEEE TCSC可扩展计算专委会2020年度“杰出青年学者奖”；获IEEE Systems Journal 2017、ICA3PP 2014等期刊和会议的最佳论文奖。担任Computer Communication编委、国际会议IEEE ICPADS 2019程序委员会副主席及中国计算机学会网络与数据通信专委会副秘书长。

报告题目：区块链技术与系统

报告摘要：

区块链是建设下一代互联网可信基础设施的关键技术，对于我国数字经济发展具有重要意义。本报告将分析当前区块链在数据存取、通用测评、跨链交互方面存在的技术挑战，并分享团队近期的相关工作。最后，本报告讨论区块链未来发展方向和趋势。



肖江
华中科技大学

论坛讲者6

肖江，华中科技大学计算机学院教授、博士生导师。主要研究领域为区块链、分布式系统等。具有丰富项目经验，主持了国家重点研发计划青年科学家项目、国家自然科学基金项目、湖北省重点研发计划项目等。在区块链的可扩展存储、高效事务处理、共识性能优化、跨链架构等方面取得多项创新成果。在 IEEE TPDS、IEEE TKDE、IEEE JSAC、INFOCOM、SRDS、ICDCS、ACSAC 等重要国际期刊和学术会议上发表论文70余篇（含 CCF A/B类25篇、ESI 高被引论文2 篇）；单篇论文谷歌最高引用510 次；是全球唯一连续三年在分布式系统领域顶会ICDCS发表区块链研究成果的学者；获国内外发明专利授权10 多项；获国内外最佳论文奖5 次。2014 年博士毕业于香港科技大学计算机科学与工程系（师从香港科技大学（广州）创校校长、IEEE Fellow倪明选教授），2017 年入选中国计算机学会CCF-Intel青年学者提升计划，2018年入选湖北省青年科技晨光计划，2019年入选ACM中国（武汉）新星奖，2022 年入选武汉市青年曙光计划。肖江是ACM/IEEE会员，CCF高级会员。担任区块链前沿学术刊物IEEE Blockchain Technical Briefs创始编委，《Blockchain Research and Applications》编委，《Connection Science》编委，《Frontiers of Computer Science》青年编委，也是CCF区块链专委会、分布式计算与系统专委会、普适计算专委会的委员。曾任 IEEE BlockSys2021 国际会议TPC主席、IEEE CCGrid 2020、IEEE MDM 2019、IEEE ICPADS 2017 和 2018等国际会议的区块链研讨会主席。

报告题目：可扩展图式区块链存储系统的研究与实践

报告摘要：

图式区块链天然具有并行特点，有效提升系统吞吐量，已成为区块链领域的重要发展趋势。随着链上数据量呈现爆炸式增长，现有图式区块链系统面临存储可扩展性瓶颈难题。本报告阐述了区块链存储可扩展性的问题内涵，分析了研究现状与挑战，介绍了团队自主研发的可扩展图链存储系统的弹性数据模型、并发事务处理机制和高效查询检索方法。

分论坛日程

论坛6:新黄金时期下的处理器安全

时间:2023年8月4日13:30 - 17:30

论坛简介:

John Hennessy 和 David Patterson 在 ISCA 2018 的图灵演讲(Turing Lecture)中认为“熔断”、“幽灵”等处理器安全漏洞的发现促成体系结构领域研究即将进入“新的黄金时期”(On the cusp of another Golden Age)的关键动因之一,具有划时代的意义。

处理器作为网络信息系统的硬件基石,长期以来,学术界和产业界以其为“可信基”,如 Intel SGX 和 ARM TrustZone 等,一旦这些处理器硬件出现安全漏洞,加解密算法赖以执行的可信硬件执行环境、以处理器为可信基的大量安全应用、以处理器为硬件基础的网络安全措施等,失去根本基石,网络空间安全秩序和体系架构面临被撼动的风险。

面向未来的“新黄金时期”的安全处理器,本论坛汇聚了本领域具有代表性、引领性的学术界和产业界专家,探讨如何识别与检测处理器硬件安全风险、如何考虑安全处理器架构设计等两个方面内容,期待在处理器安全相关学术研究和产业发展方面,能够促进各位参会专家、学者深入交流,碰撞新的火花,再上新的高度。

具体日程安排(论坛6)

时间	报告内容	报告嘉宾/单位
13:45 - 14:10	领域专用处理器关键技术研究	曾晓洋
14:10 - 14:35	安全芯片关键技术	苏金树
14:35 - 15:00	面向工业互联网安全的高性能计算	李肯立
15:00 - 15:20	茶歇	
15:20 - 15:45	处理器安全与安全处理器	汪东升
15:45 - 16:10	异构计算安全架构	侯锐
16:10 - 16:35	数据流追踪与处理器硬件安全	金意儿
16:30 - 17:30	Panel: 处理器安全产学研合作:路在何方?	

论坛主席:



吕勇强
清华大学

博士, 副研究员, 主要研究处理器漏洞挖掘, 发现 Intel 和 ARM CPU 的电压管理的重大安全漏洞——“骑士”漏洞, 获得亚太硬件安全大会 AsianHost 2019 唯一 Best Paper 奖、第 33 届 ACM CHI 会议(CCF A)最佳论文提名奖, 担任 CCF A 类领域顶级期刊 ACM IMWUT 副主编, 获 2020 年 CCF 自然科学奖一等奖、2019 年电子学会科技进步一等奖、2014 年教育部科技进步一等奖、2015 年国家科技进步奖二等奖。



胡伟
西北工业大学

博士，副教授，博导，在加州大学圣迭戈分校有 5 年的研究经历，研究方向包括集成电路设计安全、处理器安全、形式化安全验证、密码侧信道分析等。在领域顶级和重要期刊、会议上发表论文 50 余篇，出版网络空间安全专业“十三五”规划教材 1 部、专著 1 部。任 IEEE TCAD 客座编辑、硬件安全领域旗舰会议 HOST 和 AsianHOST 组委会委员、ICCAD 分会主席，作大会邀请和分会报告 10 余次。研究成果获 2019 国家保密局保密科学技术奖和 2020 中共中央办公厅党政机要密码科学技术奖。

讲者报告详细内容

论坛讲者：



曾晓洋
复旦大学

论坛讲者1

曾晓洋，复旦大学特聘教授，微电子学院副院长，兼任中国电子学会电路与系统分会副主任委员；曾任专用集成电路与系统国家重点实验室执行主任，IEEE SSCS Shanghai Chapter Chair，IEEE ISSCC TPC Member 等。长期从事高效系统芯片研究与应用工作，发表学术论文 300 余篇，拥有发明专利 120 多件；培养超过 100 位博士硕士研究生，为学术界和产业界输送高级创新人才。曾获国家杰出青年科学基金、国家万人领军人才、长江学者特聘教授、CIE Fellow 等荣誉奖项。

报告题目：领域专用处理器关键技术研究

报告摘要：

随着摩尔定律发展趋缓而应用需求急剧扩张的集成电路发展趋势，芯片架构创新将迎来新黄金时代。受限于传统通用计算架构在能效方面越来越明显的技术瓶颈，领域专用计算和相应处理器系统架构(DSA, Domain-Specific Architectures)得到了学术界和工业界的广泛关注和深入研究应用。本报告从后量子密码领域专用处理器、超低功耗物联网(IoT)安全领域专用处理器、高效动态手势识别领域专用处理器、多核安全处理器平台及其主动防御技术等若干特定领域典型应用为例，重点报告了领域专用处理器系统的架构设计方法及其关键技术等内容。



苏金树
国防科技大学

论坛讲者2

苏金树，国防科技大学教授，博士生导师，“网络技术”教育部创新团队带头人，973项目首席科学家，银河/天河计算机系统副总设计师。

长期从事高性能网络技术研究，在高性能路由器、网络安全和超算网络方面，做出系统性和创造性贡献。提出属性网络体系结构、硬件并行软件分布的路由器系统结构，强制安全控制等模型和方法，攻克了路由交换、安全控制、多路径路由等关键技术，研制了高性能路由器、高密度数据中心交换机、高速网络安全设备等重要装备，在国家和国防信息化建设中取得广泛应用。获国家科技进步特等奖1项、国家科技进步一等奖2项、国家科技进步二等奖4项。获得授权发明专利 46 项，在 SCI/EI/一级学科刊物/等发表学术论文 260 余篇。

报告题目：信息网络基础设施安全挑战与对策

报告摘要：

微处理器，操作系统，网络系统是信息网络基础设施的核心要素，也是信息网络基础设施安全的基石。报告将计算机体系结构变化，谈硬件安全面临的新挑战；从信息网络基础设施安全防护面临的新场景，新对象两方面，提供了生动的案例，分析内

在技术机制, 依据带领团队开展安全可信网络基础设施的亲身体会, 从提高网络空间整体安全性的角度, 提出创新网络新架构, 新机器的建议。



李肯立
湖南大学

论坛讲者3

李肯立, 湖南大学副校长, 信息科学与工程学院院长, CCF 会士。教育部高效能计算学科创新引智基地负责人、高性能计算应用软件技术教育部工程研究中心主任。国家超级计算创新联盟副理事长、新一代人工智能产业技术创新联盟专家委员会委员、高性能计算国家重点研发计划总体专家组成员、IEEE 高级会员、CCF 长沙主席、湖南省计算机学会副理事长、IEEE-TC/TSUSC/TII 副编辑、《计算机研究与发展》编委。主持国家重点研发计划、基金委重大仪器项目、基金委重点项目、JWKJW 重点项目等国家和省部级项目 40 余项。获国家科技进步一等奖创新团队奖 (排名 13) 等奖励 4 项。主要研究领域为并行与分布式处理、超级计算与云计算、面向大数据和人工智能的高效能计算等。

报告题目: 面向工业互联网安全的高性能计算

报告摘要:

近年来在工业互联网国家战略推动下, 大规模复杂工业应用层出不穷, 对于算力的要求与日俱增, 工业互联网应用也成为了高性能计算的重要应用领域。本报告将探讨工业互联网与智能制造应用背景下, 高性能计算面临的挑战、关键技术与未来发展方向。



汪东升
清华大学, 教授

论坛讲者4

汪东升, 清华大学计算机系 长聘教授, 主要从事处理器安全和计算系统架构等方面研究, 主持研发当时国内主频最高的微处理器 THUMP-107, 2019 年率领团队发现 INTEL、ARM 等现代主流处理器重大漏洞 -VoltJokey (CVE-2019-11157), 曾主持完成国家自然科学基金重点项目、973 项目课题、863 重点项目、国家重点研发项目等多项国家级项目, 在顶级国际会议和期刊上发表论文 100 余篇、获得国家发明专利 80 余项。曾获得国家科技进步二等奖 1 次, 省部级一等奖 3 次、二等奖 3 次。

报告题目: 处理器安全与安全处理器

报告摘要:

处理器安全漏洞不断曝光, 处理器安全问题日益严峻。如何检测和识别处理器漏洞、如何发现新的漏洞, 成为网络空间安全领域和处理器设计领域共同关注的新热点难题。本报告将主要介绍清华大学团队在该领域的研究工作, 包括相关重要漏洞的发现, 以及面向未来自动化漏洞挖掘的一些思考。



侯锐
中国科学院
信工所研究员

论坛讲者5

侯锐, 中国科学院信息工程研究所信息安全国家重点实验室副主任、研究员、博士生导师, 获得国家杰青、优青项目资助。长期从事处理器芯片架构设计及芯片安全等方面的研究工作。

报告题目: 异构计算安全架构

报告摘要:

异构计算已经成为高效能计算的重要架构和优化核心方法。数十年来, 异构计算的核心问题始终是性能优化。近年来, 异构计算的安全性越发重要、亟待解决。本报告将简要介绍异构安全架构的进展, 分享报告人团队的研究进展。



金意儿
华为公司

论坛讲者6

金意儿, 华为可信计算首席科学家, 佛罗里达大学名誉教授, IEEE硬件安全与可信专委会联席主席。他的研究领域主要涉及软硬件协同安全和新兴集成电路安全, 包括硬件支持的系统安全, 集成电路产业链安全, 以及可信自动化等。他撰写了集成电路安全一书, 同时在国际知名期刊和杂志上发表了超过200篇论文, 是亚洲硬件安全年会的联合创始人。他目前是IEEE设计自动化委员会的杰出讲师。他的论文获得了多项最佳论文奖, 他本人也获得包括美国能源部和美国海军在内的多项杰出青年教授奖。

报告题目: 数据流追踪与处理器硬件安全

报告摘要:

数据流追踪(Information Flow Tracking)技术被提出四十多年来, 已经被广泛地应用于复杂程序中信号的安全性和完整性保护。但是, 信息流追踪在硬件安全领域的应用远不如它在软件领域那么流行, 在本次讲座中, 我们会首先介绍数据流追踪在硬件安全领域的学术界和业界的最新尝试。接着, 我们会以模拟电路层面的信息流追踪和体系结构层面的动态信息流追踪为例, 介绍信息流技术可以帮助硬件安全研究人员检测硬件木马或者提升系统安全。其中, 模拟电路层面的信息流追踪技术可以帮助识别最新的模拟木马电路, 而这类木马电路很难被传统的数字电路领域的安全检测方法所发现。体系结构层面的动态信息流追踪技术, 克服了当前动态信息流追踪往往伴随着较大的软件或者硬件的开销这一问题, 是一种基于协处理器的细粒度动态信息流追踪技术, 这一技术利用硬件协处理器提供细粒度的信息流追踪服务, 极大地提升了细粒度数据流追踪能力, 且总体开销较小。



黄琳
蚂蚁集团

Panel嘉宾

黄琳, 博士, 蚂蚁集团基础安全全部, 资深安全专家, 目前负责蚂蚁集团端平台体系的整体安全保障工作, 涵盖移动端APP、小程序、IoT等方向, 同时也负责形式化验证组、硬件侧信道安全组等研究小组的工作。牵头集团级客户端软硬件产品内部安全标准和规章制度, 以及上线前安全检测系统的技术建设, 并主持上线前安全检测、上线后安全巡检和应急等业务工作的开展。擅长端安全、无线安全, 具有十余年安全、通信、电子、芯片的多领域的技术与实战经验。



邱朋飞
北京邮电大学

Panel嘉宾

邱朋飞, 博士, 北京邮电大学网络空间安全学院副研究员, 博士生导师。主要从事处理器安全研究, 包括漏洞挖掘和检测、安全架构和安全模型设计等, 发现的通用处理器“骑士”漏洞是国内研究团队发现的首个处理器硬件漏洞, 影响大量已应用的处理器, 此后又相继发现处理器的“频率隐藏通道”、“PMU侧信道”和“瞬态执行侧信道”漏洞。邱朋飞以第一作者身份在CCS、DAC、IEEE TCAD等国际会议和期刊上发表12篇学术论文, 获得中国计算机学会科学技术奖自然科学一等奖、湖南省自然科学奖二等奖、清华大学优秀博士学位论文奖、Intel漏洞奖励等。主持国家自然科学基金青年基金项目一项; 作为骨干参与科技部重点研发、JKW基础加强、国家自然科学基金面上、华为公司横向合作等多项项目。



张福新
龙芯

Panel嘉宾

张福新, 博士, 中国科学院计算技术研究所正高级工程师, 博士生导师, 龙芯实验室主任, CCF体系结构专委会常务委员。主要研究方向包括微处理设计、二进制翻译与操作系统等, 长期致力于龙芯系列处理器设计与生态建设, 曾主持核高基重大专项、863重点项目等国家科技项目, 曾获中国科学院杰出成就奖、中国青年五四奖章和国务院特殊津贴专家等荣誉。

分论坛日程

论坛7:高性能图计算体系结构和系统软件

时间:2023年8月4日13:30 - 17:30

论坛简介:

近年,随着大数据和人工智能的蓬勃发展,数据之间的关联关系变化速度日益加快,数据自身及其关联关系的附属信息也日益丰富。为了从这些数据中获取有用信息,新型图计算任务(例如新型图神经网络、流图计算、超图计算)不断涌现,图计算需求日益复杂多样。复杂多样化的图计算需求给现有图计算体系结构和系统软件带来了巨大挑战。如何设计新型高性能图计算体系结构和系统软件满足复杂多样的图计算需求是一个亟待解决的难题。为此,本论坛将组织相关专家讨论新型图计算模型以及新型高效图计算体系结构和系统软件设计,以高效支撑不断涌现的复杂图计算需求,助力图计算体系结构和系统软件发展和生态成熟。

具体日程安排(论坛7)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	数通与计算业务场景的Process In Memory (PIM) 技术机会与挑战	蒋力
14:00 - 14:30	增量图计算的技术与系统	张岩峰
14:30 - 15:00	面向不规则计算的硬件优化: 稀疏矩阵加速与近存计算架构	高鸣宇
15:00 - 15:30	基于可扩展范式的图神经网络系统	杨智
15:30 - 15:50	茶歇	
15:50 - 16:20	稀疏图计算支撑高效点云	戴国浩
16:20 - 16:50	面向图算法的“软硬兼施”加速研究	严明玉
16:50 - 17:20	亚秒级点到点图查询系统	章明星

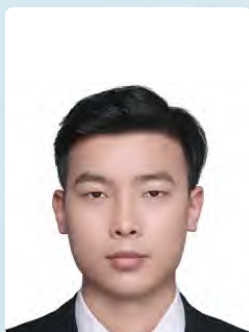
论坛主席:



张宇

华中科技大学计算机科学与技术学院

张宇, 华中科技大学计算机科学与技术学院副教授, 博士生导师, CCF高级会员, CCF高性能专委会委员、CCF体系结构专委会委员、CCF分布式计算与系统专委会委员, 之江实验室特聘专家, 2016年6月在华中科技大学获得计算机系统结构博士学位, 主要研究高性能图计算、体系结构和系统软件, 主持国家自然科学基金和企事业单位项目10余项, 参与国家重点研发计划项目、国家自然科学基金重点基金、863计划课题等, 在ISCA、HPCA、ASPLOS、SC、DAC、USENIX ATC等CCF A/B类和IEEE/ACM Transactions上发表论文60余篇, 获CCF A类会议USENIX ATC 2018最佳论文提名和最佳存储相关论文等, 成果应用于华为等多家龙头企业, 获评CCF高性能计算“卓越青年”、ACM中国新星奖等。



赵进

之江实验室图计算研究中心

赵进, 之江实验室图计算研究中心助理研究员, 主要研究面向图计算的高性能体系结构和系统软件。2022年9月在华中科技大学获得计算机系统结构博士学位, 入选第八届“长江学子”大学生就业创业人物“开拓创新类”。主持CCF-蚂蚁科研基金和之江实验室青年基金等, 在ISCA、SC、DAC、ACM TOS/TACO、IEEE TC/TKDE/TBD、JCST等会议和期刊上发表或录用学术论文10余篇, 入选CCF A类期刊IEEE TC的Featured Paper of the Month等。

讲者报告详细内容

论坛讲者:



蒋力
上海交通大学
华为技术有限公司

论坛讲者1

蒋力, 上海交通大学 / 华为技术有限公司, 副教授 / 异构通信网络处理器技术首席科学家。从事计算机体系结构与芯片设计自动化领域的研究。在计算机体系结构及芯片设计自动化领域的顶级 / 主流国际会议和期刊, 如 DAC, TCAD, ISCA, MICRO, DATE, TC, TVLSI, 以及人工智能领域顶级会议, 如 AAAI, ICCV, SIGIR 等上发表论文 96 篇, 其中获得 DATE22, DATE23 最佳论文, 以及 3 次最佳论文提名。在 MICRO, DATE, ASP-DAC, ITC-Asia, ATS, CFTC, CTC 等国际和国家会议中担任联席主席和 TPC 成员。他是 IET Computers & Digital Techniques (SCIE/SSCI), Integration the VLSI Journal (SCIE) 期刊编委。CCF FTC/CTC, ITC-Asia 等多个学术会议的 TPC 主席, 领域主席, 以及 DATE, MICRO, ASP-DAC 等国际会议 TPC 成员。CCF YOCSEF 上海主席 (2020-2021)。曾获 ACM 上海新星奖, CCF 集成电路 Early Career Award, 吴文俊人工智能科学技术奖 芯片专项 二等奖, 教育部 - 华为智能基座“突出贡献奖”。

报告题目: 数通与计算业务场景的Process In Memory(PIM)技术机会与挑战

报告摘要:

服务器 SSD 盘控的过滤、解 / 压缩, 计算后台和通信(网络)处理器的图数据库查询、安全规则匹配、转发查表等元数据操作, 都存在大量访存密集但算力小的算子, 通过 PIM 卸载能极大提升 CPU 的效率和算力。但也存在异构计算, 异构存储等系统架构挑战。



张岩峰
东北大学计算机
科学与工程学院

论坛讲者2

CCF数据库、信息系统、分布式计算与系统专委会执行委员, 东北大学教授, 博士生导师, 东北大学与美国麻省州立大学联合培养博士, 国家级青年人才计划入选者。主要研究方向为图计算系统、分布式数据库、机器学习系统等。承担或参与国家重点研发计划、国家自然科学基金等多项国家级科研项目, 承担华为、阿里巴巴、百度等校企合作项目。研究成果发表在SIGMOD、VLDB、PPoPP、SOCC等重要期刊会议上。曾获云计算国际会议ACM SOCC等会议的优秀论文奖、华为年度优秀项目奖、辽宁省科技进步奖等奖励。

报告题目: 增量图计算的技术与系统

报告摘要:

真实世界中采集的图数据不断更新和变化, 顶点和边的属性动态更新, 顶点和边也会被插入和删除。图数据和图结构的动态变化导致图分析的结果需要动态更新, 这对实时图分析提出了挑战。针对动态图数据, 传统面向静态图的图计算系统需要在更新后的图快照上重新执行图分析算法, 会导致大量的冗余计算。增量图计算利用旧图上的计算结果, 期望只计算增量变化的图数据, 来增量更新已有图分析结果, 这能够减少大量的冗余计算, 大大提高图分析的效率。面对每日新增的海量动态图数据, 大规模增量图分析同样也是工业界的重要需求之一。本报告将介绍一系列近年来提出的增量图计算技术与系统, 同与会专家探讨增量图计算未来的机遇与挑战。



高鸣宇
清华大学交叉
信息研究院

论坛讲者3

清华大学交叉信息研究院助理教授, 博士生导师。于美国斯坦福大学电子工程系获博士学位。研究方向为计算机体系结构与系统, 主要关注针对人工智能和大数据分析等数据密集型应用的新型存储架构、专用计算系统、硬件系统安全等方向。已发表多篇国际顶级学术会议 (ISCA, ASPLOS, HPCA, OSDI, PACT 等) 论文, 曾获得IEEE Micro 2016年度计算机系统结构最佳论文奖 (Top Picks)、福布斯中国2019年30 Under 30等荣誉。他也是MICRO、ASPLOS、ISCA等多个会议的TPC和ERC委员会成员。

报告题目: 面向不规则计算的硬件优化:稀疏矩阵加速与近存计算架构

报告摘要:

Sparse computations on matrix/tensor data are widely used in deep learning and graph applications, but efficiently supporting sparse computations on existing systems is challenging due to irregular data distribution and poor access locality. We argue that, since the irregular data distribution is difficult to predict beforehand, it would be necessary to make the hardware architecture aware of and dynamically adaptive to the specific workload at runtime. We present two example designs in this talk. The first one is a sparse matrix accelerator that uses adaptive dataflow based on data sparsity patterns. The second one is a near-data processing system that co-optimizes remote access and load imbalance issues. These two designs are both published in ASPLOS 2023.



杨智
北京大学计算机
学院

论坛讲者4

北京大学信计算机学院副研究员，博士生导师、国家高层次青年人才。主要研究方向为深度学习模型的分布式并行、计算优化、算法和系统设计。在 OSDI、ATC、EuroSys、SIGMOD、VLDB 等高水平国际学术会议上发表论文 50 多篇，获得过 VLDB 2022 最佳论文奖、WWW 2022 最佳学生论文奖、高校科学研究优秀成果奖、湖北省科技进步一等奖、中国电子学会科技进步一等奖、北京大学黄廷方 / 信和青年杰出学者奖等。

报告题目: 基于可扩展范式的图神经网络系统**报告摘要:**

图神经网络(Graph Neural Networks, GNNs)在各种基于图的任务中取得了最先进的性能。然而，由于主流 GNNs 是基于神经信息传递机制设计的，它们在数据规模和信息传递步骤上无法很好地扩展。尽管目前的框架中应用了采样和分布式训练，仍然面临高通信成本的问题。该报告介绍一种新的 GNN 范式和系统，它提供了一种系统地构建和探索可扩展 GNNs 设计空间的原则方法。通过对信息传递机制进行解耦，提出了一种新颖的可扩展图神经网络体系结构范式(Scalable Graph Neural Architecture Paradigm, SGAP)，以及一个由 150k 种不同设计组成的通用架构设计空间。按照这一范式，我们实现了一个自动搜索引擎，可以通过多目标优化自动搜索性能良好且可扩展的 GNN 架构，以平衡多个标准(例如准确性和效率)之间的权衡。



戴国浩
上海交通大学
清源研究院

论坛讲者5

上海交通大学长聘教轨副教授，清源研究院定制计算中心(CCC, Customized Computing Center)Co-PI，人工智能设计自动化创新实验室(DAI Group, Domain-specific Artificial Intelligence & Design Automation Innovation Group)负责人。分别于 2014 年和 2019 年在清华大学电子工程系获得工学学士和博士学位，发表电路设计自动化、异构计算、体系架构等领域高水平国际期刊、会议论文 30 余篇，谷歌学术施引 800 余次。发表论文荣获 ASP-DAC 2019 最佳论文奖，DATE 2023/DAC 2022/DATE 2018 最佳论文提名，WAIC 2022 优秀青年论文奖。个人曾获 WAIC 2022 云帆奖、NeurIPS21 BIGANN 竞赛全球冠军、北京市优秀博士毕业生、清华大学优秀博士毕业生、清华大学优秀博士毕业论文等荣誉。戴教授曾参与指导学生获 ACM 2021 SRC 全球第三，MICRO 2020 SRC 全球第一。作为项目负责人，戴教授承担了包括国家自然科学基金青年项目等在内的多项纵向项目。

报告题目: 稀疏图计算支撑高效点云**报告摘要:**

随着人工智能在更广泛场景的应用与落地，3D点云神经网络已经在物体检测、语义分割等各种3D感知应用中发挥着不可替代的重要作用。目前，已经出现了一些基于GPU或基于专用架构的点云计算方案。然而，点云数据具有高度稀疏的特点，并且稀疏性来源于实际物体本身的深度和表面信息，需要在运行时动态处理稀疏点云。在现有点云网络的推理计算过程中，存在访存量、访存效率和计算单元利用率低的问题。并且随着算法的不断更新，感知范围扩大、场景扩大，点云数据变得越来越庞杂；并且需要同时运算多帧、处理融合多模态的数据。在这种背景下，访存和计算单元利用率低的问题会越来越严峻，无法满足端侧应用的实时性、低延时、低能耗的需求。因此，目前3D点云神经网络处理系统仍需要1-2个数量级的效率优化。面向3D点云神经网络，我们在计算流图、并行算子、硬件架构三个层面，从访存和计算两个角度，对3D点云计算系统进行细粒度优化，形成了快速高性能的GPU点云计算系统和架构设计，最终将3D点云算法计算性能提升1.5至4倍。未来，我们将继续开展广泛人工智能场景下快速高性能稀疏计算的系统与芯片研究，支撑人工智能向下一个时代的发展与进步。



严明玉
中国科学院计算
技术研究所

论坛讲者6

CCF专业会员、中国科学院计算技术研究所副研究员。2020年于中国科学院计算技术研究所获得工学博士学位，并留所工作。研究方向是图算法的“软硬兼施”加速。在包括MICRO、HPCA、AAAI、IJCAI等在内的领域顶级会议或期刊上发表论文20余篇。入选CCF青年人才发展计划、中国科学院青促会、中国科学院青年团队计划，曾获北京市技术发明一等奖。博士论文《图处理加速架构》入选2022年度CCF优秀博士学位论文。

报告题目：面向图算法的“软硬兼施”加速研究

报告摘要：

图(Graph)是描述和分析实体之间关系/交互的通用语言，被广泛应用于表示各种日常生活数据和科研数据，例如地铁路线图、分子拓扑等等。图算法具有超强的图数据处理能力，已经被应用于路线导航、蛋白质功能预测、知识推理等重要领域。例如，图神经网络算法就已经被应用于许多知名企业的推荐系统中。随着图数据规模的不断增大，图算法的运行时间也不断增长。因此，图算法的加速研究就变得尤为重要。该报告将会介绍我们团队通过“软硬兼施”的研究路线加速图算法的相关工作。



章明星
清华大学计算机
科学与技术系

论坛讲者7

清华大学博士。主要在计算机系统机构领域开展创新性和实用性相结合的研究工作，提出了系列新思路和新方法，相关成果在包括 OSDI、SOSP、ASPLOS、HPCA、FSE、VLDB、ATC、EuroSys 等国际顶级会议和期刊上发表论文二十余篇。其中发表在 A 类会议 FSE 2014 的论文获 ACM SIGSOFT 杰出论文奖，先后获得清华大学、北京市、IEEE TCSC、ACM SIGOPS 等组织颁发的优秀博士学位论文奖。获得国家自然科学基金青年科学基金及博士后基金一等项目资助。CCF 推荐中国科协青年人才托举计划入选，科技部重点研发项目课题负责人。

博士毕业后进入清华大学与深信服联合培养的博士后计划，负责创新业务孵化，担任深信服首席算法技术专家，后任职创新研究院院长。先后主导基于机器学习的恶意软件识别引擎(SAVE)、安全大数据平台、RDS 等创新项目孵化。其中 SAVE 引擎服务于上万家客户，2018 年获得 SKD Awards，2019 年获得 Google VirusTotal、德国 AV-Test 等多项国际机构认可。

2022 年加入清华大学计算机系高性能所，任职教研系列助理教授。

报告题目：亚秒级点对点图查询系统

报告摘要：

尽管当前实时数据仓库的发展使得在大规模关系型数据上进行亚秒级别分析查询已经可行，对于同等规模的图数据，以同样的时延响应图分析查询却任然是一项巨大的挑战。主要的原因在于图数据之间存在广泛且复杂的关联关系，单个图查询就可能需要访问大部分的图数据。在我们的研究中，我们发现剪枝技术对于消除冗余计算以提高查询效率的重要性。但是，目前已有的仅仅基于上界的剪枝技术在处理大规模图时的效率还远未达到预期。因此，我们提出了一种基于下界剪枝的方法，能有效地提升查询处理速度。在这个方法的基础上，我们开发了 SGraph 系统，它能够在亚秒级别响应并处理图上的动态点对点查询，展现出远超现有系统的出色性能。

分论坛日程

论坛8:高性能互连论坛

时间:2023年8月4日13:30 - 17:30

论坛简介：

高性能互连网络是高性能计算机系统和云数据中心系统的关键基础设施，直接决定系统性能和均衡扩展能力。近年来工业界和学术界在高性能融合网络协议、在网计算、高吞吐芯片设计、片上与片间通信链路等方面开展了大量工作。高性能互连论坛将从芯片内互连，芯片间互连，应用驱动互连等不同维度，探讨本领域的前沿技术。论坛邀请国内著名企业和学术机构的专家学者，分享实际部署经验和创新性研究成果。

具体日程安排(论坛8)

时间	报告内容	报告嘉宾/单位
13:30 - 13:40	嘉宾介绍	
13:40 - 14:20	大模型训练的网络优化技术研究	李丹/清华大学
14:20 - 15:00	Tensor, Collective-Op & Flow -- AI大集群横向扩展的通信挑战	付斌章/阿里巴巴
15:00 - 15:20	茶歇	
15:20 - 16:00	面向未来超算系统的先进互连技术展望	齐星云/国防科大
16:00 - 16:40	AI算力时代的高性能网络互联系统和理论	朱夏/华为
16:40 - 17:20	面向模型训练的在网计算协议设计、开发框架和管理系统	吴文斐/北京大学
17:20 - 17:45	Panel:高性能互连前沿技术探究	

论坛主席:



董德尊
国防科技大学

董德尊, 国防科技大学计算机学院研究员、博导, 教育部长江学者、全国百篇优博、国防科技卓青, 天河互连副主任设计师, 军科委创新特区专家, 长期参与银河 / 天河系列高性能计算机系统研制工作, 获军队科技进步一等奖、湖南省自然科学一等奖、湖南省教学成果特等奖。主持 10 余项国家和军队科研项目。在 CCF 推荐的国际刊物发表论文 100 余篇。在国内外会议 / 论坛担任程序 / 组织 / 宣传委员会主席 / 委员等 40 余次。



雷斐
国防科技大学

雷斐, 国防科技大学计算机学院高级工程师, 2011 年本科毕业于同济大学, 2013 年、2018 年于国防科技大学计算机学院取得硕士、博士学位。长期从事高性能互连网络和聚合通信库的研究, 在体系结构和高性能计算领域高水平会议与期刊发表论文十余篇。

讲者报告详细内容

论坛讲者:



李丹
清华大学

论坛讲者1

李丹, 清华大学计算机系长聘教授, 教育部长江学者特聘教授, 国家 973 计划项目首席科学家, 国家重点研发计划项目负责人, 国家十四五重点研发计划“网络空间安全治理”专家组副组长。主要研究方向为网络安全、网络智能、数据中心网络。发表学术论文 100 余篇, 获授权专利 30 余项。获教育部“青年科学奖”, 以第一完成人获中国通信学会技术发明一等奖。

报告题目：大模型训练的网络优化技术研究**报告摘要：**

在大模型的分布式训练过程中，网络通信往往成为系统性能的瓶颈。针对分布式机器学习训练的特点，可以从网络拓扑构建、网络流量调度、参数同步算法等方面进行网络技术优化，从而提升分布式大模型训练的整体性能。本报告主要介绍演讲人在该方向的部分研究成果。



付斌章
阿里巴巴

论坛讲者2

付斌章，博士，阿里云基础设施网络研发事业部高性能网络业务负责人，CCF高性能专委会委员，在高性能互连网络领域发表高水平论文40余篇。

报告题目：Tensor, Collective-Op & Flow -- AI大集群横向扩展的通信挑战**报告摘要：**

大规模分布式机器学习应用的快速发展要求更多的高性能异构计算集群的体系结构创新。在该领域，阿里云具有丰富的研发、建设和运营经验。为了给客户提供更加专业的服务，阿里云发布了全新的灵骏智算产品，可以在提供更强大算力的同时提供更多弹性以及更可预期的集群稳定性。在整个集群架构中，高性能网络是核心技术之一。本次汇报将聚焦灵骏智算高性能网络解决方案PredFabric，并重点介绍如何通过网络拓扑、负载均衡算法、拥塞控制算法以及通信库的联合创新解决大语言预训练模型（例如通义千问）的并行扩展难题，以及相关的未来工作。



齐星云
国防科技大学

论坛讲者3

齐星云，国防科技大学计算机学院副研究员，硕士生导师，计算机系统结构研究室主任。长期从事国产高性能计算机系统和自主高性能互连网络的研究，研制成功的高性能互连网络芯片组性能指标达到国际先进水平。主要研究方向为高性能计算机系统结构，高性能互连网络，大规模集成电路芯片设计等。获省部级科技进步一等奖2项，以第一作者在国内外学术期刊和会议发表学术论文10余篇，授权专利20余项。主持军队重点科研项目2项，参与国家重点研发计划等国家重点科研项目2项。

报告题目：面向未来超算系统的先进互连技术展望**报告摘要：**

为了应对未来国产超算系统对系统规模快速增长的需求，具有低延时、高带宽、易管理、高可靠、低成本等特性的下一代先进互连技术将起到关键作用。结合“天河”系列超算大规模互连网络系统的研制实践，我们从多个不同的维度展望了未来高速互连系统的关键技术和发展趋势，包括芯片内的高性能片上网络、封装内的D2D短距互连技术、芯片间的光电共模互连技术、系统内的低延时高带宽互连技术等。这些技术将可以针对不同层次的互连需求，很好地满足未来高性能计算系统和数据中心大规模通信和互连的需求。



朱夏
华为

论坛讲者4

朱夏，华为技术有限公司网络技术实验室高级研究员。2014年博士毕业于东南大学计算机科学与工程学院，同年加入华为技术有限公司。现担任华为中央研究院网络技术实验室网络理论/高性能网络能力域主任，网络基础技术项目经理。在公司期间主持研究了10+项关键预研项目，申请专利40+件。

报告题目：AI算力时代的高性能网络互联系统和理论

报告摘要:

随着AI大模型时代的到来,云上云下都呈现出AI重算力的爆发式增长。AI应用和系统对网络提出了极高要求。针对AI计算和云等新兴业务的网络研究呈现出“承上启下”的趋势。一方面,网络需要加大与上层计算应用和系统的深度融合,在联合调度/协同优化等方面考虑系统化设计,进而提升计算效率。另一方面,网络研究范围需要向内(主机总线/OS软件)向下(物理拓扑)延伸,通过延展网络边界提升网络传输效率。本报告将系统阐述AI算力时代的诸多网络挑战,并给出网络系统设计和理论的潜在研究方向。



吴文斐
北京大学

论坛讲者5

吴文斐,北京大学计算机学院研究员,博士生导师。长期从事计算机网络和分布式系统方向的研究。研究兴趣包括:分布式系统体系结构和计算机网络管理。在计算机网络和体系结构领域发表论文50余篇,包括CCFA类论文18篇。主持研发的在网计算加速的分布式机器学习系统获得NSDI21最佳论文奖(中国首次);主持研发的在网计算加速的大数据系统获得ASPLOS23杰出论文奖;其他成果获得IPCC19最佳论文提名、SoCC13最佳学生论文奖等。

报告题目: 面向模型训练的在网计算协议设计、开发框架和管理系统

报告摘要:

高性能计算中的集合通信在部署到特定网络拓扑时,往往产生一对多、多对一的通信模式,造成通信和系统整体效率的瓶颈。现代可编程交换机具备高速计算和传输能力,可以完成集合通信中的归约和复制操作,从而消除瓶颈。本报告将介绍高通量、高可靠、高资源利用率的在网计算协议设计方法,适配异构应用的在网计算协议开发框架,和能够为多应用做运行时资源管理的在网计算资源管理系统。

分论坛日程

论坛9:量子计算系统论坛

时间:2023年8月4日13:30 - 17:30

论坛简介:

使用面向特定领域的专用协处理器是计算系统提升算力的有效手段,近年来量子计算取得了诸多令人兴奋的发展,有望成为面向量子化学模拟、优化问题等领域的新型高速协处理器。一个完整的量子计算系统需要量子算法、量子编程和编译、量子测控系统、量子芯片等多个层次的协同,并使用量子容错技术或噪声缓解技术来抵抗或接纳噪声的影响。本论坛旨在向大家介绍全栈量子计算系统的整体工作原理。为此,本论坛将邀请六位讲者依次介绍以上五个领域,并提供一个完整的量子计算系统示例。希望本论坛有助于在量子计算系统的构建上形成一些共识,并推动量子软件与硬件之间、量子计算与经典计算之间的更紧密的结合。

具体日程安排(论坛9)

时间	报告内容	报告嘉宾/单位
13:45 - 14:10	超导量子计算	朱晓波/中国科学技术大学
14:10 - 14:35	量子算法:量子世界的计算之法	李绿周/中山大学
14:35 - 15:00	量子程序设计与软件开发	周立/中国科学院软件研究所
15:00 - 15:20	茶歇	
15:20 - 15:45	量子控制微体系结构并行化研究	张孟禹/腾讯量子实验室
15:45 - 16:10	量子纠错和量子错误缓解	李颖/中国工程物理研究院
16:10 - 16:35	一种光量子计算系统	刘英文/国防科技大学
16:35 - 17:35	Panel 讨论:NISQ时代下的量子计算系统研究	

论坛主席:



付祥

国防科技大学
计算机学院

付祥, 国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室 (HPCL) 副研究员。2011 年本科毕业于清华大学电子工程系, 2014 年硕士毕业于国防科学技术大学计算机学院, 2018 年底博士毕业于荷兰代尔夫特理工大学(研究方向量子控制体系结构)。付祥的主要研究目标是以可编程的方式控制量子芯片, 代表工作包括第一个量子控制微体系结构 QuMA、可执行的量子计算指令集 eQASM 和全面的量子 - 经典异构编程框架 Quingo, 曾获体系结构顶级会议 MICRO-50 最佳论文奖, 并入选 2017 体系结构会议论文最佳选录 (Top Picks' 17)。付祥最近感兴趣的问题包括量子程序语言、量子编译和量子计算体系结构等。



劳玲玲

西北工业大学
计算机学院

劳玲玲, 西北工业大学计算机学院教授、硕士生导师。于 2012 年获得哈尔滨工业大学学士学位, 2014 年获得西北工业大学硕士学位, 2019 年获得荷兰代尔夫特理工大学博士学位。2020 年至 2022 年于英国伦敦大学学院进行博士后研究。主要研究方向为容错量子计算, 量子纠错, 量子计算机体系结构, 量子编译等。发表论文 10 余篇, 包括 2 篇 CCF A 类会议 ISCA、1 篇 CCF A 类期刊 TCAD、2 篇中国科学院一区 TOP 期刊 QST 等。

讲者报告详细内容

论坛讲者:



朱晓波

中国科学技术
大学

论坛讲者1

朱晓波, 中国科学技术大学教授。主要从事超导量子计算以及超导约瑟夫森结系统的研究。在磁通量子比特与金刚石中的 NV 色心的量子混合系统上做出了一系列的重要工作。先后创造了超导量子比特最大纠缠数目纪录。研制了超导量子计算原型机“祖冲之号”, 实现了“量子优越性”。

报告题目: 超导量子计算

报告摘要:

由于量子计算在某些问题的处理能力上相比于经典计算机有着压倒性的优势, 被普遍认为是下一代的计算技术, 因而引起了广泛的关注。超导方案因具有良好的可扩展性, 目前备受关注, 各大公司纷纷投资进入该领域。本次报告将主要讲解超导量子计算的现状及近期和中远期目标, 并介绍我们在该方向上取得的一系列进展。



李绿周
中山大学

论坛讲者2

李绿周, 中山大学计算机学院教授、量子计算与软件研究所所长、计算机科学系主任、中国计算机学会(CCF)量子计算专业组副主任、CCF杰出会员、CCF杰出演讲者。长期从事量子计算研究, 主要研究兴趣为量子算法与复杂性、量子计算模型、量子机器学习、量子编译与优化等, 发表学术论文70余篇, 出版学术专著1部。曾获广东省杰出青年基金、湖南省自然科学二等奖、CCF优秀博士学位论文奖、中山大学青年教师授课大赛一等奖等。

报告题目: 量子算法:量子世界的计算之法

报告摘要:

量子计算领域的两大核心问题是:如何将量子计算系统扩展到能够解决实际问题的规模?如何在科学和工业中最大程度的利用量子计算的强大算力?解决第二问题的关键在于能否针对实际问题设计足够聪明的量子算法, 即“量子计算的有用性有赖于量子算法”。本报告将介绍量子算法研究的几个层面、核心问题、历史发展脉络等。



周立
中国科学院软件
研究所

论坛讲者3

周立, 中国科学院软件研究所副研究员。毕业于清华大学计算机系, 之后于马克斯普朗克安全与隐私研究所进行博士后研究。主要研究方向为量子程序设计方法学, 包括对量子程序的验证、计算机辅助证明、测试与调试等。近年来多个原创性工作发表于顶级会议如POPL, PLDI, LICS等。曾与合作者共同获得过OOPSLA'20和LICS'22的杰出论文奖。与合作者共同开发首个实用的后量子密码协议验证工具EasyPQC(CCS'21)以及首个经过验证的量子程序验证平台CoqQ(POPL'23)。

报告题目: 量子程序设计与软件开发

报告摘要:

本报告将简要介绍量子程序设计与软件开发。量子程序设计旨在用高级结构化的语言描述量子算法, 充分利用和发挥量子硬件独有的强大计算能力, 是量子计算走向实用化的关键技术之一。我们将讨论了量子软件开发的各个层面, 包括语言设计和静态分析、动态测试与调试、编译与优化技术等核心组成部分, 分析与经典程序开发的关键区别, 探讨其中所面临的挑战与限制, 总结近年来该领域的主要进展和突破。



张孟禹
腾讯量子实验室

论坛讲者4

张孟禹, 腾讯量子实验室研究员, 2018年研究生毕业于代尔夫特理工大学, 同年加入腾讯。主要研究兴趣:量子控制微体系架构, 量子指令集, 测控系统硬件以及纠错解码架构等。

报告题目: 量子控制微体系结构并行化研究

报告摘要:

量子控制体系结构做为量子计算机的中间层, 起着帮助硬件层和算法层进行协作和交互的作用。随着量子比特数量和可执行电路层数的不断增长, 这一架构的效率和可扩展性也受到了挑战。因为超导量子比特相对较短的退相干时间(在多比特系统中目前常见为几十微秒), 量子控制微体系架构中任何的处理不及时都会导致额外累计的退相干错误, 从而使得计算结果的正确性降低。本报告将介绍我们为了解决该问题所做的工作, 以及腾讯量子实验室内关于测控系统硬件, 硬件纠错解码器等相关工作的进展与成果。



李颖
中国工程物理
研究院

论坛讲者5

李颖, 中国工程物理研究院研究生院副教授。从事量子计算理论研究, 研究兴趣包括量子纠错、量子错误缓解、量子计算的网路架构和量子算法等。2006年本科毕业于南开大学, 2009年硕士毕业于南开大学, 2013年博士毕业于新加坡国立大学。2013年到2017年在牛津大学从事博士后工作, 2017年回国到中国工程物理研究院研究生院工作。

报告题目: 量子纠错和量子错误缓解

报告摘要:

通用量子计算可以通过一系列的量子门实现任意的量子算法。为了实现通用量子计算, 我们需要量子纠错来防止由退相干导致的计算错误。然而, 量子纠错需要大量的量子比特, 因此在近期内还无法实现。量子错误缓解是另一种实现高精度量子计算的方法, 并且由于不需要大量量子比特而适用于近期的量子计算技术。在这次报告中, 我将对量子纠错和量子错误缓解进行介绍。



刘英文
国防科技大学

论坛讲者6

刘英文, 博士, 国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室助理研究员, CCF体系结构专委会委员。从事量子计算方向研究, 参与研制银河鲲鹏QW2020量子计算系统、QUANTA VQS16量子计算系统等, 主要研究领域包括光量子芯片、量子光源等方面, 主持国家自然科学基金青年基金一项、参与国家自然科学基金重点项目, 在Optics Letters、Optics Express等期刊发表多篇学术论文。

报告题目: 一种光量子计算系统

报告摘要:

本报告将介绍一种专用的光量子计算系统的具体实现, 深入分析系统的组成结构和工作原理。报告从大规模光量子计算芯片, 到量子芯片测控与编程, 再到量子算法的实现, 以全栈的视角展示一台完整的量子计算系统。

分论坛日程

论坛10:Tutorial: 量子计算

时间:2023年8月5日13:30 - 17:30

论坛简介:

量子计算有望加速解决经典计算机难以解决的一些问题, 如人工智能、生物医药、量子化学模拟等。因此, 与GPU、TPU类似, 量子计算系统很有潜力成为向特定领域的新型协处理器。近年来, 量子计算取得了实现量子霸权、展示量子纠错的可行性等突破性进展, 激发了更广泛研究群体对量子计算的兴趣。量子计算也在量子编程编译、量子噪声缓解、量子任务调度等体系结构相关的方面取得了可喜的进展。理解量子计算的基本原理和使用量子计算模拟器等基本工具是入门量子计算的有效方式。本教程包含四个环节:量子计算基础介绍、量子模拟器的使用介绍、量子编程框架的使用介绍和量子计算体系结构前沿进展介绍。本教程旨在向大家介绍量子计算并带领大家一起学习基础的量子计算基础工具的使用、了解量子计算体系结构前沿动态, 期望有助于引导大家开展量子计算体系结构相关的研究。

具体日程安排(论坛10)

时间	报告内容	报告嘉宾/单位
13:40 - 14:20	量子计算基础	吴俊杰/国防科技大学
14:20 - 15:00	量子计算模拟器的基本原理及使用	刘雍/国防科技大学
15:00 - 15:20	茶歇	
15:20 - 16:00	基于Quingo框架的量子-经典异构编程	付祥/国防科技大学
16:00 - 16:40	量子计算体系结构前沿进展	劳玲玲/西北工业大学
16:40 - 17:10	Panel: 量子计算基础问题讨论	

论坛主席:



刘英文
国防科技大学助理
研究员

刘英文, 博士, 国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室(HPCL)助理研究员, CCF体系结构专委会委员。从事量子计算方向研究, 参与研制银河鲲鹏QW2020量子计算系统、QUANTA VQS16量子计算系统等, 主要研究领域包括光量子芯片、量子光源等方面, 主持国家自然科学基金青年基金一项、参与国家自然科学基金重点基金项目, 在Optics Letters、Optics Express等期刊发表多篇学术论文。



刘雍
国防科技大学助理
研究员

刘雍, 博士, 国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室(HPCL)助理研究员, 2013年、2015年和2020年于国防科技大学分别取得学士、硕士和博士学位。研究领域包括量子计算高性能模拟、量子计算机体系结构、量子算法、光量子计算等, 在Physical Review Letters、National Science Review、Quantum Science and Technology等期刊上发表学术论文十余篇。

讲者报告详细内容

论坛讲者:



吴俊杰
国防科技大学
研究员

论坛讲者1

吴俊杰, 博士, 研究员、博士生导师, 现任国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室(HPCL)主任。中国计算机学会(CCF)杰出会员, CCF量子信息专业组常务委员, CCF体系结构专业委员会秘书长(2011-2019)、中国电子学会量子信息分会委员、全国量子计算与测量标准化技术委员会委员等。从事量子计算方向研究, 主持研制成功银河鲲鹏QW2020量子计算系统、QUANTA VQS16量子计算系统等, 在量子霸权标准、光量子芯片等方面的研究成果多次登上科技日报等国内外科技媒体。在Science Advances、National Science Review、Physical Review Letters等发表论文, 获湖南省自然科学一等奖等。

报告题目: 量子计算基础**报告摘要:**

薛定谔的猫、量子纠缠,听起来就很神秘;量子加速、量子霸权,听起来就很神奇。到底什么是量子?什么是量子计算?量子计算机会取代电子计算机吗?本次报告将深入浅出地介绍量子物理和量子计算的原理与基础,带你走进既神秘又神奇的量子世界!



刘雍
国防科技大学
助理研究员

论坛讲者2

刘雍,博士,国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室(HPCL)助理研究员,2013年、2015年和2020年于国防科技大学分别取得学士、硕士和博士学位。研究领域包括量子计算高性能模拟、量子计算机体系结构、量子算法、光量子计算等,在Physical Review Letters、National Science Review、Quantum Science and Technology等期刊上发表学术论文十余篇。

报告题目: 量子计算模拟器的基本原理及使用**报告摘要:**

量子计算模拟器是当前开发量子算法必不可少的工具之一,它的基本原理是什么?又是如何实现量子-经典混合算法的模拟呢?除了刻画量子算法中量子态的演化行为,能否进一步描述量子计算系统的行为呢?本报告我们将一起探讨这些问题。



付祥
国防科技大学
副研究员

论坛讲者3

付祥,博士,国防科技大学计算机学院量子信息研究所兼高性能计算国家重点实验室(HPCL)副研究员。2011年本科毕业于清华大学电子工程系,2014年硕士毕业于国防科学技术大学计算机学院,2018年底博士毕业于荷兰代尔夫特理工大学(研究方向量子控制体系结构)。付祥的主要研究目标是以可编程的方式控制量子芯片,代表工作包括第一个量子控制微体系结构QuMA、可执行的量子计算指令集eQASM和全面的量子-经典异构编程框架Quingo,曾获体系结构顶级会议MICRO-50最佳论文奖,并入选2017体系结构会议论文最佳选录(Top Picks'17)。付祥最近的研究兴趣集中于量子编程编译和量子测控软硬件系统等。

报告题目: 基于Quingo框架的量子-经典异构编程**报告摘要:**

量子程序设计语言是控制量子计算机的入口。那么量子程序长啥样,我们该如何进行量子编程?量子程序的描述与经典程序又有何不同?下载安装Quingo量子程序设计语言,实现并在模拟器或真机上运行你的第一个量子算法吧!



劳玲玲
西北工业大学
教授

论坛讲者4

劳玲玲,博士,西北工业大学计算机学院教授、硕士生导师。于2012年获得哈尔滨工业大学学士学位,2014年获得西北工业大学硕士学位,2019年获得荷兰代尔夫特理工大学博士学位。2020年至2022年于英国伦敦大学学院进行博士后研究。主要研究方向为容错量子计算,量子纠错,量子计算机体系结构,量子编译等。发表论文10余篇,包括2篇CCF A类会议ISCA、1篇CCF A类期刊TCAD、2篇中国科学院一区TOP期刊QST等。

报告题目: 量子计算体系结构前沿进展**报告摘要:**

量子计算机的实现需要量子算法、量子处理器和量子计算控制系统等多个技术的协同。然而在量子计算机上实现量子算法存在很多技术难题,例如量子比特状态易坍塌、量子逻辑门错误率高、量子算法与量子处理器拓扑结构不兼容等。为解决以上问题,一系列量子计算机体系结构方法被提出,包括量子编译技术、量子指令集和微体系结构、量子纠错和容错计算机理论。这些方法可以有效地减少量子电路门数量、缩短算法执行时间、增强量子噪声消除能力,从而提高量子计算的保真度。本报告将从计算机领域的角度介绍量子计算体系结构的组成和前沿发展,探讨量子与经典计算机体系结构的相似与不同之处。

分论坛日程

论坛11:新一代数据中心网络

时间:2023年8月5日13:30 - 17:30

论坛简介:

为贯彻落实国家建设新型数据中心的战略部署,论坛聚焦数据中心网络的发展趋势和新技术,为国内外工业界和学术界搭建高端对话交流平台,汇集数据中心网络与系统领域的顶级专家学者就新一代数据中心网络热点议题开展主题报告与思辨讨论,分享与探讨最新的技术研究成果和实践经验,进一步推动数据中心网络研究领域的创新和进步。

具体日程安排(论坛11)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	高效可扩展的数据中心RDMA网络	陈凯/香港科技大学
14:00 - 14:30	微软云的下一代数据中心网络创新与实践	熊勇强/微软亚洲研究院
14:30 - 15:00	分布式MoE模型训练和推理的加速方法	徐宏/香港中文大学
15:00 - 15:20	茶歇	
15:20 - 15:50	AI DCN技术发展趋势	郝卫国/华为
15:50 - 16:20	高性能数据中心网络拥塞控制和故障管理	张娇/北京邮电大学
16:20 - 16:50	面向AI的新型数据中心智算网络体系	翟恩南/阿里云
16:50 - 17:30	Panel 面向大模型场景的数据中心网络技术挑战与发展趋势	

论坛主席:



黄家玮
中南大学

黄家玮教授现为中南大学计算机学院教授,主要从事计算机网络协议与算法、数据中心、机器学习、流媒体领域的研究,主持国家自然科学基金重点/面上项目、湖南省重点研发计划、华为企业合作项目等研究项目,在TON、TCOM、TC、TCC、TMM、软件学报等期刊和INFOCOM、EuroSys、ICNP、ICDCS等国际会议上发表100多篇论文,获得CCF技术发明二等奖等奖励。



崔勇
清华大学

崔勇现任清华大学长聘教授,教育部长江学者特聘教授,任互联网体系结构全国重点实验室副主任,主要研究计算机网络体系结构、下一代互联网和网络安全,曾获首届青年长江学者、国家优秀青年科学基金、教育部新世纪人才和中创软件人才奖,先后担任中国通信标准化协会理事、国际互联网标准化组织 IETF IPv6 过渡工作组主席,IEEE TPDS、IEEE TCC、IEEE Network 及 IEEE Internet Computing 期刊编委。

讲者报告详细内容

论坛讲者:



陈凯
香港科技大学

论坛讲者1

香港科技大学教授、iSING实验室主任, 主要从事数据中心网络、高性能网络、机器学习系统和硬件加速的研究, 在SIGCOMM、NSDI、OSDI和TON等多个顶级期刊和会议发表论文100余篇, 担任APNet的指导委员会主席, SIGCOMM、NSDI、CoNEXT 等会议程序委员会成员。

报告题目: 高效可扩展的数据中心RDMA网络

报告摘要:

随着数据中心网络带宽的不断增长, 传统的TCP/IP软件协议栈已经成为通信瓶颈。因此, 远程数据直接内存访问RDMA现在已经成为了数据中心网络的主流通信机制。然而, 由于RDMA依赖于无损网络传输, 在有损数据中心网络中RDMA面临着效率和扩展性的严峻挑战。

在本次报告中, 我将介绍: (1) 过去10年来本领域为应对这些挑战所做的研究工作; (2) 讲解我们在港科大iSING实验室为数据中心建立可扩展和高效的RDMA网络所做的努力。



熊勇强
微软亚洲研究院

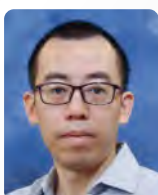
论坛讲者2

担任微软亚洲研究院网络研究组首席研究员。主要从事系统和网络以及网络安全的研究, 包括互联网路由协议、移动自组织网络和对等网络, 专注于数据中心网络系统, 尤其是架构设计、优化调度问题、交换机构建方法, 以及DDoS攻击等安全问题, 在SIGCOMM、NSDI等会议与期刊发表论文80余篇。

报告题目: 微软云下一代数据中心网络的创新与实践

报告摘要:

在云计算不断崭露头角的背景下, 微软正致力于推动下一代数据中心网络技术的创新与技术落地, 以满足客户不断增长的性能需求。在下一代云数据中心的创新和实践中, 微软公司和SONiC社区紧密合作, 推出了Disaggregated API for SONiC Hosts (DASH), 它第一次从云供应商的角度定义了云应用在可编程硬件上的业务抽象, 并提供了行为模型和测试用例保证云业务的一致性和性能要求。同时, 微软正在以DASH为基础, 搭建下一代DASH智能交换机, 这种智能交换机融合了交换机芯片和DPUs, 可以直接和SONiC软件栈结合到一起, 直接部署在数据中心的不同位置, 为云数据中心提供加速服务。



徐宏
香港中文大学

论坛讲者3

香港中文大学副教授, 主要从事计算机网络和系统, 尤其是大数据系统和数据中心网络的研究。在SIGCOMM、ATC、DAC、EuroSys和TON等多个顶级期刊和会议发表论文100余篇。获得了ACM SIGCOMM 2022和IEEE ICNP 2015的最佳论文奖。

报告题目: 分布式MoE模型训练和推理的加速方法

报告摘要:

分布式MoE模型训练采用数据和专家并行方式, 通过集合通信缓解了通信瓶颈问题。本报告将介绍一个提升MoE训练效率的系统Lina。该系统将基于优先级的微操作通信调度与流水线驱动的专家打包相结合, 并采用了一种基于量子分区的新调度方案, 有效地减少了通信数据量。



郝卫国
华为

论坛讲者4

华为数据通信战略与业务发展专家,东南大学硕士,2004年加入华为,从事过Bras、网络操作系统、sdn控制器、sd-wan、DCN产品和解决方案等研发工作,在IETF发布6篇RFC和多篇工作组草案,拥有60篇左右专利。目前从事数据通信产业与业务发展工作,携手学术界,共同定义Net5.5G产业代际。

报告题目: AI DCN技术发展趋势

报告摘要:

随着AI大模型时代到来, AI数据中心网络又面临一次重大技术架构变革。为提高AI算力集群并行计算效率, 对网络带宽、负载均衡、拥塞控制、计算与网络协同等提出了新的要求。在新的浪潮中, 有全栈IB和开放解耦以太之争, 有主流芯片和设备厂商的新一轮军备竞赛。本次报告会对本轮技术趋势和产业界面临的挑战做一次初步的总结交流。



张娇
北京邮电大学

论坛讲者5

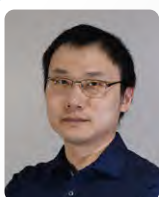
北京邮电大学教授, IEEE/CCF 高级会员, 2014 年获得清华大学博士学位, 曾于加州大学伯克利分校联培一年。长期致力于云数据中心网络、网络传输协议研究。在 ACM SIGCOMM、USENIX NSDI、ACM EuroSys、IEEE/ACM Trans. On Networking 等网络领域重要会议 / 期刊发表学术论文 70 余篇, 授权国家发明专利 30 余项, 专利转化 5 项。主持国家重点研发计划(青年科学家项目)、国家自然科学基金面上 / 青年基金和华为、字节跳动创新基金等 20 余项项目。担任 ACM CoNext、IEEE/ACM IWQoS 重要网络国际会议分项联合主席及 Multimedia、INFOCOM、ICPP 等技术委员会委员。获中国计算机学会优秀博士学位论文奖、北京市优秀博士毕业生、首届中国科协青年人才托举工程、中国通信学会技术发明一等奖和中国电子学会创新团队奖等荣誉。

报告题目: 高性能数据中心网络拥塞控制和故障管理

报告摘要:

以大模型为代表的人工智能发展对数据中心算力基础设施提出了更高的要求。随着数据中心节点间带宽的不断增长, 高速端侧数据处理给CPU带来巨大计算开销。因此, 近几年, RDMA网卡在数据中心网络中开始大规模使用, 以释放CPU通用算力资源。然而, 利用RDMA网卡处理高速网络数据也带来了新的挑战和问题。

在本次报告中, 我将介绍:RDMA在以太网中的应用背景和关键技术, 并围绕其中两个关键问题展开:高速RDMA网络拥塞控制和RDMA端侧故障瓶颈定位。介绍如何设计新型RDMA网络拥塞控制以及通用RDMA端侧瓶颈定位系统, 来保障RDMA网络的低时延、大吞吐和高可靠性。



翟恩南
阿里云

论坛讲者6

阿里云网络研究负责人, 阿里云资深技术专家。2015年于耶鲁大学计算机系获博士学位, 随后担任耶鲁大学研究型助理教授兼讲师, 2018年加入阿里巴巴。研究领域包括计算机网络、分布式系统、程序验证等, 先后在这些方向的国际顶级会议如SIGCOMM、OSDI、NSDI等累计发表30+篇论文。担任SIGCOMM、NSDI等国际顶级会议程序委员会委员。曾获通信学会技术发明一等奖一次。

报告题目: 面向AI的新型数据中心智算网络体系

报告摘要:

云计算通过数据中心网络将算力规模化, 从而满足服务对算力的需求。然而, 随着AI、深度学习等算力密集型技术的飞速发展, 云计算的算力遇到了前所未有的挑战。作为算力瓶颈的网络性能, 急需设计研发新型的技术进行网络加速。

本次报告主要介绍:(1)阿里云自主研发的面向AI的新型数据中心网络体系——服务器端/应用和网络融合感知的新型体系;(2)该体系下解决的关键技术挑战与解决这些挑战阿里云所研发的关键技术(主要包括拥塞控制、路径选择以及硬件编程等), 这一系列技术发表在 SIGCOMM 和 NSDI 等国际顶级会议上。

分论坛日程

论坛12:AI大模型应用与基础系统创新

时间:2023年8月5日13:30 - 17:30

论坛简介:

本次“AI大模型应用与基础系统创新”学术论坛旨在探讨人工智能大模型在应用中的创新与突破,以及基础系统的关键技术研究。论坛将邀请国内外的专家学者分享他们在人工智能领域的最新研究成果和经验,并就该领域的前沿技术进行深入探讨和交流。

具体日程安排(论坛12)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	昇思MindSpore技术创新进展和超大规模AI实践	苏腾/华为技术有限公司
14:00 - 14:30	图算融合的人工智能芯片编译器	翟季冬/清华大学
14:30 - 15:00	架构感知的稀疏张量运算AI编译插件	蒋力/上海交通大学
15:00 - 15:30	茶歇	
15:30 - 16:00	银河天璇多维度并行智能训练框架(Merak)	赖志权/国防科技大学
16:00 - 16:30	人工智能大模型四维并行训练系统(MindPipe)	崔鹤鸣/香港大学
16:30 - 17:00	面向大模型的低精度自适应数值类型	冷静文/上海交通大学
17:00 - 17:45	Panel:可持续的大模型之路	

论坛主席:



金雪峰
华为

金雪峰,昇思MindSpore首席架构师,华为2012实验室中央软件院架构与设计管理部部长,当前主要的方向MLSys,包括大规模机器学习系统、AI编译器、AI+科学计算系统等,之前曾先后担任华为的分布式数据库/大数据平台、电信基础软件平台的技术负责人,有20年系统软件设计和开发经验。



李诚
中科大

李诚,德国马普学会软件系统所(MPI-SWS)博士,中国科大计算机学院/国家高性能计算中心(合肥)副教授,博导,FGS、CCF THPC期刊青年编委。聚焦人工智能基础系统软件研究,在SOSP、OSDI、EuroSys、ATC、FAST、ASPLOS、SC、HPCA等计算机系统领域著名国际会议上发表论文40多篇。2019年入选ACM FCA成员。曾担任第14届/第21届ChinaSys程序共同主席、SOSP 2017 学术海报程序共同主席、EuroSys 2021/ACM SIGMETRICS 2023论文出版共同主席、首届CCF计算机系统大会/芯片大会宣传主席等,长期参与SOSP、FAST、Middleware、DSN、ICDCS、SRDS等系统领域著名国际会议的程序委员会。获2023世界人工智能大会云帆奖璀璨明星、2022 AI 2000 Most Influential Scholar Honorable Mention in Computer Systems、2022年CCF分布式专委会杰出青年学者、2021年ACM ChinaSys新星、2021年ACM中国新星提名等科研奖项。主讲《编译原理和技术》课程入选第二批国家级线下一流课程,获安徽省第五届青教赛工科组一等奖、第四届中国计算机教育大会计算机类教学资源建设特等奖(2项)、全国高校教学创新大赛安徽省二等奖等教学奖项。

讲者报告详细内容

论坛讲者:



苏腾
华为

论坛讲者1

苏腾, 浙江大学博士, 华为昇思MindSpore架构师, MindSpore超大规模AI技术负责人, 近两年孵化了20+大模型, 其中6个千亿参数规模以上的大模型。加入华为10年, 长期从事大规模分布式并行基础软件研发, 现主导MindSpore超大规模AI系统构建, 曾主导华为自研的分布式任务计算框架、分布式强一致性K-V存储框架、容器化资源管理与调度框架、Actor模型编程框架的设计开发, 并在多个产品落地应用, 在大规模分布式系统方向有丰富的实践经验。

报告题目: 昇思MindSpore技术创新进展和超大规模AI实践

报告摘要:

以大模型、大数据、大算力驱动的超大规模AI正成为下一代人工智能的突破口, 也是迈向通用强人工智能最有潜力的技术方向之一。昇思MindSpore构建了多维度混合并行、多维度内存优化、图算融合等关键技术, 支撑了20+大模型的训练。并在2021年联合高校及科研院所协同创新, 发布了鹏程盘古NLP大模型、中国科学院紫东太初图文音多模态大模型、鹏程神农蛋白质结构大模型、武大珞珈遥感大模型等。本次分享分享基于昇思MindSpore的大模型的最新进展和成果, 介绍昇思MindSpore超大规模AI训练、微调、压缩、部署全流程关键技术, 如何使用这些关键技术昇腾大集群上训练大模型, 以及昇思MindSpore上的大模型如何助力产业发展。



翟季冬
清华大学

论坛讲者2

翟季冬, 清华大学计算机系长聘教授, 博士生导师。现为清华大学计算机系高性能所副所长, ACM中国高性能计算专家委员会秘书长。主要研究方向包括高性能计算、性能评测和编译优化等。研究成果发表在相关领域顶级学术会议和期刊——SC、ICS、PPOPP、ASPLOS、MICRO、OSDI、ATC、IEEE TC、IEEE TPDS等。研究成果获IEEE TPDS 2021最佳论文奖、IEEE CLUSTER 2021最佳论文奖、ACM ICS 2021最佳学生论文奖等。担任NPC 2018程序委员会主席、IEEE Cluster 2021领域主席、SC 2022领域副主席, SC、ICS、PPOPP、PACT等国际学术会议程序委员会委员。目前担任IEEE TC、IEEE TCC等多个国际学术期刊编委。担任清华大学学生超算团队教练, 指导的团队十三次获得世界冠军。获教育部科技进步一等奖、中国计算机学会自然科学一等奖、CCF-IEEE CS青年科学家奖、国家杰出青年科学基金。

报告题目: 图算融合的人工智能芯片编译器

报告摘要:

随着人工智能领域对算力需求的持续增长, 人工智能芯片成为当前处理器发展的一个主流趋势。为了满足对算力的巨大需求, 硬件公司推出了各种人工智能芯片, 例如寒武纪Cambricon、华为昇腾、阿里巴巴含光、百度昆仑等。如何开展面向人工智能芯片的自动编译技术对推动我国人工智能芯片的发展具有重要意义。本报告讨论当前在人工智能芯片上实现自动编译面临的主要挑战以及最新的研究进展。



蒋力
上海交通大学

论坛讲者3

蒋力, 上海交通大学计算机科学与工程系副教授, 博导。华为科学家, 异构通信处理器研究高级专家。研究方向为计算机体系架构与设计自动化(EDA)。在计算机体系结构及芯片设计自动化领域的顶级/主流国际会议和期刊, 如ISCA、MICRO、DAC、DATE、TCAD、TC, 以及AI领域顶级会议, 如AAAI、ICCV、SIGIR等上发表论文90余篇, 获得2次DATE最佳论文奖, 及多次最佳论文提名。曾获ACM上海新星奖, CCF集成电路Early Career Award, 吴文俊人工智能奖(芯片类)二等奖。曾担任CCF YOCSEF上海主席, CFTC2021程序委员会主席。

报告题目: 架构感知的稀疏张量运算AI编译插件

报告摘要:

深度学习模型(包括大模型)都是极度稀疏的。为了利用这种稀疏性,学术界研究了诸多支持稀疏张量运算的硬件加速机制。由于其设计复杂性,在当前的GPGPU和AI处理器上鲜有落地商用。这次报告介绍一种将稀疏张量运算转化为稠密张量运算的编译插件,根据芯片计算调度多层次的粒度充分利用稀疏计算中的局部性完成算子拆分/融合/调度优化。不仅可以直接作用于无稀疏性支持的AI处理,而且可以便捷地嵌入到常见AI编译框架中。这一稀疏张量运算编译插件嵌入MindSpore后,针对稀疏CNN, Transformer模型中几种业界常用的稀疏pattern,可以在V100 GPU与Ascend 910加速卡上分别获得1.93x和2.21x性能提升。



赖志权
国防科技大学

论坛讲者4

赖志权,国防科技大学计算机学院副研究员,主要研究方向为高性能计算、并行智能计算、分布式图计算等,国家重点研发计划青年科学家项目负责人,银河、天河系列高性能计算系统技术骨干,参与国家重点研发计划、装备预先研究、国家“核高基”研究计划等项目10余项。研究成果在IEEE TPDS、IEEE TC、ICPP、CLUSTER 相关领域国际会议和期刊发表论文30余篇,获军事科学技术进步一等奖1项。

报告题目: 银河天璇多维度并行智能训练框架(Merak)**报告摘要:**

随着智能计算技术的飞速发展,以GPT-3为代表的超大规模智能模型的参数量达到千亿、万亿规模,对智能计算框架的可编程性和并行训练效率提出严峻考验。TensorFlow、PyTorch等基础深度学习框架已不适应大规模智能训练的可编程性和扩展性需求,并行智能计算框架成为国内外框架生态竞争的新焦点。为了解决大规模智能模型训练内存开销大、并行训练难等问题,国内外研究机构推出了微软DeepSpeed、英伟达Megatron-LM等并行智能计算框架,但仍然存在并行训练编程难、并行效率低等技术挑战。本报告将围绕大规模智能模型训练问题展开,介绍银河天璇多维度并行智能训练框架的设计思路和关键技术,以及该框架的未来发展规划。



崔鹤鸣
香港大学

论坛讲者5

崔鹤鸣博士是香港大学计算机科学学院的副教授。崔博士专注于构建构建软件系统和工具,以极大地提高现实世界应用软件的可靠性、安全性和性能很感兴趣。2015年在哥伦比亚大学获得博士学位后,崔博士加入港大作为助理教授、博士生导师,独立组建了一个并行分布式系统团队,目前在读全日制博士生约20人,主要来自国内最著名几个大学的计算机本科毕业尖子生。他最近的研究促成了一系列开源项目,并在广泛领域的国际顶级会议和期刊上发表了文章,包括SOSP、TPDS、MICRO、NSDI、ASPLOS、ATC、ICSE、EuroSys。近三年,崔博士至少担任一次国际顶级系统/网络会议的程序委员会委员,包括OSDI、SIGCOMM、ASPLOS、NSDI、ATC、EuroSys。他担任ACM APSys 2016和2021的主席,以及ACM ChinaSys 2023的程序主席。作为研究项目负责人,他的研究项目共获得约1亿港元的资助,包括国家科技部(重点研发计划)、香港研究资助局(例如2023年的RGC研究影响基金)、香港创新科技署和裘槎基金会。崔博士最近的安全系统论文(例如[Uranus AsiaCCS 2020]和[Cronus MICRO 2022])及其成果的Ubiquitous Trusted Execution Environments(UTEE)项目已成为华为可信智能云服务的核心软件系统。崔博士于2021年开始研发的并行大AI模型训练系统MindPipe(论文发表于[Fold3D TPDS 2023]、[NASPipe ASPLOS 2022]、[vPipe TPDS 2021]),已经在PyTorch库和Nvidia的GPU上实现,支持通用的大型AI模型(例如,Transformer、GPT、CPM和Pan-Gu);他的系统也被集成到华为的MindSpore主分支中,作为其大模型并行训练系统。崔博士在清华大学获得学士和硕士学位,在哥伦比亚大学获得博士学位;均为计算机专业。

报告题目: 人工智能大模型四维并行训练系统(MindPipe)**报告摘要:**

大模型(例如Transformer和GPT)不断增强的建模能力在各个AI领域取得了前所未有的成功,包括理解视觉和自然语言。大模型的高建模能力主要源于其不断增加的复杂性(每层具有更多的神经元层和更多的神经元运算符)和动态性(在训练期间经常激活/停用每一层中的神经元运算符,例如神经架构搜索或NAS)。这种复杂性和动态性很容易使大型DNN超过

现代 GPU 的计算和内存容量, 因此训练大型 DNN 往往需要通过多个维度将 DNN 拆分为多个 GPU, 包括数据并行、张量并行和流水线并行。崔博士的演讲将介绍他最近的论文 (例如 [Fold3D TPDS 2023]、[NASPipe ASPLOS 2022] 和 [vPipe TPDS 2021]), 这些论文解决了现有多维并行训练系统的主要局限性, 包括 GPipe、Pipedream、和 Megatron。例如, vPipe 专注于解决严重的负载不平衡和低 GPU 计算利用率; NASPipe 将展示超网并行性, 这是一种新的并行训练维度, 适用于以超网方式设计的高度动态的大型 DNN (例如, 进化的变压器和神经架构搜索)。



论坛讲者6

冷静文为上海交通大学计算机科学与技术系长聘副教授, 主要研究方向为面向人工智能的新型计算系统的设计以及性能、能效、可靠性优化, 主持和参与了多项自然科学基金以及龙头企业横向项目。在国际一流的会议和期刊上发表了四十多篇论文和相关国内国际专利, 获得过DAC, PACT等多个国际会议的最佳论文提名奖。

冷静文
上海交通大学

报告题目: 面向大模型的低精度自适应数值类型

报告摘要:

大模型的算力需求两年内增长了240倍, 远超遵循摩尔定律的芯片制程所带来的提升。因此, 计算架构的演进以及计算数值格式的革新成为了计算效率提升的关键。在本次分享中, 我们深入分析了大模型中的数值分布特征, 并设计了能够自动调整数值类型以贴近大模型中不同的张量分布, 从而将大模型的计算数值位宽缩减到了4比特。在此基础上, 我们也为自适应数值设计了对于存储子系统友好的编码格式, 能够以极低的硬件开销发挥自适应数值类型的计算效率优势。

分论坛日程

论坛13:安全计算系统论坛

时间:2023年8月5日13:30 - 17:30

论坛简介:

近些年, 计算机安全问题频发引发人们对计算机系统安全的重点关注, 计算系统安全已经成为学术界及工业界的热点研究话题。不同于网络安全, 计算系统安全更关注底层系统软件和体系结构的安全问题。由于这些基础软硬件是支撑业务运行的基础, 因此其安全问题危害高, 解决难度大。本论坛旨在汇聚国内系统安全学术界与工业界相关的最新研究成果, 探讨漏洞挖掘、漏洞利用、软硬协同设计、机密计算等领域内热点技术话题, 以及领域未来的发展方向。

具体日程安排(论坛13)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	多版本Linux内核的漏洞可利用性评估	张源/复旦大学
14:00 - 14:30	软硬件协同的系统软件漏洞防护	张超/清华大学
14:30 - 15:00	DuVisor:软硬协同的用户态Hypervisor	糜泽羽/上海交通大学
15:00 - 15:20	茶歇	
15:20 - 15:50	Arm机密计算架构(CCA)	张锋巍/南方科技大学
15:50 - 16:20	可信安全基础设施技术和实践	闫守孟/蚂蚁集团
16:20 - 16:50	安全韧性虚拟化系统技术	谭寅/华为
16:50 - 17:30	Panel: 安全计算系统的技术演进与未来趋势	

论坛主席:



王喆
中国科学院计算
技术研究所

王喆, 中国科学院计算技术研究所处理器芯片全国重点实验室副研究员。长期从事系统安全、操作系统和虚拟化的研究, 在漏洞挖掘、漏洞利用、程序分析和安全防护等方面取得了一系列的科研成果。在包括IEEE Security and Privacy、ACM CCS、USENIX Security、USENIX ATC、ACM ISSTA、TDSC和TSE等安全、软工和系统等领域的国际会议和期刊上发表/录用CCF-A类论文10余篇。现担任CCF THPC期刊的青年编委, 曾担任国内外著名会议和期刊的审稿人。主持1项国家自然科学基金青年项目, 作为分课题负责人参与国防装备预研项目1项。同时作为核心人员, 参与多项国家863计划、国家自然科学基金重点项目、北京市科委重大专项、国防装备研发计划等。曾获2022年安全顶会CCS最佳论文提名奖, 中国科学院大学“领雁”金奖和银奖、中国科学院计算技术研究所“卓越之星”、中国科学院计算技术研究所“新学术百星”等奖项和称号。



申文博
浙江大学

申文博, 浙江大学百人计划研究员, 博士生导师, 浙江大学计算机科学与工程系副主任。研究方向为操作系统安全, 云原生系统安全, 软件供应链安全。在计算机安全、系统、网络国际顶级会议、期刊发表论文 30 余篇, 获得 3 项杰出论文奖(NDSS 16, AsiaCCS 17, ACSAC 22)。主持国家自然科学基金、科技部重点研发课题等多项科研项目。研究成果被应用于保护超过亿部设备系统安全。

讲者报告详细内容

论坛讲者:



张源
复旦大学

论坛讲者1

复旦大学计算机科学技术学院副教授、博导, 白泽战队指导教师。主要研究方向为软件安全、程序分析、安全攻防技术。研究工作主要发表于网络安全顶会, 获得 ACM CCS 2018 亮点论文、ACM CCS 2020 最佳论文提名、USENIX Security 2022 杰出论文奖等荣誉。带领白泽战队获得国内外顶尖安全攻防赛事冠军 20 余次, 包括全国大学生信息安全竞赛创新实践能力赛 2020 年 /2021 年 /2022 年的三连冠, 发现的多个软硬件高危漏洞得到华为、腾讯、蚂蚁金服、百度、谷歌等公司致谢, 获得 2021 年国家信息安全漏洞库最具价值漏洞奖。

报告题目: 多版本Linux内核的漏洞可利用性评估

报告摘要:

Linux内核由于庞大的代码空间、重要的安全价值, 有大量的安全漏洞被发现, 对内核漏洞的治理带来巨大挑战。评估Linux漏洞的可利用性有助于内核开发者评估漏洞补丁的开发优先级, 也有助于Linux下游厂商评估漏洞补丁的部署优先级。然而, 由于Linux内核多分支版本之间、上下游版本之间存在显著的代码差异, 针对一个版本的漏洞利用程序难以有效评估该漏洞在其他内核版本的可利用性。本次报告通过深入分析现有生成式漏洞可利用性评估方法的不足, 分享一种新的基于迁移式的漏洞可利用性评估方法。



张超
清华大学

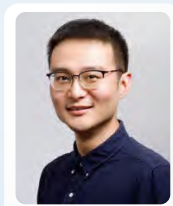
论坛讲者2

清华大学长聘副教授，华为冠名教授，蓝莲花战队教练，CCF杰出会员。曾获得清华大学学术新人奖、国家级青年人才、MIT TR35 China、求是杰出青年学者等奖励和荣誉。主要研究软件和系统安全，尤其是智能攻防方向。

报告题目：软硬件协同的系统软件漏洞防护

报告摘要：

软件漏洞是网络攻击的突破口，具有高危害、高价值、高增长、难根治的特点。攻击者可以利用未知的0day漏洞突破系统，实施对传统防护手段的“降维打击”，实现各种攻击目的。如何防范未知漏洞，是业界广受关注的难题。近年来，研究人员提出了软硬件协同的解决方案，通过扩展处理器指令集提供硬件特性，并扩展编译器实现软件加固，在软件运行过程中进行安全状态自检，阻断漏洞攻击路径实现“内生安全”。本次报告将分享团队利用现有硬件特性以及自主设计硬件特性实现的多个软硬件协同防护方案。



糜泽羽
上海交通大学

论坛讲者3

糜泽羽，上海交通大学助理研究员，硕士生导师，主要从事操作系统和系统虚拟化的研究，作为负责人和技术骨干主持和参与国家自然科学基金青年科学基金项目、重点项目、科技部重点研发计划等国家级科研项目。在SOSP、OSDI、ASPLOS、EuroSys、USENIX Security、IEEE TC等操作系统顶级会议和期刊发表十余篇学术论文。

报告题目：DuVisor:软硬协同的用户态Hypervisor

报告摘要：

目前主流的系统虚拟化系统由两个部件组成：一个是使用硬件虚拟化的内核驻留程序，一个是提供虚拟机管理和I/O虚拟化的用户态进程。然而，这种虚拟化架构在安全性和性能方面都存在问题。此前的相关工作通过将内核功能卸载到用户态的方式最小化内核的部分，这种设计却面临着安全和性能之间的权衡：虽然更多的卸载可能会减少内核的攻击面，但却增加了内核态与用户态之间的模式切换，从而增加了虚拟化系统的性能开销。本次报告将介绍一种软硬件结合的新型虚拟化系统设计，它将内核驻留程序参与的控制平面与用户态进程的数据平面完全分开，从而消除了内核在虚拟机运行时的干预。基于这种设计而实现的用户态虚拟机监控器DuVisor，可以直接在用户态处理几乎所有的虚拟机下陷，包括在用户态配置虚拟机寄存器，管理第二阶段页表并实现高效的I/O虚拟化。我们已经在开源的RISC-V CPU上实现了硬件扩展，并构建了基于Rust的虚拟机监控器。在FireSim平台上的评估数据显示，DuVisor给应用带来的开销低于5%。



张锋巍
南方科技大学

论坛讲者4

南方科技大学计算机科学与工程系指南针（COMPASS）实验室负责人，副教授，正高级研究员，博导，IEEE和CCF高级会员。在加入南科大之前，担任美国韦恩州立大学计算机系助理教授（2015-2019），主要研究领域是系统安全和硬件辅助安全。在计算机安全“四大”顶会上都担任过PC和发表过论文，研究工作曾获批多项美国自然科学基金项目，以及2021年国家级青年人才项目。

报告题目：Arm机密计算架构(CCA)

报告摘要:

随着各行业计算技术的发展,以及云端和终端用户的增加,人们将大量数据存储在各种计算机设备中。在行业发展的同时,人们对于设备和数据安全的关注也在日益增加。TrustZone 是一种有望用于将机密数据分割到可信执行环境(TEE)中的安全技术。不幸的是,第三方开发者对 TrustZone 的可访问性受到限制。这是因为 TEE 供应商需要严格验证这些可信应用程序以保持其安全性。此外,基于 TrustZone 的系统随着 TCB 的增大,出现了影响 TEE 软件系统的漏洞,可能导致整个系统被攻击。

最近 Arm 提出了机密计算架构(CCA),CCA 创建了一个名为 Realm world 的新物理地址空间,用于保护第三方用户数据的机密性和完整性。CCA 的当前版本主要基于 Realm world 虚拟化技术进行隔离,并没有提供用户级别的隔离环境。我们提出了 Shelter,旨在构建一个在 Normal world 非安全世界的用户态 TEE;它也是 CCA Realm VM 架构的一个扩展。Shelter 允许第三方开发者在非安全世界的用户态空间中,为用户态应用部署机密计算环境。Shelter 基于 CCA 机密计算架构的硬件特性实现对用户态可信应用的内存隔离管理,从而为用户态可信应用创建机密、隔离的计算环境,并消除了对主机操作系统、虚拟机监控程序或特权软件(例如 Secure OS 或 Realm 虚拟机监控程序)的信任需求。我们已经实施和评估了 Shelter,结果表明 Shelter 能够在实际工作负载下以适度的性能开销保证应用程序的安全性。



闫守孟
蚂蚁集团

论坛讲者5

蚂蚁集团研究员、资深总监,可信安全计算部负责人和技术研究院计算系统实验室主任。他的研究兴趣聚焦于操作系统、系统安全、和软硬协同计算技术。他带领团队开源了业界最受欢迎的 TEE LibOS——Occlum,并研发了信任根与CPU厂商解耦的HyperEnclave国产TEE。Occlum和HyperEnclave已发表论文,并在蚂蚁、阿里巴巴、微软Azure、以及诸多隐私计算企业得到广泛应用。加入蚂蚁之前,在Intel从事基础技术研究,为Intel产品贡献了多项成果。他发表了多篇顶会论文,并拥有30余件专利。他在西北工业大学获得计算机应用技术博士学位。

报告题目:可信安全基础设施技术和实践

报告摘要:

本报告将分享蚂蚁集团在可信安全基础设施方面的技术探索和实践经验,涉及可信硬件、安全容器、机密计算等全栈安全可信技术。



谭寅
华为

论坛讲者6

华为2012系统安全技术实验室主任。北京大学博士,新加坡南洋理工大学博士后,加拿大滑铁卢大学ECE系助理教授,研究方向密码学、可信执行环境、操作系统安全等。2015年入职华为加拿大研究所,先后负责自研操作系统AOS安全、区块链安全、数据安全与隐私保护技术、SGX等TEE技术等。

报告题目:安全韧性系统的关键技术研究的前沿实践

报告摘要:

系统安全已经取得了长足的进展,但是由于包括Linux内核在内的系统软件变得越来越复杂,攻击面持续上升,虽然大量的系统安全防护方法被工业界所采纳,还是常常有高危漏洞利用导致系统一触即溃。在这样一个背景下,面向ICT等关键基础设施、智能车等新产业的智能复杂系统、CPU+xPU+分布式的新型计算架构等,本报告抛砖引玉,探讨从源头上构筑高安和韧性的系统,把学术界提出的各类方法进行快速验证并应用到产品中的部分研究成果和实践。

分论坛日程

论坛14:分布式工业网络与计算论坛

时间:2023年8月5日13:30 - 17:30

论坛简介:

工业互联网是将传统制造业与互联网技术相融合,实现生产过程数字化、网络化、智能化的一种新型工业模式,也是我国数字经济发展的关键保障。目前,工业互联网研究在基础设施建设、标准制定、技术创新等方面已经取得了一定的进展,但是随着工业系统规模的扩大,工业互联网在朝着更加智能、高效、安全、可靠的方向发展过程中,面临着分布式网络架构设计、网络协议、数据质量、高效协同计算、网络安全等诸多问题与挑战。分布式工业网络与计算学术论坛旨在为工业互联网领域的学者、工程师和研究人员提供一个交流和合作的平台,分享最新的研究成果和技术进展,探讨未来的发展方向和挑战,促进学术界和产业界的合作与创新。

具体日程安排(论坛14)

时间	报告内容	报告嘉宾/单位
13:30 - 14:15	新一轮科技革命、产业变革与工业互联网	王兴伟/东北大学
14:15 - 15:00	星地协同计算赋能6G智能资源管控技术的思考	盛敏/西安电子科技大学
15:00 - 15:20	茶歇	
15:20 - 16:05	人工智能与数字孪生驱动的高效可靠工业互联网	闵革勇/英国埃克塞特大学
16:05 - 16:50	高性能计算服务化与工业应用	唐卓/湖南大学
16:50 - 17:30	Panel: (嘉宾:肖依/中山大学; 谢高岗/中科院计算机网络信息中心; 李哲涛/暨南大学)	

论坛主席:



姜楠
华东交通大学

姜楠,现任华东交通大学信息工程学院(人工智能学院)副院长、教授、博士生导师。享受省政府特殊津贴,入选江西省“双千计划”科技创新高端人才、省杰青、省主要学科学术和技术带头人、省百千万人才工程、省青年科学家(井冈之星)等人才计划和荣誉。担任20余个国际会议程序委员会主席、TPC Member、SCI期刊副主编和客座编辑,以及中国计算机学会物联网专委会、网络与数据通信专委会和中国铁道学会信息化委员会等多个专委会委员。以第一或通讯作者在IEEE INFOCOM、IEEE TKDE、ACM ToIT等国内外期刊会议上发表论文近百篇;以第一发明人申请发明专利21项,授权中国发明专利8项、美国发明专利1项,获软著7项,出版专著2部和教材3部,培养博士硕士30余人。主持20余项科研项目,包括国家自然科学基金5项、国家重点研发计划子课题2项等;获江西省自然科学奖二等奖1项(排名第一)、中国人工智能学会“吴文俊人工智能科学技术奖”科技进步奖二等奖1项(排名第一)和自然科学三等奖1项(排名第一)、中国铁道学会科学技术奖二等奖1项(排名第一)、詹天佑铁道科学技术奖青年奖等多项奖励。



邱铁
天津大学

邱铁,天津大学智能与计算学部教授,博士生导师,入选国家级青年人才称号、天津市杰出青年基金获得者,天津大学北洋学者计划、同年入选天津市131人才计划,ACM中国理事会天津分会秘书长;曾获天津市科技进步一等奖1项(排名第1)、大连市科学著作二等奖2项、ACM天津新星奖,中国计算机学会CCF杰出会员、IEEE/ACM高级会员、网络与数据通信专委会常委。曾在美国爱荷华州立大学作访问学者。主要从事物联网技术及大数据等相关研究工作。目前为国际期刊IEEE/ACM Transactions on Networking (ToN)等五个SCI检索国际期刊编委。作为多个国际会议的大会主席、程序委员会主席及程序委员会成员。编写并出版专著及专业教材10部;现已获得授权国家发明专利30余项,计算机软件著作权5项。已发表学术论文150余篇,部分论文发表在IEEE/ACM ToN, TMC, TKDE, JSAC, IEEE INFOCOM等国际著名期刊或会议上,其中18篇论文被列为ESI高被引论文。主持国家重点研发计划、国家自然科学基金重点项目、面上项目及各类项目20余项。

讲者报告详细内容

论坛讲者:



王兴伟
东北大学

论坛讲者1

王兴伟, 现任东北大学副校长, 博士生导师、国家杰出青年科学基金获得者、国务院政府特殊津贴获得者、教育部新世纪优秀人才、辽宁杰出科技工作者、兴辽计划特聘教授、全国工程专业学位研究生教育指导委员。主要研究方向为互联网、云计算和网络空间安全等。获国家科技进步二等奖2项、教育部科技进步一等奖2项、中国通信学会科学技术一等奖1项、教育部技术发明二等奖1项、辽宁省技术发明二等奖1项、湖南省自然科学二等奖1项; 在IEEE Trans等著名学术期刊和IEEE ICDCS等著名学术会议上发表论文100余篇, SCI收录100余篇; 出版学术著作9部; 取得国家发明专利授权27项; 国家级和省部级人才培养奖励20余项。

报告题目: 新一轮科技革命、产业变革与工业互联网

报告摘要:

本报告主要介绍工业互联网的由来、参考架构和层次结构以及带来的一系列挑战 and 对策, 并详细介绍工业互联网的发展。讲座内容主要包括讲述世界现代化与科技革命的关系、历次工业革命的特点及发展、中国制造2025的一条主线和四个转变、工业互联网的参考架构和层次结构。之后王从网络规模、网络异构、网络性能、网络安全以及应用的多样化、个性化与定制化等方面介绍了工业互联网带来的挑战与相应的对策。



盛敏
西安电子科技大学

论坛讲者2

盛敏, 西安电子科技大学教授、博士生导师。现为空天地一体化综合业务网全国重点实验室主任、国家科技部6G总体专家组成员; 国家自然科学基金创新研究群体负责人、科技部重点领域创新团队负责人; 教育部长江学者特聘教授, 国家自然科学基金委杰出青年基金获得者; 中国电子学会会士、中国通信学会会士。主要从事天地一体异构网络融合、无线自组织网络、空间信息网络等领域的研究工作; 获国家技术发明二等奖2项。

报告题目: 星地协同计算赋能6G智能资源管控技术的思考

报告摘要:

为满足6G全场景服务需求, 星地一体化已成为未来网络的重要形态。本次报告以星地协同计算为出发点, 探究星地计算资源如何高效使用, 实现“以网协算、以算强网”的智能资源控管方法, 从而助力未来6G网络资源控管的按需智能分配。



闵革勇
英国埃克塞特大学

论坛讲者3

英国埃克塞特大学计算机学科教授, 高性能计算与网络实验室主任。一直致力于计算机网络和分布式系统的研究, 在移动互联网、云计算与边缘计算、大数据、人工智能等领域取得了一系列创新性研究成果, 在国际重要学术期刊/会议发表论文共计380多篇。作为项目的总负责人或首席科学家主持过20多项重要科研和产业化项目。

报告题目: 人工智能与数字孪生驱动的高效可靠工业互联网

报告摘要:

工业互联网通过创建数字化、网络化、智能化的新一代工业系统, 对提升生产效率、降低运营成本及推进制造业的转型升级至关重要。随着规模的快速扩大, 工业互联网面临着如何有效处理海量大数据、管理大规模设备连接, 以及满足个性化和定制化应用需求等一系列挑战。本报告将介绍人工智能和数字孪生技术在工业互联网中的应用, 深入探讨如何设计高效的网络建模和数据分析方法、及分布式数据处理平台, 旨在提升工业互联网的故障检测和预测的效率和精度, 实现可靠的工业互联网。

**论坛讲者4**

唐卓, 湖南大学岳麓学者特聘教授、博士生导师。国家杰出青年科学基金获得者, 教育部青年长江学者, 湖南省杰出青年科学基金获得者。现任湖南大学信息科学与工程学院副院长(主持工作), 国家超级计算长沙中心总工程师, 算力网络湖南省工程研究中心主任, 湖南大学先进技术协同创新研究中心常务副主任。

唐卓
湖南大学

报告题目: 高性能计算服务化与工业应用

报告摘要:

工业信息物理系统(CPS)由设备层、感知层、网络层、认知层和控制层组成, 形成人、产品、物理空间和信息空间的深度融合, 是实现智能制造系统的基础。本报告将针对基于国家超算的高性能计算服务化平台的构建, 以及如何为工业应用建立基于数据采集、融合、分析、计算的CPS模型提供算力支撑。回答两个问题: 如何支持HPC资源的弹性管理与分配, 建立工业算网, 让用户能按需使用计算/存储/网络资源; 如何在超算上提供具有工业计算SAAS支撑能力的云原生应用环境, 让用户透明使用各类计算工具与函数库。

分论坛日程

论坛15: 教学论坛

时间: 2023年8月5日13:30 - 17:30

论坛简介:

图灵奖获得者John Hennessy和David Patterson共同发表文章《计算机体系的新黄金时代》, 期待着未来十年将会有一场新计算机体系结构的迅猛发展。近年来, 开源指令集、通用GPU、领域专用加速器等多项技术的成熟、应用推广和国产化, 吸引着越来越多的年轻人投身体系结构领域。这也给计算机体系结构方向的教学和人才培养提出了新的挑战: 教学内容、教学方法与模式、能力培养、实践环境条件等应进行怎样的变革? 本论坛的主题为“产学合作的计算机体系结构人才培养”, 邀请国内长期从事计算机体系结构教学和人才培养的学术界和工业界专家, 与国内各高校计算机体系结构方向的骨干教师汇聚一堂, 共同就以上关键问题展开探讨。

具体日程安排(论坛15)

时间	报告题目	讲者/单位
13:30 - 13:35	论坛及嘉宾介绍	
13:35 - 14:10	计算机体系结构一流课程建设	张晨曦/同济大学
14:10 - 14:45	变革实践教学体系, 强化系统能力培养	钟将/重庆大学
14:45 - 15:20	面向系统能力提升的计算机系统课程建设	王党辉/西北工业大学
15:20 - 15:45	立足自主创新, 培养致力于国产芯片的一流人才	张承义/飞腾信息技术有限公司
15:45 - 16:10	龙架构教育开源规划与进展	汪文祥/龙芯中科技术股份有限公司
16:10 - 16:35	“一生一芯”计划计算机系统能力培养	余子濠/中国科学院计算技术研究所
16:35 - 17:00	产学合作—CCF TCArch可定制计算挑战赛的人才培养尝试	王超/中国科学技术大学

论坛主席:



沈立
国防科技大学

CCF杰出会员, 体系结构专委会秘书长, 湖南省计算机学会常务理事。长期从事计算机体系结构方向的教学和科研工作, 是国家精品课/一流课程“计算机体系结构”、“计算机原理”的主讲教师, “计算机体系结构”湖南省优秀研究生导师团队成员。教学科研成果曾获得首届全国优秀教材一等奖、国家教学成果二等奖(2项)、湖南省教学成果一等奖、教育部自然科学二等奖、湖南省自然科学二等奖等奖励。



李涛
南开大学

CCF杰出会员、理事, 体系结构专委会副主任, 中国计算机学会通讯专题编委, CCF南开大学学生分会督导主任。长期从事计算机体系结构方向的教学和科研工作, 入选2019年度高校计算机专业优秀教师奖励计划和天津市教学名师奖后备人选, 南开大学智能物联网教学团队负责人, 计算机系统能力培养建设高校南开大学负责人, 曾参与和主持获得多项天津市教学成果特等奖、天津市科技进步一等奖和南开大学教学成果二等奖等。

讲者报告详细内容

论坛讲者:



张晨曦
同济大学

论坛讲者1

1992-2004年在国防科技大学任计算机教授, 2005年至今在同济大学任特聘教授。国家杰青, 全军优秀教师, 上海市教学名师。主要从事计算机体系结构的研究, 是我国第一台智能计算机的第二研制人。先后主持过一个国家973课题和5项国家自然科学基金项目, 发表学术论文100多篇。曾获部委级科技进步奖一等奖两项, 国家发明专利3项。负责主讲的“计算机系统结构”课程先后被评为国家级精品课程、国家级精品资源共享课、国家级一流本科课程。先后出版了10多部教材, 其中两部被被评为国家级精品教材, 3部被评为十二五国家级规划教材, 5部被评为十一五国家级规划教材。出版专著两部, 其中一部在North-Holland出版社出版, 另一部获国家教委优秀专著特等奖。2013年获上海市教学成果一等奖(1/7), 2022年获上海市教学成果二等奖(2/7)。计算机系统结构资源获全国计算机教育大会课程资源建设特等奖(2021年)。

报告题目: 计算机体系结构一流课程建设

报告摘要:

(1) 系统结构课程目前存在的问题及对策; (2) 体系结构课程的资源建设。包括内容体系和框架、教材、教学方法、PPT课件、动画课件、教学视频、线上课程等; (3) 实验方案设计。包括虚拟实验和物理实验; (4) 国家级一流课程建设; (5) 课程思政建设。



钟将
重庆大学

论坛讲者2

博士、教授，主要从事大数据分析、自然语言处理等研究，近年来承担了国家重点研发项目课题、军工863课题和国家自然科学基金等课题4项，重庆市自然科学基金以及横向科研项目20余项，发表学术论文60余篇，获得重庆市科技进步二等奖、重庆市政府发展研究一等奖、军队科技进步三等奖各1项。担任重庆大学优质课程《计算机组成原理》和重庆市一流课程《硬件综合设计》的课程负责人，负责学院计算机系统能力培养课程群的建设，连续六届指导本科生参与全国“龙芯杯”CPU设计大赛并获奖，作为项目负责人承担了教育部新工科研究与实践项目，重庆市高等教育教学改革研究重大项目。在《中国电化教育》和《计算机教育》等期刊上发表了多篇教学改革论文，入选2019年度“高校计算机专业优秀教师奖励计划”。

报告题目： 变革实践教学体系，强化系统能力培养

报告摘要：

本报告主要分享重庆大学计算机学院在实施系统能力培养改革工作过程中，如何开展基于OBE理念，通过课程群协同改革来强化系统能力培养，并将创新能力培养融入专业课程教学活动的探索、初步成效和改革经验。重点介绍在计算机核心课程的实践教学情况，特别是通过导入全国系统能力培养大赛教学资源 and 评价方法，形成“以赛促教、以赛促学”机制，推进相关课程群的建设。



王党辉
西北工业大学

论坛讲者3

博士，副教授，就职于西北工业大学计算机学院。分别于1997年/2000年/2005年在西北工业大学获得计算机软件/计算机系统结构/计算机应用技术专业的学士/硕士/博士学位。于2007年7月-9月赴比利时鲁汶大学和欧洲微电子中心进修数字系统设计，于2012年8月-2013年8月赴美国匹兹堡大学计算机科学系和电子与计算机工程系做访问学者。近年来，作为项目负责人承担了国家自然科学基金面上项目2项、陕西省自然科学基金2项；作为技术负责人或技术骨干，承担了核高基重大专项、国家预研项目、基础研究项目10多项；获得省部级科技进步奖一等奖2项。发表研究论文80余篇，包括IEEE TC、IEEE TCAD、IEEE TVLSI、DAC、ICCAD等顶级期刊和顶级会议。目前的主要研究方向包括计算机系统结构、新型存储系统、高效能计算系统、嵌入式系统、人工智能加速器设计等。近年来一直承担本科生“计算机组成与系统结构”、硕士生“高计算机系统结构”等课程的教学工作，获得陕西省教学成果一等奖3项、二等奖1项；入选2021年度“计算机专业优秀教师奖励计划”；指导本科生参加“全国大学生计算机系统能力培养大赛”，获得团体一等奖2项、团体二等奖5项、团体三等奖5项，其中学生设计的CPU性能多次排名第一。

报告题目： 面向系统能力提升的计算机系统课程建设

报告摘要：

针对本科生计算机系统核心课程体系缺乏对系统设计和应用能力培养的整体考虑，学生难以形成对计算机系统的全面认识的问题，西北工业大学计算机专业设置了“计算机系统课程”综合实践课程，该课程横跨4个学期，旨在让学生自主设计完成一个CPU、一个操作系统内核、一个编译器和一个网络协议栈，切实提升学生的计算机系统设计能力，为解决我国“缺芯少魂”问题培养后备力量。



张承义
飞腾信息技术
有限公司

论坛讲者4

博士，正高级工程师，毕业于国防科技大学计算机学院，现任飞腾信息技术有限公司副总经理，中国计算机学会杰出会员、计算机工程与工艺专委会秘书长。长期从事国产高性能微处理器芯片研制和推广工作，参与多项国家级重大专项项目，获国家科技进步特等奖1项、一等奖1项和省部级科技进步一等奖多项。

报告题目： 立足自主创新，培养致力于国产芯片的一流人才

报告摘要：

飞腾公司是国产CPU芯片设计领域的龙头企业，产品在国家关键基础行业具有广泛的应用，在国产高性能微处理器设计领域积累了丰富的培养人才经验。报告结合飞腾公司技术创新和产品研发的实践，介绍国产CPU企业对芯片设计人才的能力需求，提出对高校计算机体系结构教学内容、教学模式、实践课程等方面改革的意见建议。



汪文祥
龙芯中科技术
股份有限公司

论坛讲者5

博士,正高级工程师,龙芯中科技术股份有限公司处理器核首席架构师,中国科学院大学岗位教授。主要研究方向为处理器体系结构设计、处理器验证及计算机系统性能分析优化等,参与了多项国家“核高基”、863和973项目,在国内外各种期刊会议发表文章十余篇,申请专利数十项,授权专利十余项。2012年起任龙芯中科芯片研发部IP组组长,负责龙芯系列CPU IP核的开发研制及维护,参与制订龙芯指令系统架构规范。著有《CPU设计实战》、《计算机体系结构》、《计算机体系结构基础》等著作。

报告题目: 龙架构教育开源规划与进展

报告摘要:

自主开源指令集龙架构(LoongArch)已被国际开源社区广泛接受,基于龙架构的产品已在市场中广泛应用。龙架构中的32位精简版分支针对计算机体系结构方向教学和人才培养需求定制,面向教育领域开源。在开源推广过程中,龙芯公司除了联合国内教育界同仁共同开发基于龙架构的课程、教材,还推出了“百芯计划”,旨在为高校师生提供覆盖处理器芯片前后端设计的全流程平台,助力各高校建设基于自研芯片的学科教研体系。本报告将介绍龙架构在教育领域开源的工作规划与当前的进展情况。



余子濠
中国科学院计算
技术研究所

论坛讲者6

博士,中国科学院计算技术研究所工程师。“一生一芯”计划的培养方案设计者和首席讲师,参与建设大规模人才培养流程。教学版模拟器NEMU和南京大学“计算机系统基础”课程实验PA的设计者,多年担任该课程实验的助教。南京大学全系统教学实验Project-N的联合设计者之一。参与编写《计算机系统基础第2版》、《计算机组成与设计(基于RISC-V架构)》、《数字逻辑与计算机组成》等教材。

报告题目: “一生一芯”计划计算机系统能力培养

报告摘要:

“一生一芯”计划已历五期,自第四期开始将计算机软硬件系统能力加入教学方案,指导学生完成一款可流片的RISC-V处理器芯片,并运行自己编写的简易操作系统和真实游戏《仙剑奇侠传》,在芯片设计的过程中帮助学生建立程序、运行时环境、编译器、操作系统、指令集架构和微结构设计之间的关联认识。本报告将介绍“一生一芯”计划当前的教学方案,并通过若干案例展示学生如何全面认识计算机系统。



王超
中国科学技术大学
CCC竞赛组委会

论坛讲者7

博士,CCF杰出会员,中国科大副教授,博导,软件学院副院长。CCF体系结构专委会常务委员,江苏省电子学会副秘书长,中国科学院青年创新促进会优秀会员,ACM TODAE、TCBB等国际期刊的编委。

报告题目: 产学合作—CCF TCArch可定制计算挑战赛的人才培养尝试

报告摘要:

定制计算算法实现挑战赛(Customized Computing Challenge, CCC)是中国计算机学会体系结构专委会所举办的旨在挖掘计算机体系结构人才、激发学生对计算机体系结构的研究兴趣、培养学生创新精神的一项赛事。赛事从2021年开始连续举办了3届,累计覆盖了国内60多所高校,数百位同学。本次报告将就此次赛事的产学合作模式,比赛成果等多方面探讨赛事的人才培养尝试。

分论坛日程

论坛16:RISC-V生态关键技术论坛

时间:2023年8月5日13:30 - 17:00

论坛简介:

RISC-V已成为一个全球关注的开源硬件平台,其灵活、可定制、免授权费用的特性赋予了其重要的战略价值。越来越多的科研机构和企业正积极探索和参与到RISC-V生态的建设和发展中,推动着RISC-V的技术进步和生态繁荣。本次论坛将全面探讨RISC-V架构的前沿技术和应用,包括JIT编译技术的应用、服务器级与嵌入式RISC-V芯片的设计与应用、敏捷设计方法在处理器开发中的实践、异构SoC设计的开源IP和敏捷设计策略,以及在RISC-V架构下实现内存隔离和安全的方法。本次论坛将组织嘉宾们分享他们的最新研究成果,探讨RISC-V生态中的关键技术和发展趋势,为RISC-V生态的进一步发展注入新的动力。

具体日程安排(论坛16)

时间	报告内容	报告嘉宾/单位
13:30 - 14:00	JIT编译技术在RISC-V软件生态中的应用	吴伟/中国科学院软件研究所
14:00 - 14:30	开放的 RISC-V:开创科研和创新的无限可能	陆吉年/算能科技
14:30 - 15:00	“香山”开源高性能RISC-V处理器与敏捷开发方法实践	石侃/中国科学院计算技术研究所
15:00 - 15:20	茶歇	
15:20 - 15:50	基于开源IP和敏捷设计的领域定制SoC芯片	贾天宇/北京大学
15:50 - 16:20	RISC-V可信执行环境的内存隔离机制探索	杜东/上海交通大学
16:20 - 17:00	Panel: RISC-V生态关键技术	

论坛主席:



包云岗
中国科学院计算
技术研究所

包云岗,研究员,现任中国科学院计算技术研究所副所长,中国科学院大学计算机学院副院长,中国开放指令生态(RISC-V)联盟秘书长。主要研究领域为开源芯片与敏捷设计、数据中心体系结构等,带领团队在国内率先开展了一系列开源芯片实践。发表70余篇学术论文,包括ASPLOS、HPCA、ISCA、MICRO和Communication of the ACM等国际会议期刊。曾获“CCF-IEEE CS”青年科学家奖、共青团中央“全国向上向善好青年”等荣誉称号。



夏虞斌
上海交通大学

夏虞斌,上海交通大学教授,博导,上海市优秀学术带头人,“国家级青年人才计划入选者,挑战杯”特等奖优秀指导老师。主要研究领域为操作系统与系统架构,论文发表在 OSDI、SOSP、ISCA、ASPLOS、ToCS 等国际会议与期刊,曾获 DSN“时间检验奖”、教育部技术发明一等奖、CCF NASAC 青年软件创新奖,《麻省理工科技评论》中国“隐私计算科技创新人物”等。

讲者报告详细内容

论坛讲者:



吴伟
中国科学院软件研究所

论坛讲者1

中国科学院软件研究所PLCT实验室项目总监、LLVM基金会理事会理事、RISC-V大使。在RISC-V国际基金会(RVI)担任技术委员会(TSC)委员、中国联络人、Infra HC 主席、RVP TG 联席主席等职务。HelloGCC及CNRV(RISC-V中国)社区核心成员。深耕编译技术领域,于2019年创立PLCT实验室,团队在RISC-V基础软件领域有多项贡献被 RISC-V International、GNU Toolchain、LLVM Project、Google V8、Mozilla Spidermonkey、OpenJDK、QEMU等重要开源社区接收,并在多个开源社区中承担RISC-V架构维护(Maintainers)角色,开源贡献已惠及全球超百亿RISC-V相关产品部署。2021年、2022年获得RISC-V国际基金会技术领导力及社区贡献等多项奖项。

报告题目: JIT编译技术在RISC-V软件生态中的应用

报告摘要:

JIT编译技术是在程序代码执行期间对代码进行编译、优化、生成的技术,对于Java、JavaScript等语言的代码执行性能至关重要。RISC-V 的开源软件生态环境在过去四年有了新的突破,其中最主要的原因是 OpenJDK、V8、ART Runtime、Spidermonkey等基础软件中内嵌的JIT编译器完成了 RISC-V 的适配和优化,使得 Java 生态和 JavaScript 生态得以嫁接到 RISC-V。本次演讲中将介绍JIT编译技术的原理和前沿实现,并会介绍在V8、OpenJDK等虚拟机中的 RISC-V JIT 移植过程。



陆吉年
算能科技

论坛讲者2

算能产品经理,20年半导体工作经验,涉及市场,营销,产品规划等职位。先后在云天励飞任职芯片市场中心总经理;SiFive China市场开发总监;国科微电子副总裁;全志科技营销总监等职位。

报告题目: 开放的 RISC-V:开创科研和创新的无限可能

报告摘要:

这次报告将介绍两种RISC-V芯片,一种是服务器级别芯片,用于不同类型的硬件产品,包括服务器、开发者套件以及集群。另一种则是可以运行Linux的开发板,主要面向嵌入式等领域。本次报告将深入探讨这些基于RISC-V架构的芯片的设计和性能特点,并着重介绍它们在云和端侧应用中的优势和应用场景。同时,也将分享算能在教学和研究方面的实践经验和成果,以及如何将其应用于各种领域,包括人工智能、物联网、机器人等。



石侃
中国科学院计算技术研究所

论坛讲者3

中国科学院计算技术研究所副研究员,RISC-V国际基金会技术委员会委员,国际开源硬件组织OpenHW亚洲工作组常委。博士毕业于伦敦帝国理工学院电子工程系,曾任英特尔芯片工程师,兼任研究科学家。他有超过十年的芯片研究和从业经历,聚焦开源芯片敏捷设计与验证、FPGA的数据中心网络加速等领域的研发和创新工作。曾获IEEE Micro Top-Picks论文奖、HiPEAC 论文奖等。

报告题目: “香山”开源高性能RISC-V处理器与敏捷开发方法实践

报告摘要:

芯片敏捷设计方法能大幅提升芯片设计效率,缩短芯片开发周期,但现有的芯片敏捷设计方法仍然存在两个主要问题,导致其在实际应用中的使用仍然有限:1)缺乏支持敏捷芯片设计的工具链和开发框架,尤其是对于大规模的现代处理器而言;2)传统的验证方法不够敏捷,使之成为整个芯片开发过程的主要瓶颈。为了解决这两个问题,我们提出了一个支持敏捷处

理器开发流程的开源平台MINJIE,它整合了一系列工具,用于逻辑设计、功能验证、性能建模、硅前验证和调试,以进一步提升最先进的处理器设计的开发效率。通过该平台与工具,我们构建了两代名为“香山”的开源超标量乱序执行的RISC-V处理器,目前在GitHub上获得超过3200个星标,形成超过400个分支,成为国际上最热门的开源硬件项目之一,得到国内外企业和研究者的积极支持。在本报告中,我将介绍MINJIE这个处理器敏捷开发平台,包括其中的关键技术与研究成果,以及“香山”处理器的设计演进与发展方向。



贾天宇
北京大学集成
电路学院

论坛讲者4

北京大学集成电路学院研究员。美国西北大学博士,哈佛大学博士后,前卡耐基梅隆大学助理研究教授。研究兴趣为数字集成电路设计与计算机体系架构,包括人工智能加速器设计、SoC 芯片架构设计、敏捷设计方法等,其在旗舰会议 ISSCC、VLSI、JSSC、MICRO、DAC 等发表文章 40 余篇。其博士工作获得了美国西北大学最佳博士论文奖,以及国际固态电路学会 IEEE SSCS Pre-Doctoral Achievement Award。

报告题目: 基于开源IP和敏捷设计的领域定制SoC芯片

报告摘要:

摩尔定律的变慢和多核处理器有限的并行性推动了现代计算系统向异构的SoC架构发展。然而,异构体系架构增加了芯片设计和验证的复杂性。开源硬件和敏捷芯片设计可以通过促进设计复用来缓解这一芯片设计挑战。本次报告将介绍基于开源IP和敏捷设计的12nm领域专用RISC-V SoC。SoC采用灵活可扩展的设计方法,基于完全开源的RISC-V处理器和高层次设计加速器,以上方法使得设计人员在4个月内完成复杂的SoC芯片设计。本报告还将介绍服务于领域定制的异构SoC建模和设计空间探索方法,实现定制化芯片的快速设计和优化。



杜东
上海交通大学
IPADS

论坛讲者5

上海交通大学并行与分布式系统研究所(IPADS)助理研究员,博士。长期从事体系结构与操作系统相关研究。工作发表在ISCA、ASPLOS、SOSP、OSDI等国际学术会议和ACM TOCS等国际期刊。代表论文发表在2020年在操作系统与体系结构领域顶会ASPLOS,在该会议当年论文引用排名第一。作为主要负责人领导的开源软件蓬莱项目合入鸿蒙与欧拉操作系统。担任SOSP 23 AEC共同主席,RISC-V国际基金会SPMP工作组Chair等职。

报告题目: RISC-V可信执行环境的内存隔离机制探索

报告摘要:

随着RISC-V在边端等场景的广泛应用,越来越多的RISC-V设备需要保存和处理安全敏感数据,如用户隐私数据、地理信息、生物信息等。如何设计、扩展和利用RISC-V内存隔离能力,并且基于此提供和设计机密计算系统成为一个重要的问题。本报告将探讨RISC-V架构下的内存隔离机制和可信执行环境设计,探讨不同特权级、不同粒度下的隔离机制,并介绍基于蓬莱可信执行环境系统的多形态的安全域内存隔离方案。

分论坛日程

论坛17:隐私计算软硬协同加速技术

论坛冠名赞助商: 蚂蚁集团

时间:2023年8月5日13:30 - 17:20

论坛简介:

随着大众隐私保护意识的增强,以及相关隐私保护法规的推出,隐私计算技术日益受到学术界和产业界的重视。目前以MPC、同态加密、零知识证明为代表的基于密码学的隐私计算技术,由于涉及到复杂的密码运算,在性能上还难以支撑实际场景的应用;而且往往涉及需密码学家深度介入的应用改造过程,所以其通用性和易用性也存在亟待解决的问题。结合美国DARPA DPRIVE项目以及国内相关先进计算研究计划来看,目前领域共识认为,通过硬件加速技术、编译技术等软硬协同设计技术,将是一条解决上述性能问题和易用性问题的有效路径。本论坛将组织领域研究者分享他们的最新研究成果,探讨隐私计算软硬协同加速的关键技术和发展趋势,吸引更多生力军投入这一重要领域,共同推动数据密态时代的到来。

具体日程安排(论坛17)

时间	报告内容	报告嘉宾
13:30 - 14:00	CCF-蚂蚁科研基金软硬协同专项课题介绍	闫守孟
14:00 - 14:30	基于FPGA的全同态加密计算加速器设计	鞠雷
14:30 - 15:00	全同态处理器和算子库的设计探索	路航
15:00 - 15:20	茶歇	
15:20 - 15:50	隐私计算与硬件加速:密码学加速器与加速器可信执行环境	高鸣宇
15:50 - 16:20	线性系统的通用优化与高性能模乘设计	张博
16:20 - 16:50	同态加密软硬件协同设计	边松
16:50 - 17:20	可信执行环境的密码学拓展	金意儿

论坛主席:



闫守孟
蚂蚁集团

闫守孟, 蚂蚁集团研究员、资深总监, 可信安全计算负责人和蚂蚁技术研究院计算系统实验室主任。他在可信安全技术领域开源了一系列知名项目 (Occlum TEE OS、HyperEnclave 国产 TEE、铜锁国密密码库等)。他也领导了隐私计算加速器、隐私计算一体机、蚂蚁 Linux 操作系统、NanoVisor 安全容器等研发工作。加入蚂蚁之前, 在 Intel 从事基础技术研究。他发表了 10 余篇论文 (包括 PLDI、ASPLOS、USENIX ATC、CCS 等), 并拥有 30 余件专利。他在西北工业大学获得计算机应用技术博士学位。



路航
中国科学院计算
技术研究所

路航, 中国科学院计算技术研究所处理器芯片全国重点实验室副研究员、硕士生导师, 中国科学院青年创新促进会成员, 计算所新百星。主要研究方向为全同态处理器设计。他发表论文20篇 (MICRO、HPCA、DAC、ICCAD、TCAD、TVLSI、ICPP、ICCD等)。作为负责人/执行负责人, 负责国家自然科学基金青年基金、面上项目, 参与国家科技部重点研发子课题和中国科学院XD项目; 负责华为、OPPO等企事业单位合作项目3项; 获得2022年北京市经信局、北京市大数据中心举办的北京市大数据技能大赛隐私计算赛道季军; 获得2022年电子学会科技进步二等奖。

讲者报告详细内容

论坛讲者:



鞠雷
山东大学

论坛讲者1

鞠雷, 教授, 博士生导师, 山东大学网络空间安全学院副院长, 泰山学者青年专家。担任山东省高等学校实验室“EDA数学科学实验室”副主任、山东大学-蚂蚁集团隐私计算联合研究中心负责人。担任中国计算机学会体系结构专委会委员, 容错计算专委会委员。

研究方向为应用驱动的计算机系统软硬件协同设计、分析与优化方法, 具体的研究内容包括面向人工智能与数据安全的领域专用处理器架构与存储系统设计、面向新型计算机体系结构的系统软件研究、以及实时嵌入式系统的设计与分析等。在HPCA、DAC、TCAD等芯片设计自动化、嵌入式与实时系统领域的国际高水平会议期刊上发表论文50余篇。

报告题目：基于FPGA的全同态加密计算加速器设计**报告摘要：**

全同态加密技术可确保数据处理权与数据所有权的分离,是最具广阔应用前景的隐私保护技术之一。然而,相比明文状态高达5至6个数量级的计算与存储开销是全同态加密在实际应用场景中落地面临的主要障碍。采用FPGA等异构计算架构对全同态算子和应用进行加速实现已展示了巨大潜力,但性能方面仍受到片上计算存储资源和访存带宽的限制。与此同时,全同态加密加速器设计受到安全性、计算精度、性能等多维度设计目标的相互制约,庞大的设计优化空间对于全同态应用端到端的自动化部署产生挑战。本次报告将围绕全同态加密技术在FPGA上的加速,介绍计算并行与访存优化技术。同时,结合全同态应用,例如矩阵乘法和CNN推理,针对多种规格的硬件平台,对安全性、性能、计算精度等多目标优化的设计空间探索技术进行探讨。



路航
中国科学院计算
技术研究所

论坛讲者2

路航,中国科学院计算技术研究所处理器芯片全国重点实验室副研究员、硕士生导师,中国科学院青年创新促进会成员,计算所新百星。主要研究方向为全同态处理器设计。他发表论文20篇(MICRO、HPCA、DAC、ICCAD、TCAD、TVLSI、ICPP、ICCD等)。作为负责人/执行负责人,负责国家自然科学基金青年基金、面上项目,参与国家科技部重点研发子课题和中国科学院XD项目;负责华为、OPPO等企事业单位合作项目3项;获得2022年北京市经信局、北京市大数据中心举办的北京市大数据技能大赛隐私计算赛道季军;获得2022年电子学会科技进步二等奖。

报告题目：全同态处理器和算子库的设计探索**报告摘要：**

业界普遍认为阻碍全同态加密(FHE)商业落地的主要原因是其硬件性能问题,因此设计专用处理器(DSA)是提高FHE硬件性能的可行方案之一。在硬件设计上,首先要对FHE的算子进行合理拆分,利用可重构的方式组合算子硬件模块;在算子库设计上,要合理利用硬件特性,设计合适粒度的算子软件实现以及驱动;在易用性设计上,要提供参数化配置减少用户在密码学原理学习上的时间开销。本报告汇报团队在全同态专用处理器——“张江壹号”和全同态算子库——“Poseidon”的实现探索,力求解决软硬件协同设计瓶颈并促进FHE的商业化落地。



高鸣宇
清华大学

论坛讲者3

高鸣宇,清华大学交叉信息研究院助理教授,博士生导师。于美国斯坦福大学电子工程系获博士学位。研究方向为计算机体系结构与系统,主要关注针对人工智能和大数据分析等数据密集型应用的新型存储架构、专用计算系统、硬件系统安全等方向。已发表多篇国际顶级学术会议(ISCA、ASPLOS、HPCA、OSDI、PACT等)论文,曾获得IEEE Micro 2016年度计算机系统结构最佳论文奖(Top Picks)、福布斯中国2019年30 Under 30等荣誉。他也是MICRO、ASPLOS、ISCA等多个会议的TPC和ERC委员会成员。

报告题目：隐私计算与硬件加速:密码学加速器与加速器可信执行环境**报告摘要：**

在当前的人工智能与大数据时代,数据处理不仅要求越来越高的性能,同时也需要完善的隐私保护机制确保计算过程中的私密性和完整性。性能方面,FPGA、GPU、ASIC等专用硬件加速平台被广泛采用;安全方面,隐私保护的计算可以基于多方安全计算、同态加密、零知识证明等现代密码学算法,或是基于硬件可信执行环境所提供的软硬件隔离保护。在本报告中,我将讨论以上两种隐私计算方式与硬件加速平台的结合,分为“accelerators for security”,即利用专用硬件架构加速现代密码学算法,和“security for accelerators”,即在硬件加速平台上构建可信执行环境。



张博
蚂蚁技术研究院

论坛讲者4

蚂蚁技术研究院副研究员。美国南加州大学博士。研究兴趣为数字集成电路设计与全同态加密,包括硬件加速器设计、高性能计算等,其在IEEE Transactions on Computers、DATE等期刊或会议发表文章10余篇。

报告题目：线性系统的通用优化与高性能模乘设计**报告摘要：**

基于乘法和加法的线性系统广泛存在于硬件加速等应用场景。本次演讲将会介绍线性系统的通用优化，将乘法简化为压缩，移除全位宽加法。保证线性系统正确性的同时，优化面积和时间。对于同态加密等隐私计算而言，其硬件加速需要例化大量模乘单元。本次演讲也会在线性系统优化的基础上，进一步介绍模乘的优化，移除取整和取模等非线性操作的影响，为隐私计算提供高性能原语算子。



边松
北京航空航天大学

论坛讲者5

边松于2014年毕业于威斯康辛麦迪逊分校，分别于2017年和2019年在京都大学取得硕士与博士学位，于2019年任京都大学助理教授，于2021年任北京航空航天大学网络空间安全学院副教授。他的主要研究方向是同态加密、隐私保护计算与密码软硬件协同加速，以第一或通讯作者于USENIX Security、TIFS、VLDB及DAC等CCF-A类期刊及会议发表论文9篇。获国家自然科学基金委、日本文部科学省战略创造研究、日本学术振兴会特别研究员项目支持。他是多个国际会议与期刊的程序委员与审稿人，包括AAAI、ICML、IEEE TDSC、TIFS及TCAD。他是IEEE与CCF的会员，CCF容错专委会执行委员。

报告题目：同态加密硬件协同设计**报告摘要：**

全同态加密是一种可以在不传输解密密钥的前提下在密文上执行任意(图灵完备)计算的密码算法，在外包计算及双方安全计算场景中有广泛应用。尽管全同态加密拥有最优交互轮数及较低在线通信量，在全同态加密密文上的计算具有计算效率低的特点，阻碍了全同态加密在许多数据安全保护场景中的应用落地。本次报告将以全加密数据监测等场景为例，介绍近期在全同态加密的应用设计、算子设计、算法实现及硬件加速等几个方面的研究成果。



金意儿
华为

论坛讲者6

金意儿，华为可信计算首席科学家，佛罗里达大学名誉教授，IEEE硬件安全与可信专委会联席主席。他的研究领域主要涉及软硬件协同安全和新兴集成电路安全，包括硬件支持的系统安全，集成电路产业链安全，以及可信自动化等。他撰写了集成电路安全一书，同时在国际知名期刊和杂志上发表了超过200篇论文，是亚洲硬件安全年会的联合创始人。他目前是IEEE设计自动化委员会的杰出讲师。他的论文获得了多项最佳论文奖，他本人也获得包括美国能源部和美国海军在内的多项杰出青年教授奖。

报告题目：可信执行环境的密码学拓展**报告摘要：**

随着社会信息化程度的不断提高，数据正在成为越来越重要的资产，对数据进行高效、安全地控制也成为企业的核心竞争力之一。已经有了各类安全和可信技术，包括可信执行环境，软件沙箱，还有各类传统的和新兴的密码技术，从传统加解密方案，到同态计算、多方计算等。但是目前缺乏一个单一的技术能从性能、安全、可信等角度都满足需求，这就要求研究人员从技术组合的角度重新审视数据安全。在这个讲座中，我们会介绍利用密码学知识对硬件可信执行环境进行边界拓展的工作进展，基于这一框架，数据安全的保护会根据性能、安全性等约束条件进行动态分配，在硬件安全边界和密码学边界之间进行交互，保证数据的运算(无论是以明文形式在可信执行环境内部还是以密文形式在密码学框架内部)始终处在受保护的受保护的状态下。我们也会同步介绍“孟极”框架的进展，这一框架是2022年度胡杨林基金可信计算专项的一个重要成果。

赞助单位

单位名称		展位编号
 钻石	 蚂蚁技术研究院 Ant Research	蚂蚁集团 2
	 中科曙光 Sugon	曙光 1
 铂金	 龙芯中科 LONGSON TECHNOLOGY	龙芯 3
 黄金	 HUAWEI	华为 14
	 阿里云	阿里云
 白银	 RICORE	中科睿芯 4a
	 趋动科技 VirAI Tech	趋动科技 4b
	 AVNET	AVNET 8、10、12
	 奇安信	奇安信

蚂蚁技术研究院

探索未来前沿科技

为进一步强化和完善蚂蚁技术探索和创新机制，提升技术布局的前瞻性，体现蚂蚁在前沿硬核技术领域全力以赴、攻坚克难的决心，于2021年4月成立蚂蚁技术研究院。



交互智能实验室

交互智能实验室聚焦基础模型研究、开发通用人工智能算法架构，包括内容生成、多模态理解、数字化技术、人机交互等关键技术。



密码学实验室

聚焦于前沿密码学技术的研究，包括但不限于安全多方计算、同态加密、差分隐私等，并将其应用到蚂蚁的场景中，实现企业和消费者用户的数据安全和隐私保护。



程序设计与编译器实验室

支持隐私计算加速芯片等，形成软硬件一体优化能力。



数据库实验室

云时代的金融级分布式HTAP数据库关键技术。



计算系统实验室

硬件一体化体系结构、芯片和操作系统（包括TEE）研究。



图计算实验室

高性能实时图计算，图数据库和图学习。

蚂蚁技术研究院致力做有用、有想象力的科研。面向数字化、智能化未来，瞄准世界科技前沿，推进关键核心技术攻关，促进“产学研用”深度融合，为做强做大做优我国数字经济努力贡献力量。

欢迎加入蚂蚁技术研究院



产学研合作与前沿技术交流
AntResearch@antgroup.com



扫码关注蚂蚁技术研究院



微信号



微信公众号



曙光计算云 | Sugon

曙光计算云百万核时计划

《计算服务免费试用》



首次免费体验30天

10000核·时

1000卡·时



免费领取

超大规模且持续扩展的算力池



省时

省钱

省力

🔍 | AC.sugon.com :



龙胜男 18610272731



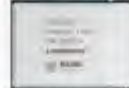






















*活动最终解释权归曙光信息产业股份有限公司所有

龙芯中科

龙芯中科面向国家信息化建设需求，面向国际信息技术前沿，以创新发展为主题、以产业发展为主线、以体系建设为目标，坚持自主创新，全面掌握 CPU 指令系统、处理器 IP 核、操作系统等计算机核心技术，打造自主开放的软硬件生态和信息产业体系，为国家战略需求提供自主、安全、可靠的处理器，为信息产业的创新发展提供高性能、低成本的处理器和基础软硬件解决方案。

龙芯中科主营业务为处理器及配套芯片的研制、销售及服务，主要产品与服务包括处理器及配套芯片产品与基础软硬件解决方案业务。目前，龙芯中科基于信息系统和工控系统两条主线开展产业生态建设，面向办公与业务信息化、工控及物联网、网络安全等领域与合作伙伴保持全面的市场合作，系列产品在能源、交通、电子政务、金融、电信、教育等行业领域已获得广泛应用。

龙架构芯片家族

龙芯3号 CPU系列 面向桌面和服务器应用	LS3A5000 LS3B5000  2.5GHz 4 LA464	LS3C5000  2.2GHz 16 LA464	LS3D5000  2.2GHz 32 LA464	LS3A6000  2.5GHz 4 LA664	LS3C6000 LS3D6000 					
龙芯2号 SoC系列 面向工控和终端应用		LS2K1000LA  1.0GHz 2 LA264	LS2K0500  500MHz 1 LA264	LS2K1500  1.0GHz 2 LA264	LS2K2000  1.5GHz 2 LA364	LS2P0500  打印机 专用芯片	LS2K3000 LS2K0300 			
龙芯1号 MCU系列 面向嵌入式 专门应用				LS1C102  32MHz LA132	LS1C103  32MHz LA132	LS100100  8MHz LA132	LS1号系列 			
龙芯配套芯片	LS7A1000  配套桥片	LS7A2000  配套独显桥片	LS8T41505  时钟芯片 时钟发生器	LS8P74401  电源芯片 LDO	LS8P2050  电源芯片 LDO	LS8P80103  电源芯片 LDO	LS8P63601  电源芯片 DC-DC	LS8T2309  时钟芯片 时钟Buffer	LS8T2305  时钟芯片 时钟Buffer	配套芯片 
	2018	2021	2022	2023	未来					





☎ 18625210050

🌐 www.ri-core.cn

📍 苏州工业园区星湖街328号创意产业园11-303

苏州睿芯集成电路 科技有限公司

苏州睿芯集成电路科技有限公司,于2020年8月落地苏州,是苏州工业园区重大领军人才企业。公司核心团队来自于中科院计算所,主要从事高通量芯片的自主研发和生产,并多次获得科技部重大项目和国家级重大专项的支持。公司于2022年获得了苏州市“独角兽”企业的荣誉,于2023年获评全国“潜在独角兽”企业。公司的研发成果已具备良好的产业生态,芯片产品化应用成果将直接应用到高通量计算的各类应用场景中。

苏州睿芯的高通量加速芯片和边缘服务器芯片,基于RISC-V架构和数据流设计方法,拥有强大的数字信号处理、人工智能、科学计算以及实时视频处理能力。目前公司芯片已经完成量产流片,在多家客户产品上得到应用,客户应用包括智慧交通、智慧城市、音视频智能分析、智能驾驶、5G基站等领域。





中国奥委会官方赞助商

奇安信简介

奇安信科技集团股份有限公司（以下简称奇安信，股票代码 688561）成立于 2014 年，专注于网络空间安全市场，向政府、企业用户提供新一代企业级网络安全产品和服务，在人员规模、收入规模和产品覆盖度上均位居行业第一。

2019 年 5 月，中国电子以 37.31 亿元战略入股奇安信，奇安信正式成为网络安全国家队。2019 年 12 月，奇安信成为北京 2022 年冬奥会和冬残奥会官方网络安全服务和杀毒软件赞助商。2020 年 7 月 22 日，奇安信在科创板挂牌上市。

2022 年 3 月 13 日，奇安信圆满完成了北京冬奥会和冬残奥会网络安全保障工作，兑现了北京冬奥网络安全“零事故”的承诺，为我国关键信息基础设施和重大活动的网络安全保障提供示范样本和有益经验。

奇安信技术研究院

奇安信技术研究院是专注于网络空间安全相关技术的研究机构，聚焦网络空间安全领域基础性或前沿性的研究课题，结合国家和社会的实际需求，开展创新性和实践性的技术研究。

研究院员工多数来自清华、中科院等知名研究机构，有多年的学术研究背景，结合奇安信集团在工程能力和安全大数据的优势，研究成果可以直接应用于产品和服务。

研究院与国内众多知名的高校和科研院所建立了紧密的合作，包括清华大学、上海交通大学、浙江大学、复旦大学、中国科学院、山东大学、东南大学、中山大学等等。

研究院主要研究方向包括互联网基础设施安全（域名、路由、PKI 等）、IPv6 下一代互联网安全体系和关键技术、5G 安全体系和关键技术、网络协议的漏洞分析与安全测量、网络威胁检测与态势感知、漏洞挖掘与漏洞攻防研究等。

2019 年起，研究院先后成立了羲和实验室、星图实验室、天工实验室三大实验室，分别专注于互联网基础协议、软件和供应链、物联网等安全研究。

软件定义算力技术领导者

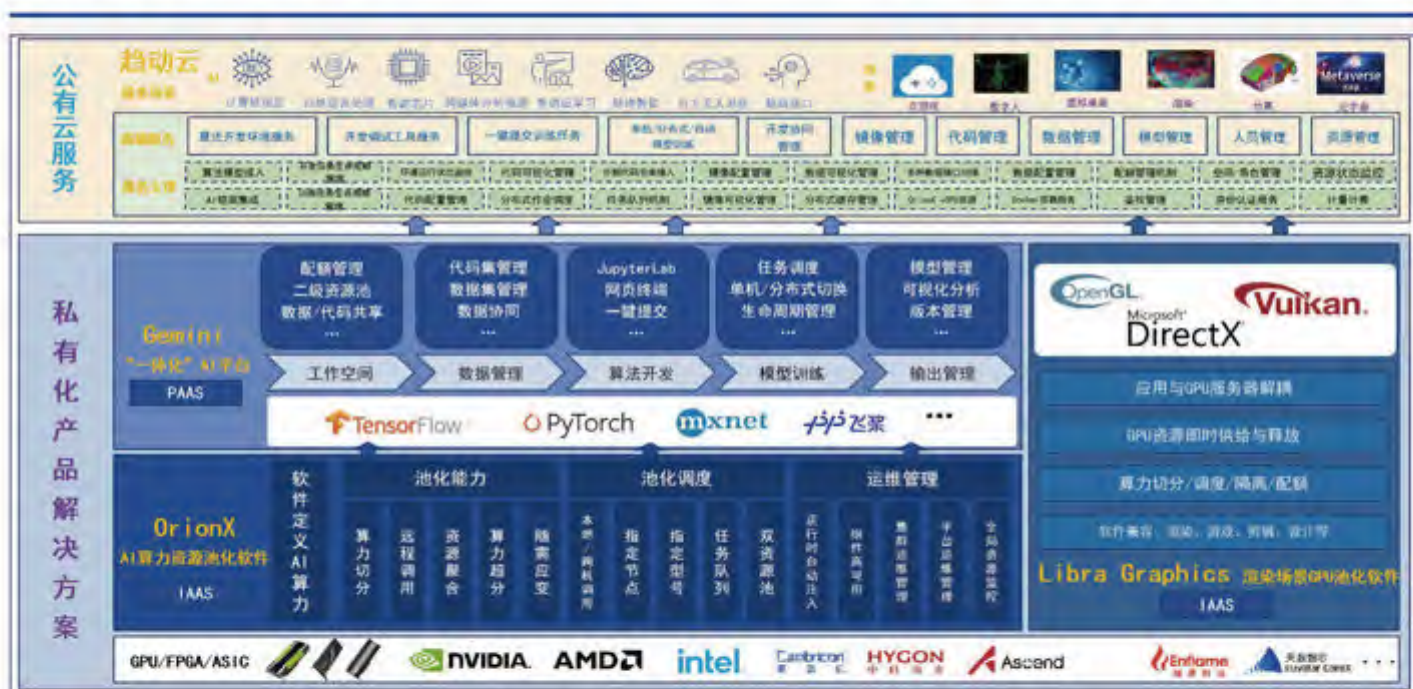
为客户提供算力池化解决方案

关于趋动科技

趋动科技作为软件定义AI算力技术的领导厂商，专注于为全球用户提供国际领先的数据中心级AI算力虚拟化和资源池化软件及解决方案，已完成中关村高新、国高新、“专精特新”等企业认证。趋动科技的 OrionX 猎户座 AI 算力资源池化软件能够帮助用户提高资源利用率和降低TCO，提高算法工程师的工作效率。趋动科技的双子座 GEMINI AI 训练平台，为客户提供强大的AI算力管理服务以及高效的算法开发和训练支持，能够化繁为简，帮助企业建好AI平台、管好GPU、用好 AI 服务。依托全球领先的AI算力池化技术，趋动科技重磅推出趋动云 VirtAI Cloud，为万千企业和AI开发者带来又便宜、又好用的AI算力池化云服务。

凭借标准化、可复制的产品架构，趋动科技得到了包括互联网、金融、电信运营商、自动驾驶、能源、科研机构 and 高校等大量行业头部客户的认可。资本市场对于趋动科技的发展充满信心——趋动科技成立两年多已经完成近亿美元的融资，顶级的投资机构持续支持趋动科技的发展，包括国开装备基金、沙特阿美旗下多元化风投基金 Prosperity7 Ventures、元禾重元、招银国际、顺为、高瓴、嘉御、戈壁、讯飞和涌铎在内的国内外顶级VC正在见证趋动科技锐意进取的脚步。

趋动科技产品矩阵



趋动科技官方微信

BEIJING 北京总部
 地址：北京市海淀区中关村大街1号海龙大厦10层北
 电话：010-62560919

SHENZHEN 深圳
 地址：深圳市南山区粤海街道科苑南路3099号中国储能大厦1319

SHANGHAI 上海
 地址：上海市徐汇区漕溪北路595号上海电影广场C幢12层
 电话：021-62263581

HANGZHOU 杭州
 地址：杭州市余杭区文一西路1818-2号人工智能小镇3幢704室

