CONFERENCE ABSTRACT

2025 9TH CONFERENCE ON

HIGH PERFORMANCE COMPILATION, COMPUTING AND COMMUNICATIONS

HP3C 2025

Jinan, China | August 27-29, 2025

Sponsored by

齐鲁工业大学 QILU UNIVERSITY OF TECHNOLOGY

山东省计算中心（国家超级计算济南中心） SHANDONG COMPUTER SCIENCE CENTER (NATIONAL SUPERCOMPUTING CENTER IN JINAN)

Co-sponsored by

国家超级计算济南中心 NATIONAL SUPERCOMPUTER CENTER IN JINAN

Supported by

广东工业大学 Guangdong University of Technology

东北电力大学 NORTHEAST ELECTRIC POWER UNIVERSITY

TABLE OF CONTENTS

ORGANIZING COMMITTEE

▶ Honorary Chair
Yinglong Wang, National Supercomputer Center in Jinan, China

▶ International Advisory Committee
Witold Pedrycz, IEEE Life Fellow, University of Alberta, Edmonton, Canada
Qingfu Zhang, City University of Hong Kong, Hong Kong, China
James Lin, Shanghai Jiao Tong University, China
Jason Dai, Principle Engineer and CTO, Big Data Technologies, Intel

▶ Conference Chair
Meihong Yang, Qilu University of Technology (Shandong Academy of Sciences), China

▶ Conference Co-chairs
Xiaoming Xiong, Guangdong University of Technology, China
Steven Guan, Xi'an Jiaotong-Liverpool University, China

▶ Technical Program Committee Chairs
Xiaoming Wu, Qilu University of Technology (Shandong Academy of Sciences), China
Shahadat Hossain, University of Northern British Columbia, Canada

▶ Technical Program Committee Vice-chair
Yoshifumi Manabe,Kogakuin University, Japan
Jingshan Pan, Qilu University of Technology (Shandong Academy of Sciences), China
Jianqiang Huang, Qinghai University, China

▶ Local Organizing Chair
Wei Ding, Qilu University of Technology (Shandong Academy of Sciences), China

▶ Publication Chairs
Yuedong Yang, Sun Yat-sen University, China
Wei Zhang, Qilu University of Technology(Shandong Academy of Sciences), China

▶ Publicity Co-chairs
Ying Guo, Qilu University of Technology (Shandong Academy of Sciences), China
Ljiljana Trajkovic, Simon Fraser University, Canada
Jian Wang, Kunming University of Science and Technology, China
Dayang Norhayati Abang Jawawi, Universiti Teknologi Malaysia, Malaysia
Siti Zaiton Mohd Hashim, Universiti Teknologi Malaysia, Malaysia
Beligiannnis Grigorios, University of Patras, Greece

▶ Special Session Chairs
Hui Zhang, Southwest University of Science and Technology, China
Jianpo Li, Northeast Electric Power University, China

- Technical Committee Members
Daqing Yun, Harrisburg University, USA

Elankovan A Sundararajan, The National University of Malaysia, Malaysia
Hua Qu, Xi'an Jiaotong University, China
Jian Wang, Kunming University of Science and Technology, China
Jianbing Ma, Chengdu University of Information Technology, China
Kohei Fujita, The University of Tokyo, Japan
Kou Lei, Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), China
Ljiljana Trajkovic, Simon Fraser University, Canada
Mohamed Arezki Mellal, M'Hamed Bougara University, Algeria
Pei He, Guangzhou University, China
Riswan Efendi, UIN Suska Riau, Indonesia
Shengxin Zhu, Beijing Normal University at Zhuhai, China
Shu Yin, ShanghaiTech University, China
Shujie Yang, Beijing University of Posts and Telecommunications, China
Siti Zaiton Mohd Hashim, Universiti Teknologi Malaysia, Malaysia
Tan Tse Guan, Universiti Malaysia Kelantan (UMK), Malaysia
Wael Yafooz, Taibah University, Saudi Arabia
Dan Huang, Sun Yat-sen University, China
Dayang Norhayati Abang Jawawi, Universiti Teknologi Malaysia, Malaysia
JONG YUN LEE, Chungbuk National University, South Korea
Julio Sahuquillo, Universitat Politècnica de València, Spain
Khalaf Khatatneh, BAU, Jordan
Manuel Mazzara, Innopolis University
Muhammad Imran Babar, University of Southampton, Malaysia
Riktesh Srivastava, City University, Ajman, UAE
Simon Winberg, University of Cape Town
Wan Nor Shuhadah Wan Nik, Universiti Sultan Zainal Abidin, Malaysia
Wei Ding, Qilu University of Technology(Shandong Academy of Sciences), China
Farhan Aadil, COMSATS University Islamabad
Ilia Peshkov, Voronezh State University
Meng Hao, Harbin Institute of Technology, China
Nazim Choudhury, University of Wisconsin, USA
Maizatul Akmar Ismail, University of Malaya, Malaysia
Othman Ibrahim, Universiti Teknologi Malaysia, Malaysia
Anand Nayyar, Duy Tan University
Yuanhao Gong, Shenzhen University, China
Amine Khaldi, Universite Kasdi Merbah Ouargla, Algeria
Fangfang Zhang, Qilu University of Technology(Shandong Academy of Sciences), China
Fei Ma, Xi'an Jiaotong-Liverpool University, China
Layth Nabeel AlRawi, Parker Wellbore, USA
Nikki Shaw, Algoma University, Canada
Nitin Auluck, Indian Institute of Technology
Omid Mahdi Ebadati E., Kharazmi University
Quande Yuan, Changchun Institute of Technology, China
Wei Xue, Tsinghua University, China
Wei-Yu Chen, Chinese Culture University, China
Yuanzhe Cai, Shenzhen Technology University, China
Yunquan Zhang, Institute of Computing Technology, CAS, China
Muneo Hori, Japan Agency for Marine-Earth Science and Technology, Japan

WELCOME MESSAGE

You are cordially invited to join us at the 2025 Conference on High Performance Compilation, Computing and Communications (HP3C'25) in Jinan, China during August 27-29, 2025.

HP3C-2025 welcomes author submission of papers from any branch of the High Performance Compilation, Computing and Communications, and their applications or other topic areas. The areas covered by the include, but not limited to: High performance Languages and Compilers for High Performance Computing, Multicore, Many-core and Multithreaded Architectural Approaches, Parallel and Distributed System Architectures, Parallel and Distributed Software Technologies.

The accepted paper will be included into HP3C 2025 Conference Proceedings, which will be published in the International Conference Proceedings Series, which will be archived in the ACM Digital Library, and indexed by Ei Compendex and Scopus and submitted to be reviewed by Thomson Reuters Conference Proceedings Citation Index (ISI Web of Science).

This year, we are honored to have 4 distinguished speakers, they are:

- Keun Ho Ryu, Chungbuk National University, Korea
- Kwang Woo Nam, Kunsan National University, Korea
- Yongde Guo, City University of Macau, Macau, China
- Wan Nor Shuhadah Wan Nik, Universiti Sultan Zainal Abidin, Malaysia

On behalf of the conference committee, we thank all the authors, reviewers, and attendees for their contributions and participation in HP3C 2025. Their dedication and expertise enable us to prepare this high-quality program to make the conference a success. Finally, we wish all the delegates a productive and enjoyable conference.

HP3C 2025 Conference Committee

August, 2025

ONSITE INFORMATION

❖ **Time Zone: UTC/GMT+8, Beijing Time for the whole program.**

❖ **Conference Venue**

国家超级计算济南中心科技园区主楼

地址：山东省济南市历城区经十东路彩石街道国家超级计算济南中心科技园区主楼

Main Building, National Super Computing Center in Jinan, Caishi Street, Jingshi East Road, Licheng District, Jinan, Shandong Province, China

❖ **Lunch & Dinner**

东悦国际酒店(省科学院济南超算中心科技园店)

Dongyue International Hotel

地址：济南市历城区超算中心科技园,蟠龙路东 100 米

Address: North Hushan Road and Panlong Road intersection to the northeast of about 220 meters, Licheng District, Jinan, Shandong, China

❖ **Temperature**

Average Temperature in August in Jinan: 22℃- 31℃

❖ **Attention Please**

♣ Please take care of your belongings in public area. For your personal and property safety, delegates are suggested to wear representative card during conference and not to lend it to those unconcerned to enter event rooms. Conference does not assume any responsibility for loss of personal belongings of participants.

♣ Don't stay too late in the city, don't be alone in the remote area. Be aware of the strangers who offer you service, signature of charity, etc., at scenic spots. You can search more Tourist Information and Security tips online.

❖ **Oral Presentation Tips**

♣ The duration of a presentation slot is 15 minutes. Please prepare your presentation for about 12 minutes plus about 3 minutes for questions from the audience;

♣ An LCD projector & computer will be available in every session room for regular presentations;

♣ Presentations should be uploaded at the computer at least 15 minutes before the session start.

❖ Emergency Call

Ambulance: 120 Police: 110

ONLINE INFORMATION

❖ Online Conference Information

- **Zoom Link: https://us02web.zoom.us/j/86197421409**
- **Zoom: password: HP3C**

Note:
please join the online room 10-15 mins before your session starts and be prepared.
♣ For General Users: https://zoomus/support/download
♣ For Users from mainland China: https://wwwzoom.com.cn/download

Tips:

- ✧ Please unmute audio and start video while your presentation.
- ✧ It's suggested to use headset with microphone or earphone with microphone.
- ✧ E-certificate will be sent to presenters after conference by email.
- ✧ An excellent presentation will be selected from each session and announced on the website after conference. An excellent presentation certificate will be sent after conference by email.
- ✧ It's **Beijing Time** (UTC/GMT +8) for the whole schedule.

Rename your screen name before entering the room	Example
Authors: Paper ID-Name	HP0001-Sam Louis
Keynote Speaker: Keynote-Name	Keynote- Sam Louis
Invited Speaker: Invited -Name	Invited- Sam Louis
Committee Member: Position-Name	Committee- Sam Louis

❖ Duration of Each Presentation

- ✧ Keynote Speech: 40 Minutes of Presentation including Q&A.
- ✧ Invited Speech: 20 Minutes of Presentation including Q&A.
- ✧ Regular Oral Presentation: 15 Minutes of Presentation including Q&A.

Scan the QR code and send "HP3C 2025" to add conference assistant Wechat

Online Pre-test Timetable and online sign in (August 27, 2025)

Zoom ID: 861 9742 1409 Password: HP3C

Time	Items
10:00-12:00	Online Keynote Speakers & Invited Speaker & Session Chairs & Committees
15:00-16:00	Online Session 1& Online Session 2

DAILY SCHEDULE

Day 1, August 27, 2025

14:00-16:00	**Onsite Sign in and Collect Conference Materials** **国家超级计算济南中心科技园区主楼，会议室440，4楼** Room 440, Main Building, National Super Computing Center in Jinan

Day 2, August 28, 2025

国家超级计算济南中心科技园区主楼，会议室440，4楼
Room 440, Main Building, National Super Computing Center in Jinan
Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C

Host: Wei Ding, Qilu University of Technology (Shandong Academy of Sciences), China

09:00-09:05	**Opening Remarks** **Steven Guan, Xi'an Jiaotong-Liverpool University, China**
09:05-09:45	**Keynote Speaker I** **Keun Ho Ryu, Chungbuk National University, Korea** **Speech Title: Exploring Healthcare Insight through Explainable Artificial Intelligence**
09:45-10:10	***Group Photo & Coffee Break***
10:10-10:50	**Keynote Speaker II** **Kwang Woo Nam, Kunsan National University, Korea** **Speech Title: UrbanAI Data Computing: Urban Media Data meet AI**
10:50-11:30	**Keynote Speaker III** **Yongde Guo, City University of Macau, Macau, China** **Speech Title: Multi-source Remote Sensing Data Fusion: From Single Modality to Holographic Perception**
11:30-11:50	**Invited Speaker** **Wan Nor Shuhadah Wan Nik, Universiti Sultan Zainal Abidin, Malaysia** **Speech Title: Job Scheduling in Cloud and Fog Computing: A Recent Systematic Review**
11:50-14:00	***Break & Lunch*** 2楼悦潮全日餐厅，东悦国际酒店(省科学院济南超算中心科技园店) 2F, Yue Chao All-day Dining Restaurant, Dongyue International Hotel
14:00-14:40	Center Visiting (only for Chinese) | 国家超级计算济南中心参观
14:40-15:00	***Coffee Break***

15:00-17:00	**Onsite Session** ..... **Room 436	会议室436** **Topic: Model-based Intelligent Information System Design and Optimization Algorithms** **Session Chair: Dr. Xin Li, Shandong Computer Science Center (National Supercomputer Center in Jinan), China** **HPC1009, HPC3033, HPC2024, HPS1046, HPC2025, HPS2016, HPS2013, HPS2045**
17:20-19:00	***Dinner*** 2楼悦潮全日餐厅，东悦国际酒店(省科学院济南超算中心科技园店) 2F, Yue Chao All-day Dining Restaurant, Dongyue International Hotel

Day 3, August 29, 2025

ONLINE SESSIONS

Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C

10:00-12:00	**Online Session 1** **Topic: Supercomputers and High-performance Computing** **Session Chair: Assoc. Prof. Othman Ibrahim, Universiti Teknologi Malaysia, Malaysia** **HPC1003, HPC1007, HPC1008, HPC1010, HPC2015, HPC2017, HPC2026, HPC3028**
12:00-14:00	***Break Time***
14:00-16:00	**Online Session 2** **Topic: Emerging Network Technologies and Information Security** **Session Chair: Dr. Nur Haliza Binti Abdul Wahab, University of Technology Malaysia, Malaysia** **HPC1005, HPS1029, HPC5047, HPS2019, HPS2018, HPS2020, HPS2021, HPS2023**

KEYNOTE SPEAKERS

Keun Ho Ryu

Chungbuk National University, Korea

国家超级计算济南中心科技园区主楼，会议室 440，4 楼
Room 440, Main Building, National Super Computing Center in Jinan

Boi: Keun Ho Ryu is a professor at Chungbuk National University in Korea and a leader of the group database and bioinformatics laboratory as well as a vice president of the Personalized Medicine Tumor Engineering Research Center. He received a BS degree from Soongsil University in 1976 and an MS and a Ph.D from Yonsei University, Korea, in 1980 and 1988 respectively. He worked not only at the University of Arizona as a post-doc and research scientist but also at ETRI (Electronics and Telecommunications Research Institute), Korea. He has served on numerous international journal reviewers and program committees including the IEEE International Conference on Advanced Information Networking and Applications (AINA), the Very Large Data Base (VLDB), and so on. He is a member of the IEEE and a member of the ACM since 1983. He has served on numerous program committees, including roles as the Demonstration Co-Chair for the VLDB, the Panel Co-Chair and the Tutorial Co-Chair for the APWeb, and the FITAT General Co-Chair. In 2008, he founded the FITAT International Group, to provide a professional community the opportunities, for publications, knowledge exchange, teaming, and cooperation. His research interests include temporal databases, spatiotemporal databases, stream data processing, knowledgebase information, security, data mining, bioinformatics, and biomedical informatics.

Speech Title: Exploring Healthcare Insight through Explainable Artificial Intelligence

Abstract: Research and applications of AI technologies have been advancing rapidly. However, as we incorporate AI-generated outcomes into practical settings, many questions naturally arise: Why did this result occur? Can I trust this result? These concerns are especially critical in real-world applications, particularly in the doctor-patient relationship.

To address these challenges, we propose a DeepSHAP-based Explainable AI (XAI) framework. To demonstrate its practical applicability in clinical settings, we applied our method to Non-Communicable Diseases (NCDs).

In this keynote, we aim to convey two core messages: one related to health and the other to academia. Why two messages? The answer becomes clear when we reflect on the evolution of our research journey. In the preliminary part of this keynote, we will outline the progression of our lab's research and its connection to NCDs. Then, we will introduce our DeepSHAP-based XAI approach, detail the selection of the NCD dataset used in our experiments, and present evidence supporting the validity of our framework.

KEYNOTE SPEAKERS

Kwang Woo Nam

Kunsan National University, Korea

Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C

Boi: Kwang Woo Nam is a professor at Kunsan National University in Korea, the founder of the Spatial Information Laboratory, and the dean of the School of Computer Information and Communication Engineering. His research areas cover management of computing power network devices, resource scheduling, location services and servers in spatio-temporal database systems, GeoAI data streams, and data mining, etc.

He has served as the director of education of the Korean Society for Spatial Information, the subject editor of applied sciences, a consultant on spatial information for smart cities of the Telecommunications Technology Association (TTA), an expert of the evaluation team of the Korea Agency for Infrastructure Technology Advancement (KAIA) under the Ministry of Land, Infrastructure and Transport, and a professional commissioner of the International Organization for Standardization/Technical Committee 211 (ISO/TC 211), etc.

He has presided over 19 national-level projects and led the formulation of international standards such as TTAPG409, PG405, and ISO/TC 211. In cooperation with the South Korean company TurboSorft, he participated in the research and development of the GeoCMS subsystem under the OpenGIS of the United Nations.

Speech Title: UrbanAI Data Computing: Urban Media Data meet AI

Abstract: The rapid advancement of mobile technologies has given rise to an unprecedented volume of media data enriched with geospatial tags, including photographs, videos, and microblogs. The scope of such urban media data generated in contemporary cities is remarkably broad. Beyond the data produced through smartphones, it also encompasses 360-degree street-view imagery, dashcam videos from vehicles, real-time video streams from traffic surveillance systems, as well as diverse photos and videos disseminated via social media platforms such as YouTube and Instagram. When integrated with AI technology, these urban media data unlock possibilities for dynamic and multifaceted urban analysis that transcend the limitations of traditional static urban information systems. This keynote addresses the current state of urban media data computing and provides an overview of ongoing research activities in our laboratory.

KEYNOTE SPEAKER

Yongde Guo

City University of Macau, Macau, China

Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C

Boi: Guo Yongde, Ph.D. in Information and Communication Engineering from Tsinghua University, serves as the Program Director of the Faculty of Data Science at City University of Macau, an Associate Professor, and a Doctoral Supervisor. He is also a specially appointed expert under the national talent program and the Director of the Joint Laboratory for Spatiotemporal Big Data and Artificial Intelligence at City University of Macau-Zhuhai Orbita. As a visiting scholar at the Department of Electrical Engineering and Computer Science at Oregon State University, USA, he studied under the esteemed Professor Thomas Dietterich and participated in applied projects related to environmental sustainability computing in Africa. His long-term research focuses on image processing, computer vision, pattern recognition, deep learning, environmental sustainability computing, and the spatial analysis and application of remote sensing data.

In recent years, he has published over 10 high-level academic papers and authored two monographs. He has led two research projects funded by the Macau Foundation, one horizontal project with enterprises, and has been a key research participant in one major project supported by the National Natural Science Foundation of China.

郭永德，清华大学信息与通信工程专业博士，澳门城市大学数据科学学院课程主任，副教授，博士生导师，国家级人才项目特聘专家，澳门城市大学-珠海欧比特时空大数据与人工智能联合实验室主任。美国俄勒冈州立大学电气工程与计算机系作访问学者，师从 Thomas Dietterich 荣誉教授，参与有关非洲环境可持续计算的应用项目。长期从事图像处理、计算机视觉、模式识别、深度学习、环境可持续计算和遥感数据空间分析与应用等方向研究。

Speech Title: Multi-source Remote Sensing Data Fusion: From Single Modality to Holographic Perception

Abstract: This keynote speech focuses on the cutting-edge field of multi-source remote sensing data fusion. Traditional single-modality remote sensing data has limitations in information and struggles to comprehensively depict complex scenes. In contrast, multi-source data fusion can integrate the advantages of different sensors, overcoming the constraints of single data sources. This presentation will elaborate on the development path from single modality to holographic perception, introduce key fusion technologies such as deep learning-based feature extraction and fusion methods, showcase successful application cases in environmental monitoring and disaster assessment, and explore future directions for achieving more efficient and precise holographic perception, providing new insights for the development of the remote sensing field.

INVITED SPEAKER

Wan Nor Shuhadah Wan Nik

Universiti Sultan Zainal Abidin, Malaysia

Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C

Boi: Wan Nor Shuhadah Wan Nik is currently a senior lecturer in the Faculty of Informatics and Computing, University Sultan Zainal Abidin (UniSA), Malaysia. She received a Ph.D. in Computer Science (Distributed Systems) from the University of Sydney, Australia in 2012 before being appointed as a Deputy Director (Infrastructure & Services) at Information Technology Centre, UniSA from the year 2014 - 2017. She has been involved in more than ten research grants and led five national grants in Distributed Systems. Her main research interest includes the area of Computer Networks and Distributed Systems, including Scheduling in Grid / Cloud / Edge and Utility Computing, Wireless Sensor Networks, IoT, Heuristics and Optimization, and Blockchain.

Speech Title: Job Scheduling in Cloud and Fog Computing: A Recent Systematic Review

Abstract: Cloud and Fog computing have emerged as pivotal paradigms in the field of distributed computing, offering flexible and scalable resources for various applications. Efficient job scheduling is a critical factor in optimizing resource utilization and enhancing the performance of these systems. This systematic review aims to provide an up-to-date overview of the state-of-the-art research in job scheduling for Cloud and Fog computing environments. Job scheduling is a complex challenge in Cloud and Fog computing due to their dynamic and heterogeneous nature. The need to balance resource allocation, minimize latency, and enhance energy efficiency poses significant research questions. To address these issues, this article systematically reviews existing literature to identify trends, challenges, and recent advancements in job scheduling strategies. The objectives of this work were to: assess the current landscape of job scheduling techniques in Cloud and Fog computing; analyze the key challenges and trends in job scheduling research; and highlight recent advancements and innovations in this domain, which further provide insights for future research directions in these computing environments. We conducted an advance searching and comprehensive systematic review of peer-reviewed articles (n=48) published in 2023 from Scopus and IEEE databases based on PRISMA framework. Our search and selection criteria ensured the inclusion of relevant studies, and a rigorous analysis was performed to extract key findings and identify emerging trends. By summarizing the state-of-the-art, it offers valuable insights for researchers and practitioners in the field, guiding future research efforts to address the evolving demands of these dynamic computing paradigms.

ONSITE SESSION

- **Topic: Model-based Intelligent Information System Design and Optimization Algorithms**
- **Time: 15:00-17:00, Aug. 28, 2025 | UTC/GMT+8**
- **Location: Room 436 | 会议室 436**
- **Session Chair: Dr. Xin Li, Shandong Computer Science Center (National Supercomputer Center in Jinan), China**
- **HPC1009, HPC3033, HPC2024, HPS1046, HPC2025, HPS2016, HPS2013, HPS2045**

HPC1009
15:00-15:15

Title: Optimization of Sparse Matrix Block LU Factorization: A Hybrid MPI-OpenMP Parallel Method
Author(s): Yujiao Han, Jingshan Pan, Min Tian, Jiwei Xu, Lu Zhang
Presenter: Yujiao Han, Qilu University of Technology, China, Shandong Computing Center (National Supercomputing Center in Jinan), China

Abstract: The solution of sparse linear systems plays a crucial role in scientific computing and engineering applications, with LU decomposition being one of the core algorithms. Its efficiency and scalability directly impact overall computational performance. This paper proposes an optimized algorithm for sparse matrix block LU decomposition based on the MPI-OpenMP hybrid parallel approach, aiming to address the bottlenecks of traditional LU decomposition in high-performance computing environments. By dynamically adjusting block sizes, matrix scaling, and pivot perturbation, combined with OpenMP parallelization and manual loop unrolling, the algorithm significantly enhances computational efficiency and parallel performance. Experimental results show that the proposed algorithm demonstrates excellent scalability and computational performance across various matrix sizes, achieving a speedup ratio of 4.11 for a 1250×1250 matrix. The study provides an efficient and scalable solution for the parallel solution of sparse linear systems, with broad application potential.

HPC3033
15:15-15:30

Title: Runtime prediction model for high performance computing jobs based on dual dimensional feature fusion
Author(s): Yongqiang Tian, Xiaorong Zhang, Fengwei Yang, Wenxiang Yang, Jie Yu, Gang Xian, Liang Lai
Presenter: Yongqiang tian, Southwest University of Science and Technology, China

Abstract: To address the inefficiency in resource allocation caused by insufficient runtime prediction accuracy in High-Performance Computing (HPC) systems, this paper proposes a job runtime prediction model, DD-JRPM, based on dual-dimensional feature fusion. The model integrates temporal and user-dimensional features, leveraging Long Short-Term Memory (LSTM) networks to capture timeseries patterns in job submissions and employing an attention mechanism to dynamically weight critical features. This approach resolves the issue of traditional methods that disrupt temporal dependencies by randomly shuffling data. Experiments conducted on three real-world HPC job log datasets demonstrate that DD-JRPM significantly outperforms baseline models such as linear regression, random forest, and LightGBM in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results highlight that combining dual-dimensional feature fusion with an attention mechanism effectively enhances prediction accuracy, offering novel

HPC2024
15:30-15:45

Title: A Heterogeneous Parallel Optimization Algorithm for Batched Matrix Multiplications on the SW26010-pro Processor
Author(s): Long Zhang, Min Tian, Shui Nai, Xinyu Guo
Presenter: Long Zhang, Qilu University of Technology, China

Abstract: The optimization of batched matrix multiplication has attracted wide attention in current research and industrial practice.This paper presents a heterogeneous parallel optimization algorithm for the SW26010-pro many-core processor, leveraging its master-slave core architecture to accelerate batch dense matrix computations. This method combines the master-slave task allocation mechanism, block computing strategy and asynchronous DMA transmission optimization to improve computing efficiency. Numerical experiments demonstrate that the performance can be improved by 10%-26% when the batch size is 512 compared with the non-optimized version based on double precision floating-point matrix multiplication.

HPS1046
15:45-16:00

Title: An Enhanced Hybrid Neural Network: An Optimized Segmentation Method for Agricultural Remote Sensing Images
Author(s): Guoxun Zheng, Zhengang Jiang
Presenter: Guoxun Zheng, Changchun University of Science and Technology, Changchun Institute of Technology, China

Abstract: In the context of rapid development of agricultural modernization and precision agriculture, farmland remote sensing image segmentation, as a core technology for obtaining key information such as crop distribution and growth status, has a direct impact on the accuracy and efficiency of the science of agricultural resource management decisions. However, the prevalence of small-scale structure, ambiguous region boundaries, and the interference of complex background noise such as shadow interference in agricultural scenes have seriously restricted the practical application effect of segmentation algorithms. To address the above challenges, this study constructs an innovative hybrid neural network architecture based on the U-Net framework that integrates multi-dimensional optimization strategies. The architecture strengthens the consistent expression ability of features at different scales by introducing the generative adversarial training mechanism, and embeds the convolutional block attention module (CBAM), which combines with the hybrid dilated convolution (HDC) structure to enhance the feature capture ability and significantly suppress the artifacts that occur during the segmentation process. The experimental results show that the Mean Intersection over Union (mIoU) and mean pixel accuracy (mPA) of this method reach 52.6% and 64.8%, respectively, which are significantly better than the current mainstream comparison models, and show superior segmentation performance, thus verifying its generalization ability and practical application potential in real scenarios.

HPC2025
16:00-16:15

Title: The OPHGMRES : An Optimized Parallel Householder GMRES Iterative Algorithm
Author(s): ChunQi Dong, Min Tian, Wei Du, ChaoShuai Xu
Presenter: ChunQi Dong, Qilu University of Technology, China

Abstract: GMRES algorithm is one of the most popular Krylov subspace iterative method for solving sparse linear equations. But the Gram-Schmidt GMRES algorithm can not stabilize residual vector fluctuation and have bad numerical stability. We propose a novel OPHGMRES algorithm based on parallel Householder orthogonalization and GMRES iterative method. We improved the Arnoldi process and initial residual of the algorithm by reducing redundant vector calculations in each iteration. This optimization increases the parallel performance of Householder

GMRES. The numerical experiments demonstrate that the OPHGMRES algorithm achieved an average speedup of 1.6 while maintaining the same numerical accuracy as GMRES.

HPS2016
16:15-16:30

Title: Resource Orchestration Method for Industrial Control Software Security Crowdsourcing Platform
Author(s): Fenghua Tong, Jiming Dong, Dawei Zhao, Lijuan Xu, Fuqiang Yu
Presenter: Jiming Dong, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, China

Abstract: Among the many components of industrial control systems, industrial control software is of crucial importance. Its security directly affects the reliability and risk - resistance of the system. The industrial control software security crowdsourcing platform integrates security testers from multiple sources to jointly conduct vulnerability mining and security assessments. However, with the rapid growth in demand for industrial control software security crowdsourcing, designing a resource dynamic allocation strategy for the testers, improving resource scheduling capabilities, and ensuring efficient utilization of resources have become key challenges. Considering the asynchronous, diverse, and dynamic nature of user demands, this paper presents a resource allocation model for multi-server and multi-user scenarios, with constraints including computation, storage, and bandwidth. The model utilizes a distributed optimization approach to effectively address issues such as low resource utilization and long task response times. Experimental results show that the proposed method performs excellently in terms of task completion time and stability, providing strong theoretical and technical support for the design and implementation of industrial control software security crowdsourcing platforms. Experimental verification shows that this method has a significant advantage in the time required to complete tasks, and also demonstrates outstanding performance in terms of stability. These remarkable results provide a strong theoretical basis and technical support for the design concept and ultimate implementation of the crowdsourced security testing platform for industrial control software.

HPS2013
16:30-16:45

Title: Practical Byzantine consensus algorithm based on Kademlia algorithm
Author(s): Peng Zhao, Lei Guang Qiang, Bo Lu, Jiang Fu Yuan
Presenter: Guanglei Qiang, Taiyuan Normal University, China

Abstract: This paper proposes an improved solution (K-PBFT) based on the Kademlia and PBFT consensus algorithms, aiming to solve the problems of poor scalability and high communication overhead of traditional PBFT in large-scale networks. When the number of nodes increases, the communication complexity of PBFT increases quadratically, resulting in a decrease in efficiency. To address this problem, K-PBFT uses Kademlia's node management mechanism to optimize node division by selecting subsets through random functions. It ensures that each subset meets the PBFT operating conditions and allows subset overlap or node omission, thereby improving fault tolerance. In addition, the introduction of a reputa tion points mechanism simplifies the selection of primary nodes. Nodes that meet the reputation standards can skip the submis sion phase, further accelerating the consensus speed. Experimental results show that K-PBFT has been further improved compared with the traditional PBFT consensus algorithm and has better fault tolerance.

HPS2045
16:45-17:00

Title: ADD-TabNet: A Lightweight Tabular Deep Model for Evaluating Access Quality of Crowdtesting Subsystems
Author(s): Fuqiang Yu, Lijuan Xu, Fenghua Tong, Dawei Zhao
Presenter: Fuqiang Yu, National Supercomputer Center in Jinan, China

Abstract: Ensuring the secure and stable integration of external subsystems is critical for the reliability of crowdtesting platforms. In this paper, we present ADD-TabNet, a lightweight yet effective deep learning model designed to evaluate the access quality and security compliance of incoming subsystem modules. Although the model is inherently tailored for access quality assessment, public datasets specifically labeled for this purpose are scarce. To validate its anomaly detection capability, we benchmark ADD-TabNet on the widely-used UNSW-NB15 dataset, where it demonstrates strong performance compared to a suite of baseline models. Furthermore, we contextualize ADD-TabNet within a real-world crowdtesting platform scenario, illustrating how its architecture and output align with practical subsystem access evaluation requirements. Our results indicate that ADD-TabNet is not only efficient but also adaptable for real-time, high-confidence assessments of module integration quality.

ONLINE SESSION 1

- **Topic: Supercomputers and High-performance Computing**
- **Time: 10:00-12:00, Aug. 29, 2025 | UTC/GMT+8**
- **Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C**
- **Session Chair: Assoc. Prof. Othman Ibrahim, Universiti Teknologi Malaysia, Malaysia**
- **HPC1003, HPC1007, HPC1008, HPC1010, HPC2015, HPC2017, HPC2026, HPC3028**

HPC1003
10:00-10:15

Title: Comparison of Pólya-Gamma Method and Metropolis Hastings Algorithm in Logistic Regression Model Problems
Author(s): Chunxia Wang, Jianhua Wu
Presenter: Chunxia Wang, University of Jinan, China

Abstract: Binary data modeling is one of the most common tasks for applied statisticians and econometricians. Although Bayesian methods in this situation have existed for decades, they typically require a high level of familiarity with Bayesian statistics or suffer from issues such as low sampling efficiency. In order to enhance the accessibility of Bayesian models to binary data, we introduced a new latent variable representation based on Pólya-Gamma random variables for a series of common logistic regression models. From these latent variable representations, new Gibbs sampling algorithms applicable to binary, binomial, and multivalued logit models were derived. This article mainly introduces the advantages of using the Pólya-Gamma method in binary logit models compared to the traditional Metropolis Hastings algorithm.

HPC1007
10:15-10:30

Title: SCINet: a high-performance computing infrastructure prototype based on supercomputers and high-speed Internet
Author(s): WenjingHan, Zhenchun Huang, Dexun Chen, Lin Gan, Guangwen Yang
Presenter: Wenjing Han, Tsinghua University, China

Abstract: In recent years, computing power has emerged as a new productivity, playing a crucial role in facilitating various strategic scientific research and industrial tasks. Supercomputing Internet is a novel computing mode proposed under this new situation. It serves as a national high-performance computing (HPC) infrastructure operating in an Internet style, which aims at maximizing the utilization of supercomputing resources across the country, further offering one-stop services for the developers and users of HPC applications in various fields, and finally boosting the development of digital China strategy. However, at current stage, its construction is still immature and faces numerous challenges. To settle this, we develop and release a prototype of Supercomputing Internet, SCINet, which intends to copy with several key issues encountered during the construction of Supercomputing Internet, including resource interconnection, application service platform construction, special support for typical application areas, and standardization promotion. These efforts significantly demonstrate the technical feasibility of implementing the Supercomputing Internet.

HPC1008
10:30-10:45

Title: A Novel Instruction Scheduling Approach Based on Multi-Choice Model
Author(s): Jinwei Zhao, Jue Wang, Wei Wu, Fei Wang
Presenter: Jue Wang, Jiangnan Institute of Computing Technology, China

Abstract: In recent years, the advent and successive enhancements of deep learning methodologies have witnessed a consistent surge in the precision of code performance forecasting. For example, the predictive accuracy of code performance

for systems like Ithemal has surpassed the heuristic-based static code performance estimation techniques that are commonly employed by conventional compilers. It is evident that enhanced accuracy in performance prediction models can serve to refine instruction scheduling algorithms within compilers, thereby enabling the discovery of more optimal instruction sequences. However, several challenges persist in their practical application. The predominant challenge lies in the fact that, it is the prediction accuracy of relative performance, i.e., the performance differences under difference instruction arrangements, rather than absolute performance that instruction scheduling relies on. In other words, the prediction accuracy of relative performance matters more in instruction scheduling. To this end, we have proposed a multi-choice model to predict the optimal basic block from a pool of alternatives and designed a novel instruction scheduling technique based on the model. Initially, we acquired a large-scale basic block performance dataset by utilizing compiler instrumentation techniques. Then, we converted the basic block performance labels into multi-choice labels and trained a multi-choice model using the data. During compilation, multiple basic blocks are generated by random scheduling, and the optimal basic block under different instruction arrangements is predicted using the multi-choice model, thereby completing the instruction scheduling. Experiments on the Shenwei platform have shown that, based on the multi-choice model, the instruction scheduling achieves a performance improvement of 3.1\% on the test set, with -O2 as the baseline.

HPC1010
10:45-11:00

Title: CTIA: Optimizing Requests Processing by Comparing Tags in Advance for L2 Caches in GPUs
Author(s): Bingchao Li, Yuchen Zhu, Xiaohui Li, Jizeng Wei
Presenter: Yuchen Zhu, Civil Aviation University of China, China

Abstract: Within the memory hierarchy of GPUs, L1 cache is dedicated to individual streaming multiprocessor (SM), whereas L2 cache is shared among all SMs. Requests that fail to find the desired data in L1 caches are subsequently routed to L2 caches via the network on chip (NoC). However, when no space, such as miss buffers of L1 caches and input buffers of L2 caches, is available to accommodate these missing requests prior to accessing the L2 cache, the requests are impeded, leading to increased stalls and, consequently, suboptimal GPU performance. In this paper, we introduce comparing tags in advance (CTIA) to mitigate the number of stalls incurred by requests destined for L2 caches. Specifically, we duplicate the tag array of L2 cache and position it immediately after the output port of NoC. Consequently, requests arriving at the DRAM channel associated with an L2 cache can promptly access the duplicated tag array. This allows us to predict in advance whether a request will hit in L2 caches. Requests predicted to hit in L2 caches are processed as conventional, whereas those predicted to miss are directly routed to the DRAM access path without traversing L2 caches. As a result, CTIA alleviates resource contention before accessing L2 cache by cost-effectively optimizing the handling of potential missing requests, thereby reducing stalls caused by resource shortages. Experimental results demonstrate that CTIA enhances the geometric-mean performance of GPUs by 23%.

HPC2015
11:00-11:15

Title: SW-M3D-K: Research on Parallel Optimization of M3D-K Program Based on Sunway Bluelight II Supercomputer
Author(s): Yingxuan Shao, Tao Liu, Baofeng Gao, Qiang Guo, Ying Guo, Jingshan Pan
Presenter: Yingxuan Shao, 1.Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China 2.Shandong

Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Jinan, China

Abstract: With the increasingly serious environmental problems and the exhaustion of traditional fossil energy resources, nuclear energy has become increasingly prominent as an important role in future energy strategies. The use of high-performance computers for nuclear fusion simulation calculations has become an inevitable trend. The M3D-K program is a typical magnetohydrodynamic-kinetic hybrid model program. The model simulates the plasma system by integrating the M3D (Multilevel 3D) program and the high-energy particle module. Leveraging the architecture of the Sunway multi-core processor and the Sunway Bluelight II supercomputer, this paper introduces a fine-grained and versatile two-level parallelization strategy. The approach integrates process-level and thread-level parallelization, aiming to optimize computational performance and scalability for large-scale scientific simulations. In thread-level parallelization, a modular data flow transmission strategy is first proposed to solve the problem of discontinuous addresses after array copying in Fortran, which greatly improves the data transmission speed in Fortran. Secondly, we designed a CPEs mapping network and combined it with dependency elimination for acceleration. In process-level parallelization, a virtual vertex and small-scale tree communication algorithm is proposed based on the regional decomposition technology of the adaptive grid, which makes full use of the hardware of the Sunway Bluelight II supercomputer and improves the data transmission speed and calculation efficiency. For the first time, the M3D-K program was ported and optimized to the Sunway series supercomputer, enriching the application ecology of the Sunway series supercomputer. Compared with the original program, two-level parallelism can achieve a speedup of 18.13 times. In addition, our parallelization can run on 2048 processes of the Sunway Bluelight II supercomputer, with good scalability.

HPC2017
11:15-11:30

Title: The Research on GPU-Based Relational Join Algorithms
Author(s): Kou Wenwei, Du Jinlian, Su Hang, Luo Fangyuan
Presenter: Kou Wenwei, Beijing University of Technology, China

Abstract: Accelerating relational join operations using GPU is currently one of the hot topics in the field of database research. However, most existing GPU-based relational join algorithms assume that data can be entirely loaded into GPU graphics memory, neglecting the overhead of data transfer between disk, main memory, and GPU graphics memory. Additionally, these algorithms fail to fully leverage the collaborative computing capabilities of CPU and GPU, resulting in suboptimal overall performance. This paper focuses on large scale relational join operations and proposes an efficient GPU-based relational join algorithm. From the perspective of load balancing, a collaborative computing mechanism between the GPU and CPU is designed, fully utilizing the computational resources of both to enhance operational efficiency. By implementing a sequential cyclic join mechanism and data replacement strategy, the algorithm reduces the number of data transfers between main memory and GPU graphics memory, thereby enhancing overall efficiency. The performance of the proposed algorithm is experimentally validated by comparing with CPU Join and GPU Join operations. The experimental results demonstrate that the proposed algorithm performs well on large scale relational join operations, achieving speedups of 2.1X and 1.1X compared to CPU Join and GPU Join, respectively. This research provides new insights into leveraging GPU for large scale relational join operations and holds significant practical implications for improving the efficiency of large scale database queries.

HPC2026
11:30-11:45

Title: Parallel Optimization of Plasma Fluid Simulation Software Based on Sunway Bluelight II Supercomputer
Author(s): Yinhui Huang, Tao Liu, Baofeng Gao, Ying Guo, Jingshan Pan
Presenter: Yinhui Huang, 1. Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China; 2. Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Jinan, China.

Abstract: With the development of nuclear fusion technology, the demand for computer software simulation is also increasing, and the application of supercomputers has become an inevitable trend. Tokamak is a device used to realize controlled nuclear fusion, mainly for studying the physical properties of high-temperature plasma. The basic principle of a tokamak device is to utilize a strong magnetic field to confine a high-temperature plasma and prevent it from coming into contact with the walls of the device, thereby realizing a nuclear fusion reaction. Sunway Bluelight II supercomputer, as a new generation of supercomputer, has strong computing power, and there is a good research prospect to do nuclear fusion software optimization on this machine. Based on the Sunway Bluelight II supercomputer and the unique architecture of the Sunway many-core processor SW39000, this paper proposes a heterogeneous general parallel optimization method for the open-source plasma fluid simulation program BOUT++. Parallel optimization of the BOUT++ plasma-fluid simulation program is accomplished by assigning tasks to the computational cores within the core group, grouping CPE array, and dividing the work between CPE and MPE. This paper enriches the application ecology of the Sunway supercomputer. The experimental results show that the optimized program in this paper achieves to 6.87x speedup using a single core group compared to the original program and extends the program to 66,560 computing cores (1024 MPEs and 65,536 CPEs) with good strong and weak scalability.

HPC3028
11:45-12:00

Title: Archs: A WebAssembly Runtime for Cross-host Heterogeneous Computing in Serverless
Author(s): Aoyuan Sun, Yue Tu, Yuhao Gu, Chunyu Chen, Jiangsu Du, Xianwei Zhang
Presenter: Aoyuan Sun, Sun Yat-sen University, China

Abstract: WebAssembly (Wasm) is a new binary instruction format and run- time environment capable of executing both client side and server side workloads. With its numerous advantages, including small soft- ware artifacts sizes, efficiency, high performance, easy portability, and compatibility with the most popular programming languages today, it has the potential to revolutionize serverless computing as an alternative to traditional container. While successfully enabling high performance across platforms, Wasm itself specifies nothing for heterogeneous computations, such as GPUs and FPGAs, which plays an important role in today's computing system. To mitigate this issue and enabling cross-platform heterogeneous computing, we explore the use of an another neutral binary format, SPIR-V of OpenCL, in the context of Wasm. From the other perspective, Wasm can empower the cross-host capability for OpenCL/SPIR-V as well. This paper introduces WasmCL - our Wasm runtime(embedder) be- ing capable of executing OpenCL/SPIR-V heterogeneous programs to utilize the power of accelerators. Extensive experimental results show that WasmCL can achieve similar performance

to containerd, implying a promising future for a solution of cross-platform heterogeneous computing by integrating Wasm and OpenCL/SPIR-V.

ONLINE SESSION 2

- **Topic: Emerging Network Technologies and Information Security**
- **Time: 14:00-16:00, Aug. 29, 2025 | UTC/GMT+8**
- **Zoom: https://us02web.zoom.us/j/86197421409 password: HP3C**
- **Session Chair: Dr. Nur Haliza Binti Abdul Wahab, University of Technology Malaysia, Malaysia**
- **HPC1005, HPS1029, HPC5047, HPS2019, HPS2018, HPS2020, HPS2021, HPS2023**

HPC1005
14:00-14:15

Title: A General Framework for Reproducible Parallel Preconditioned Krylov Methods Using Three BLAS Variants
Author(s): Xiaojun Lei, Tongxiang Gu, Xiaowen Xu, Stef Graillat
Presenter: Xiaojun Lei, Graduate School of Chinese Academy of Engineering Physics, China

Abstract: Krylov subspace methods are crucial techniques for solving linear systems. To effectively tackle large-scale linear systems, parallelism techniques are frequently utilized. However, the use of parallelism can increase the non-associativity of floating-point operations, potentially resulting in non-reproducibility of the computations. In order to address the issue of non-reproducibility in parallel preconditioned Krylov subspace iterative methods, we investigated the sources of irreproducibility and solutions to achieve reproducibility. The main reason for the irreproducibility in iterative methods is attributed to distributed dot products. To tackle this, we investigated three reproducible Basic Linear Algebra Subprograms (BLAS) as replacements for standard distributed dot products. We utilized OzBLAS, ExBLAS, and ReproBLAS to ensure reproducibility in iterative methods. We combined Jacobi preconditioners and tested and verified on three iterative methods: CG, BiCGSTAB and GMRES. Numerical experiments demonstrate that the iterative method based on OzBLAS is reproducible in a multi-threaded environment. The iterative method based on ExBLAS and ReproBLAS achieves reproducibility in a multi-process environment. Our evaluation indicated that ReproBLAS performed well in terms of performance, robustness, and parallelism. Therefore, we employ a reproducible iterative method based on ReproBLAS to solve the large-scale algebraic equations resulting from the discretization of the two-dimensional three-temperature heat conduction equation, which arises from practical application problems in the field of laser fusion radiation hydrodynamics. Numerical experiments demonstrate the reproducible solution of the three-temperature heat conduction algebraic equations.

HPS1029
14:15-14:30

Title: A remote sensing small target ship rotation detection method based on an improved YOLOv11 algorithm
Author(s): Zhen Yu, Yuan Zhang, Lei Kou, Junhe Wan, Weiqiang Hou
Presenter: Zhen Yu, Qilu University of Technology (Shandong Academy of Sciences), China

Abstract: Ships in remote sensing images have the characteristics of arbitrary directions, complex backgrounds, diverse scales, and relatively small targets, which can easily lead to low accuracy, false detection, and missed detection in rotation detection. Because of the above situation, this paper proposes a remote sensing small target ship rotation detection method based on an improved YOLOv11 model, named Ship-YOLO. Firstly, in the backbone network, the spatial pyramid pooling fast convolution mixed squeeze attention mechanism (Spatial Pyramid Pooling Fast Convolution Mixed Squeeze Attention, SPPFCMSA) is constructed to significantly

improve the feature representation ability of small target ships through channel and spatial dual-dimensional attention weighting; at the same time, a simple maximum spatial grouping enhancement module (Map Spatial Group-wise Enhance, MSGE) is designed to reduce the risk of model overfitting. Then, in the neck network, the Bottleneck in the C3k2 module of the P3 and P4 layers is replaced with StarABlock to form a C3k2_StarA structure, which solves the problems of detail weakening and missed detection in the feature extraction process. Last, in the head network, a Cross-Scale Multi-Head Unit OBB Head (CMUOH) is proposed to effectively handle the detection tasks of ships at extreme rotation angles and different scales. Experimental results show that the detection accuracy (mAP50) of the Ship-YOLO algorithm reaches 82.4%, 98.7%, and 97.4% in the HRSC2016, SSDD+, and ShipRSImageNet dataset, respectively, which are 4.7%, 1.1%, and 1.3% higher than the YOLOv11 algorithm, respectively.

HPC5047
14:30-14:45

Title: OpenMP Offloading on AMD and NVIDIA GPUs: Programmability and Performance Analysis
Author(s): Ezhilmathi Krishnasamy, Pascal Bouvry
Presenter: Ezhilmathi Krishnasamy, University of Luxembourg

Abstract: The introduction of Graphics Processing Units (GPUs) to scientific computing has led to the development of various programming models designed to maximize their computational capabilities. Over the past few decades, GPU vendors have matured, resulting in a diverse array of GPU programming models and continuous advancements in compiler technologies that support these developments. Among these options, Open Multi-Processing (OpenMP) Offloading has emerged as a particularly promising programming model that targets GPUs from NVIDIA, AMD, and Intel. This study investigates the effectiveness of OpenMP Offloading on NVIDIA (H100) and AMD (MI250X) GPUs, which utilize different interconnect technologies—PCIe Gen5 for NVIDIA and Infinity Fabric for AMD. Using vendor-specific compiler toolchains, namely NVHPC (version 23.7-0) for NVIDIA and Cray Clang (version 17.0.1) for AMD, we evaluate performance across key linear algebra operations, focusing on kernel execution efficiency and data transfer strategies between the host CPU and the GPU. Additionally, this work includes a comparative analysis of OpenMP Offloading against cuBLAS and hipBLAS. Our results indicate that OpenMP Offloading offers performance comparable to the best numerical library versions for BLAS level 1 and 2 operations on H100 and MI250X (except matrix-vector multiplication). Furthermore, Infinity Fabric provides approximately 1.5 times the bandwidth of PCIe Gen5, according to our measurements. We also recommend utilizing low-level APIs to achieve improved performance for data movement between the CPU and GPU.

HPS2019
14:45-15:00

Title: An Efficient IoT Device Classification Method Based on the Integration of Variational Autoencoder and XGBoost
Author(s): Xin Zhang, Weite Liu, Jie Zhang, Ning Li, Le Ren, Zihan Liu
Presenter: Bingyang Shan, Qilu University of Technology (Shandong Academy of Sciences), China

Abstract: This paper proposes an effective IoT device classification method combining Variational Autoencoder (VAE) and an optimized XGBoost model to address challenges like large-scale data processing and sample imbalance. First, feature selection is improved by integrating mutual information analysis and domain knowledge. VAE is used for feature extraction, capturing latent representations from network traffic data. The XGBoost classifier is optimized using techniques like gradient clipping and model pruning to improve training efficiency and reduce computational overhead. This method boosts classification accuracy,

handling data imbalance and noise effectively, with experimental results showing 90% accuracy and F1 score. The proposed approach outperforms traditional algorithms in large-scale IoT device classification tasks

HPS2018
15:00-15:15

Title: Risk Assessment and Response Mechanism for University Network Information Security Based on AHP-FCE Model
Author(s): Aiwei Jiang, Xinping Liu
Presenter: Xinping, Liu, Eversec Technology CO., LTD, China

Abstract: With the acceleration of educational informatization, university network information security faces multifaceted challenges from open environments, complex architectures, and emerging threats. To address the limitations of existing research that relies on qualitative analysis and lacks dynamic quantitative evaluation, this study proposed a comprehensive assessment method for university network information security based on the Analytic Hierarchy Process-Fuzzy Comprehensive Evaluation (AHP-FCE) model. The AHP method was employed to construct a four-dimensional indicator system encompassing technical protection capability (37.55% weight), management mechanism integrity (38.56%), personnel security awareness (15.45%), and policy compliance (8.44%). The fuzzy comprehensive evaluation (FCE) method was then integrated to quantify risk levels. An empirical study applying the AHP-FCE model to evaluate a university's network information security revealed that the institution's overall security status was categorized as "good" (80.39/100). Specifically, technical protection capability scored 78.64/100, constrained by weak disaster recovery systems and delayed vulnerability remediation. The management mechanism scored 81.81/100 but required increased security budget allocation. Personnel awareness scored 80.03/100, necessitating enhanced mechanisms for proactive incident reporting. Policy compliance achieved the highest score 82.36/100, demonstrating regulatory adherence but requiring dynamic updates to align with evolving laws. By integrating subjective and objective evaluations, this study achieved precise multi-dimensional risk identification through "standard-driven, hierarchical weighting, and fuzzy quantification." The findings provide a scientific basis for universities to optimize technical protection priorities, refine management closed-loop systems, and adapt dynamically to compliance requirements. This approach facilitates the construction of a full-cycle security ecosystem featuring "monitoring-warning-response-feedback" mechanisms, promoting the development of resilient security systems in the context of higher education's digital transformation.

HPS2020
15:15-15:30

Title: A Novel Security Assessment Model for Defense Systems Based on Temporal Attention Mechanisms
Author(s): Xin Zhang, Le Ren, Yunxiao Wang, Lina Zhao, Zihan Liu, Weite Liu
Presenter: Yansong Li, Qilu University of Technology (Shandong Academy of Sciences), China

Abstract: In the evolving landscape of cyber threats, traditional security evaluation methods often fall short in adapting to dynamic attack scenarios and providing real-time feedback on the effectiveness of defense systems. This paper proposes a Dynamic Defense Effectiveness Scoring Method based on a temporal attention mechanism, which integrates multi- dimensional time-series data, convolutional feature extraction, and contextual embedding of attack scenarios. The proposed method leverages convolutional layers to capture short-term patterns in defense responses and employs a multi- head attention mechanism to emphasize critical time steps. Additionally, contextual information, including attack type, intensity, and system configuration, is embedded to enhance situational awareness. The final scoring is generated through a fully connected network, offering a real-time,

interpretable, and quantifiable evaluation of defense effectiveness. Experimental results demonstrate the method's superior accuracy and adaptability in various attack scenarios, providing actionable insights for security managers to optimize their defense strategies.

HPS2021
15:30-15:45

Title: Bi-mark: A copyright marking framework for AI-generative-text
Author(s): Liu Gongshen
Presenter: Boyuan Yu, Capital Normal University High School, China

Abstract: There are currently two kinds of watermarking method for AI generative text (AIGT), the model watermarking method and the generative text watermarking method. Independently each method marks one copyright. In many application scenarios, the text is generated collaboratively by AI tool and its user. That means two copyrights need to be marked. So, a single kind of watermarking method can not meet the application requirement. Therefore, this paper designs a dual watermark framework named Bi-mark for AIGT, which combines the model watermarking method and the text watermarking method to meet application scenarios. There are many choices of these two kinds of watermarking method in the framework, in condition that they do not interfere each other. A backdoor method and a post-model method are acceptable. Experiments show that the framework, based on Roberta , is effective to verify the two watermarks of the AI model and the AI user. Meanwhile, the AI-generative text is good of semantic consistency.

HPS2023
15:45-16:00

Title: Prediction Method for Assert Statements in Test Cases Based on Deep Learning
Author(s): Hua Huang, Ming Li, Jianfei Chen, Xingfang Cheng, Jie Zhang
Presenter: Wenkang Xiang, Qilu University of Technology (Shandong Academy of Sciences), China

Abstract: With the increasing prominence of cybersecurity, the technology for secure and controllable cyberspace attack simulation has become critical. However, developing high-quality test cases not only takes considerable time but also poses significant challenges. Although existing methods of automatic generation have somewhat alleviated the manual burden, they often fail to produce accurate assert statements, which impacts the effectiveness of the test results. In this study, a novel approach utilizing deep learning is developed to predict assert statements: Fused Encoder with Transformer Decoder (FETD). FETD first utilizes the fine-tuned CuBERT model to generate input sequence representations, which are subsequently merged with every encoder and decoder layer in the Transformer architecture through attention-based feature fusion. This effectively combines CuBERT's semantic understanding of code with Transformer's context modeling capabilities, resulting in a more rich and accurate representation. Additionally, we propose a correction method to further refine the formatting of the generated assert statements. Evaluation data indicate that compared to existing approaches, the proposed FETD achieves enhanced performance in assert statement generation, with significant improvements observed across metrics including BLEU, GLEU, and perfect prediction counts. Furthermore, ablation experiments prove that: the accuracy of generated assert statements is significantly improved by the correction method; By integrating the CuBERT pre-trained model, the prediction method achieves significantly higher accuracy; And the use of the fused encoder further increases the accuracy of generated assert statements.

NOTES