Architectural Design and Key Technologies of Cross-Regional Heterogeneous Computing Resource Interconnection Platform

Bing Sun[1], Guohuan Song[2,3]
[1]*school of information and business management, Dalian Neusoft University of Information, Dalian, China, sunbing@neusoft.edu.cn*
[2]*college of computer science, Zhejiang Normal University, Jinhua, China, guohuan.song@zjnu.edu.cn*
[3]*Beijing Northern Computing Intelligence Technology Co., Ltd., Beijing, China*

***Abstract*—The exponential growth in intelligent computing demand has created significant structural imbalances in computing power supply, characterized by geographical disparities, architectural heterogeneity, and management silos. This paper introduces a cross-regional heterogeneous computing resource interconnection platform based on a layered, decoupled cloud-native architecture. Its core contributions are two key technologies: a computing power identifier system that establishes a unified semantic framework for diverse resources, and a heterogeneous resource adapter that enables seamless integration through a dynamic, containerized plugin mechanism. The platform has been implemented and validated in a regional trial. Results demonstrate its capability for large-scale integration of heterogeneous resources, achieving a cross-domain scheduling success rate exceeding 99%. These findings validate the platform's effectiveness in building a scalable, efficient, and reliable computing interconnection system, providing critical technical support for the evolution of future computing networks.**

Keywords—heterogeneous computing, computing power networks, computing power identifier, heterogeneous resource adapter, dynamic plugin management

I. INTRODUCTION

In the digital economy era, computing power has become a crucial production factor and core infrastructure driving social progress and industrial upgrading. With AI technology breakthroughs, intelligent computing demand has grown exponentially, enabling computing-centric service models to achieve ubiquitous network deployment [1]. However, significant structural imbalances exist between computing supply and demand. First, geographical disparities persist with uneven resource distribution across regions, causing varying accessibility from user perspectives [2]. Second, heterogeneous computing architectures have become increasingly prominent, evolving from CPUs and GPUs to specialized ASIC chips, creating a complex computing ecosystem. Concurrently, computing resources are distributed among different cloud providers and data center operators, forming management silos [3]. Although cloud computing provides resource pooling, its implementation remains restricted to individual providers, lacking effective cross-domain and cross-architecture interconnection mechanisms. To optimize computing resource allocation, this paper designs a cross-regional heterogeneous computing resource interconnection platform, proposing a cloud-native system architecture, delineating functional interaction mechanisms across hierarchical layers, and investigating two key technologies: computing power identifier resolution and heterogeneous resource adaptation.

II. RELATED RESEARCH

A. Computing Power Network

With global digital transformation advancement, smart device proliferation and novel application scenarios have created unprecedented computing demands. Computing Power Networks (CPN) represent an innovative architecture for efficiently integrating and optimally scheduling distributed heterogeneous computing resources to address high-performance computing requirements, emerging as a prominent research area attracting significant academic and industrial attention [4-5].

The core CPN research focuses on achieving deep integration and collaborative scheduling between fundamentally heterogeneous computing and network resources. The evolutionary trajectory shows progression from "cloud-network synergy" toward "computing-network integration." Early research concentrated on internal resource allocation and task scheduling within cloud data centers, emphasizing enhanced resource allocation flexibility and efficiency through Software-Defined Networking (SDN) technology. SDN established programmable network resource management foundations by decoupling control and data planes, demonstrating significant advantages in cloud resource orchestration [6]. Industry has proposed centralized, distributed, and hybrid implementation approaches. Related research developed computing power network orchestration management systems using hybrid approaches, achieving precise business traffic routing control through collaborative computing and network scheduling mechanisms [2]. Furthermore, the Computing First Network (CFN) concept connects distributed computing clusters via CFN routers, making routing decisions based on service and data IDs while supporting advanced features like data prefetching, offering novel approaches for computing power routing [1]. Research indicates heterogeneous computing power network architectures significantly enhance system performance and energy efficiency, establishing crucial foundations for future intelligent computing development [7-8].

B. Heterogeneous Resource Pooling

Heterogeneous resource pooling represents the core technology for actualizing computing network value. The primary objective abstracts geographically dispersed, architecturally diverse computing devices into unified resource pools while providing efficient scheduling strategies. Early cloud computing research on virtual machines and task scheduling established theoretical foundations for resource pooling [9]. Integrating heterogeneous resources into unified, dynamically manageable pools significantly optimizes resource utilization and system performance. Numerous studies demonstrate this approach's effectiveness across various domains. Researchers investigated methods reducing cluster configuration scale through heterogeneous workloads while increasing parallel workload requests [10]. Wang et al. developed multi-resource allocation mechanisms substantially improving resource utilization and reducing job completion times [11]. Research indicates FPGA resource pools accelerate job completion up to sevenfold [12]. This pooling approach facilitates system functionality expansion while accommodating technological evolution [13]. Combined with predictive scheduling techniques, enhanced task scheduling algorithms optimize resource deployment and execution efficiency [14].

C. Computing Power Identification and Description Technology

Computing power identification and description technology enables unified management and collaborative scheduling of heterogeneous computing resources, providing fundamental support for resource discovery, matching, scheduling, and trading. Standardization and precision levels function as core indicators of computing power network maturity. This domain remains in preliminary exploratory phases. Researchers proposed innovative identification methods, including URL-based computing resource identification systems [15], and approaches combining identity resolution with trusted authentication for collaborative heterogeneous resource allocation [16]. To enhance resource description richness and security, blockchain technology strengthens system trustworthiness. Related research proposes blockchain-based heterogeneous resource allocation schemes, recording critical network information on blockchain to ensure security and transparency throughout resource allocation processes [17]. Core objectives encompass: generating unique universal computing resource identifiers; enabling trusted authentication of heterogeneous computing resources; and establishing unified resource interaction and sharing platforms.

In summary, computing power network architecture research has progressed from cloud-network synergy to computing-network integration, demonstrating evolutionary trends toward hierarchical, distributed, and intelligent development. Substantial advancements have been achieved in heterogeneous resource pooling, with scheduling strategies evolving from conventional fairness-based algorithms to efficient intelligent mechanisms. Although computing power identification and description technologies remain nascent, fundamental requirements encompass measurement standardization, information distributability, and transaction trustworthiness. However, practical implementation of wide-area heterogeneous computing resource interconnection lacks comprehensive platform architecture integrating standardized identification, unified heterogeneous adaptation, and multi-mode management mechanisms. This study proposes a technical framework incorporating identification resolution systems and multi-mode management adaptation. By designing layered, decoupled cloud-native architecture to enhance resource pooling efficiency, we provide critical technical support for constructing heterogeneous computing networks.

III. PLATFORM ARCHITECTURE DESIGN

A. Design Goals and Principles

The platform design aims to achieve four core objectives: heterogeneous compatibility, cross-domain scheduling capability, efficiency and elasticity, and security and reliability. For heterogeneous compatibility, the platform supports unified access and management of diverse computing resources—including general-purpose, intelligent, and supercomputing—by abstracting architectural differences through standardized interfaces. Cross-domain scheduling capability requires establishing unified computing resource views across geographic regions and management domains, enabling global collaborative scheduling for strategic cross-regional resource allocation. Efficiency and elasticity objectives maintain scheduling response times at second-level, support horizontal scaling to accommodate fluctuating loads, and enhance resource utilization exceeding 80% [18]. Security and reliability requirements mandate comprehensive security throughout resource access, data transmission, and operational processes, with system availability maintained at 99% or higher [19].

To achieve these objectives, the platform architecture adheres to design principles: employing microservices architecture for loose coupling between services, enabling independent development, deployment, and scaling; utilizing containerization for agile deployment with Kubernetes-based orchestration ensuring system elasticity; and implementing frontend-backend separation and application-data separation architectures to enhance system robustness and maintainability. These principles ensure stable operation and continuous evolution capabilities within complex heterogeneous environments.

B. Core Function Design

The core functions center on computing resource aggregation and efficient integration, providing foundational data support for resource monitoring and service capability development. Through unified API interfaces for resource provisioning and workflows, the platform enables rapid onboarding and differentiated adaptation of ubiquitous, distributed, multi-source heterogeneous computing resources. It constructs standardized computing resource pools, enhancing task execution efficiency, optimizing transmission performance, and achieving standardized management, access protocols, and optimized resource allocation. The platform offers multiple functions including identifier resolution, supply management, computing power aggregation, and computing power access, as illustrated in Fig. 1.

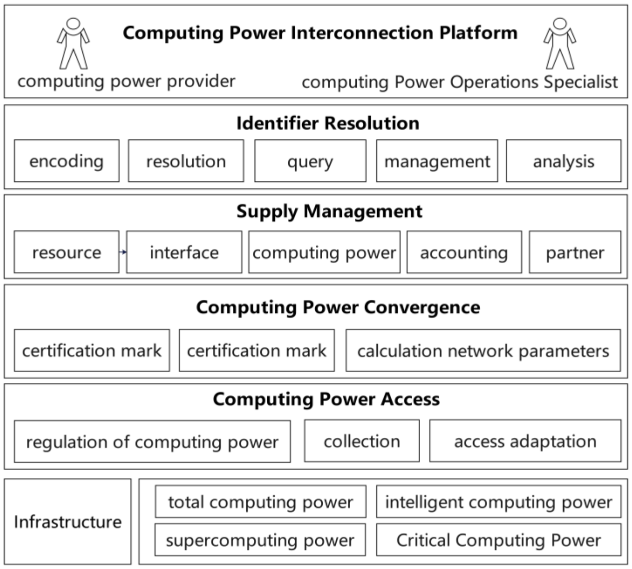

Fig. 1. Core functions of the platform

- The identifier resolution module functions as a unified identification system, utilizing multidimensional encoding for computing power identification protocols, incorporating essential attributes including geographical location, industry classification, resource type, and chip specifications. Through encoding, resolution, querying, and management functions, this module transforms non-standardized computing resources into unified format, facilitating precise resource localization, supply-demand matching, and global identifier uniqueness.

- The supply management module delivers comprehensive lifecycle management of computing resources, comprising resource management, interface management, computing metrics, accounting and settlement, and partner management components. Resource Management facilitates visual monitoring and control of diverse heterogeneous computing power, supporting resource addition, modification, and deactivation operations. Computing Power Measurement implements differentiated measurement frameworks for general, intelligent, and supercomputing, enabling unified resource normalization. Partner Management ensures regulatory compliance and service quality assurance of computing power suppliers through qualification verification and credit scoring mechanisms.

- The computing power aggregation and access module is designed for heterogeneous resource adaptation, featuring three management modes. Deep integration mode enables automated provisioning through OpenAPI interfaces. Static flow mode facilitates manual provisioning via work order systems. Cloud-Native management mode achieves computing power network integration through plugin/proxy mechanisms. The module supports multiple data collection methodologies—including manual workflow, online form, API data, and plugin data collection—ensuring real-time data acquisition and accuracy. The computing power access component utilizes heterogeneous adapter technology to transform proprietary APIs into standardized format, completing comprehensive resource integration and management.

C. Technical Architecture Design

The platform implements a five-layer decoupled technical architecture, featuring clearly defined responsibilities and standardized interfaces at each hierarchical level, as illustrated in Fig. 2.

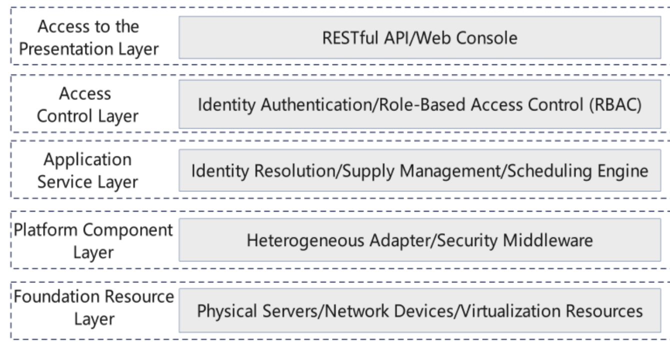

Fig. 2. General technology architecture diagram

- Foundation resource layer. The foundation resource layer functions as physical infrastructure, encompassing distributed heterogeneous computing resources including GPU server clusters, supercomputing centers, and AI computing centers. This layer utilizes virtualization technology to abstract physical resources into logically unified resources for centralized management, providing standardized access interfaces to upper layers.

- Platform component layer. The platform component layer implements modular design supporting upper-layer applications. Critical components include: heterogeneous adapters for standardized API conversion between computing power providers; security middleware for data encryption and access control; message queues for asynchronous service communication; and caching components for performance optimization.

- Application service layer. The Application Service Layer functions as the core business logic layer, implementing microservices architecture. Essential services comprise: identity resolution service for identifier encoding, resolution, and administration; supply management service for resource instantiation and lifecycle management; computing power aggregation service for supplier registration and resource reporting; and scheduling engine service for cross-domain task scheduling using intelligent algorithms [20]. services communicate via RESTful APIs, supporting independent deployment and horizontal scaling.

- Access control layer. The access control layer functions as a security gateway, implementing unified identity authentication and permission management based on Role-Based Access Control (RBAC) model. This layer delivers comprehensive security capabilities including API rate limiting, security auditing, and intrusion detection to ensure secure and controlled platform access.

- Access and presentation layer. The access and presentation layer provides web management portal and open API interfaces. The management portal facilitates resource monitoring, scheduling policy configuration, and operational data analysis, while RESTful API interfaces deliver standardized integration capabilities for third-party systems, promoting computing power ecosystem development.

IV. Key Technology Exploration

A. *Computing Power Identification and Resolution*

The computing power identification and resolution system establishes the semantic foundation for enabling cross-regional interconnection of heterogeneous computing resources. The fundamental principle involves constructing a standardized framework for describing and identifying computing resources. Through two essential processes—identification encoding and resolution—the system transforms heterogeneous computing resources into unified, recognizable digital identity representations.

- Computing power identifier. Analogous to how the Internet uses a Uniform Resource Identifier (URI) to locate files, the Computing Power Identifier (CPI) is designed to match computing tasks with suitable resources. The CPI is a field-based string code composed of two main parts: a unique identifier (prefix) and a computing power path (suffix), as illustrated in Fig. 3.

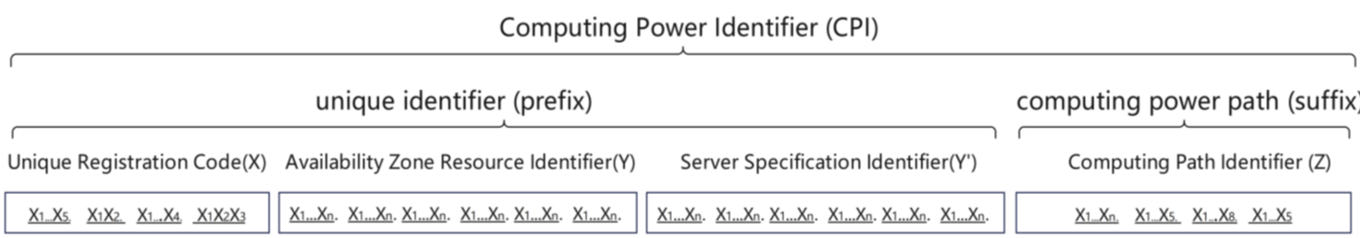

Fig. 3. Computing power identification structure

The identifier prefix is structured into three key segments:

a) X (Unique Registration Code): This code uniquely identifies the resource's provider and location, comprising four fields: enterprise code, industry code, city code, and availability zone code.

b) *Y (Availability Zone Resource Identifier):* This segment provides a macro-level overview of the resources within an availability zone. It includes six fields: resource type, service type, total compute capacity, total storage, network bandwidth, and chip model and quantity.

c) *Y' (Server Specification Identifier):* This segment details the specific server configurations. It consists of six fields: server type, CPU model, CPU core count, memory, storage, and server quantity.

The computing power path complements the prefix by pinpointing the specific route to the required resources. This path is defined by the Computing Path Identifier (Z), which specifies the exact location of a computing resource, down to the individual chip. It is composed of four fields: computing internet address, chip type, chip model, and chip unique ID.

The computing power identifier employs a hierarchical composite structure to ensure global uniqueness and rich semantic information. This design not only enables granular management of heterogeneous computing resources but also supports future field additions through reserved expansion bits. This approach guarantees standardized core information while providing flexibility for business expansion.

- Identifier resolution system. Identifier resolution is the process of parsing a computing power identifier to extract its embedded semantic information. This transforms the identifier into a structured set of resource attributes, which are stored on the platform to enable external query and invocation services. To implement this, the identifier resolution system employs a distributed architecture to deliver high-performance resolution services. Upon receiving a computing power identifier, the resolution engine executes a multi-stage process. First, it performs format validation to verify the identifier's integrity and compliance. Next, it conducts segment-based decoding, parsing each field into interpretable information according to predefined protocols. Finally, it generates a structured resource description file containing resource attributes, service capabilities, and status information. This process is supported by an identifier management system that provides three key functions: data dictionary maintenance, version management, and statistical analysis. The data dictionary ensures consistent resolution by defining the precise semantics and formats of all identifier fields. Version management facilitates incremental updates and historical traceability. Furthermore, the statistical analysis module generates multi-dimensional reports on resource distribution and utilization, providing critical data for resource optimization.

B. Flexible Heterogeneous Resource Adapter

To abstract differences in API interfaces, resource models, and asynchronous event mechanisms across heterogeneous infrastructures, the platform introduces the Heterogeneous Resource Adapter (HRA). This adapter implements bidirectional north-south conversion and hot-swappable plug-in mechanisms. The design objective is to mask infrastructure variations while providing upper layers with unified, homogeneous API interfaces. This approach enables centralized management of diverse computing resources and delivers robust technical support for cross-regional computing interconnection.

- Bidirectional North-South conversion mechanism. The bidirectional conversion mechanism between northbound and southbound interfaces is the pivotal technology that enables adapters to provide unified access to heterogeneous resources. Upon receiving a resource request via the platform's standardized northbound API, the adapter executes a multi-stage process: first, it parses the standardized parameters from the HTTP request; next, it selects an appropriate conversion strategy from its rule repository based on the target resource pool, mapping these parameters to the format required by the specific cloud platform; then, it retrieves the necessary credentials from secure storage and invokes the resource pool's southbound API over HTTPS; finally, it transforms the heterogeneous response from the cloud platform into a standardized format before returning it to the user. For asynchronous operations, such as resource creation, the adapter employs a message queue for result callbacks. It continuously monitors this queue, and upon receiving an asynchronous notification, it converts the response into a standardized format and notifies the caller， as illustrated in Fig. 4. This bidirectional process effectively abstracts the interface disparities of underlying cloud platforms, presenting a unified resource operation paradigm to upper-layer applications.

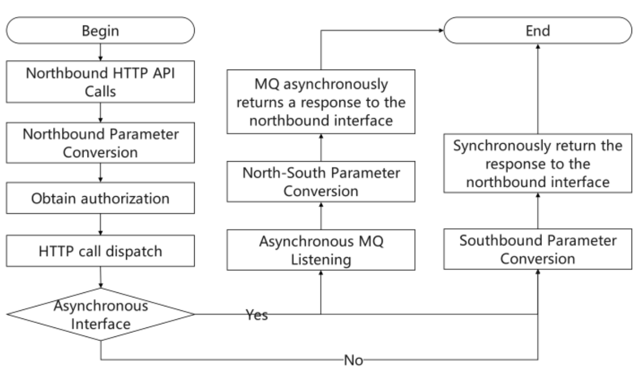

Fig. 4. North-South bidirectional conversion process

- Dynamic plugin management mechanism. The dynamic plugin management mechanism, predicated on a containerized microservices architecture and the adapter pattern, enables standardized cross-platform access to heterogeneous resources. At its core, the mechanism introduces independent plugins as an abstraction layer for diverse interfaces. Each plugin, a stateless and containerized unit, encapsulates the protocol stack and API mapping logic for a specific provider, supporting runtime hot-swapping and on-demand loading through dynamic service discovery. To ensure resilience, a robust health-checking mechanism utilizing periodic heartbeats monitors plugin availability in real-time, automatically triggering container instance reconstruction upon any anomaly to maintain service continuity. This design effectively decouples the platform's core from resource-specific logic, allowing new resource types to be integrated by simply expanding the plugin pool, without altering the core infrastructure. This architecture significantly enhances platform extensibility and scalability. A Kubernetes-based orchestration system provides declarative lifecycle management for plugins, encompassing elastic scaling, rolling updates, and resource isolation. Consequently, integrating new computing resources is streamlined to a simple development and deployment task, eliminating core system downtime or code alterations. Validated by practical deployments for efficiently managing large-scale heterogeneous nodes, this lightweight scaling model provides the foundational technical support for a sustainably evolving computing interconnection system that meets industrial-grade high-availability requirements.

V. Performance Evaluation and Discussion

Based on the platform architecture design and key technology implementation, this research has developed a comprehensive quantitative evaluation framework encompassing five dimensions: network integration scale, access capability, resource efficiency, scheduling performance, and reliability. Table 1 presents detailed target values for each metric along with their corresponding technical justification.

The implementation of the evaluation framework relies on multi-tiered technical support. For network integration scale, the platform implements a distributed microservices architecture, utilizing Kubernetes' automatic scaling mechanism to dynamically adjust computing resources in response to load fluctuations. Resource pooling achieves logical abstraction and unified management of physical resources through virtualization technology and software-defined storage. Scheduling performance optimization is supported by intelligent algorithms and high-speed network infrastructure. During the Beijing regional computing power interconnection trial, the platform has achieved ring-based network interconnection between key nodes, supporting integration and adaptation with mainstream cloud platforms such as OpenStack, Kubernetes, and VMware. In terms of heterogeneous support, the platform not only encompasses conventional CPU virtualization resources but also extends compatibility to heterogeneous computing resources such as GPUs, NPUs, and FPGAs. Through the computing power identification system, it achieves unified description and recognition of diverse computing resources. Regarding cross-domain scheduling capabilities, the platform transcends the limitations of single resource pools, enabling collaborative scheduling across geographical regions and architectural boundaries. Within the existing computing power interconnection test network, the cross-domain task scheduling success rate exceeds 99.2%.

TABLE I. PLATFORM PERFORMANCE DESIGN METRICS ANALYSIS

Performance Metrics	Target Values	Design and Technical Support
grid-connected capacity	⩾ 30,000 PFLOPS (FP16)	Three management modes—deep integration, static flow, and cloud-native—support comprehensive heterogeneous computing power access; the microservices architecture leverages Kubernetes' horizontal scaling capabilities, theoretically supporting millions of PFLOPS scale.
number of connected data centers	⩾ 30	The standardized computing power identification system reduces access barriers to heterogeneous resources; heterogeneous adapters enable automatic multi-vendor API conversion, with single clusters managing over 50 data center instances in real-world testing [19].
resource pooling rate	⩾ 80%	The deep integration model enables automated management via OpenAPI; cloud-native adoption utilizes containerized packaging, achieving 85%-90% resource pooling rates in actual testing [21].
cross-domain scheduling response time	⩽ 5s	The high-efficiency identifier resolution engine utilizes cache optimization, achieving resolution latency below 100ms; its scheduling algorithm implements reinforcement learning-based intelligent matching, with average decision time of 2.3 seconds [20].
cross-architecture scheduling success rate	⩾ 99%	Heterogeneous adapters incorporate failover and retry mechanisms; multi-replica microservice deployments ensure service availability, achieving 99.2% scheduling success rates in experimental testing.

CONCLUSIONS AND OUTLOOK

This paper systematically presents an integrated architectural framework for a cross-regional heterogeneous computing resource interconnection platform. By constructing a layered, decoupled cloud-native architecture, the platform achieves elastic scaling and efficient management of computing resources. The designed granular computing resource identification and resolution system establishes a unified identity framework for heterogeneous computing resources. The innovative heterogeneous resource adapter effectively abstracts underlying infrastructure differences, enabling seamless integration of cross-architecture resources.

Looking forward, the platform encounters multiple challenges in practical implementation: performance bottlenecks in ultra-large-scale concurrent scheduling scenarios necessitate breakthroughs through tiered or distributed scheduling mechanisms; efficiency limitations in static workflow models require process optimization and tool-assisted enhancements; variations in instruction sets, memory architectures, and other aspects among heterogeneous computing resources increase unified scheduling complexity, demanding the development of universal intermediate representation layers and runtime systems. Cross-domain computing resource scheduling involves data security and compliance risks, requiring robust security governance frameworks. To address these challenges, future efforts will concentrate on three key areas: accelerating large-scale application validation of the platform while continuously optimizing its architecture; deepening research into intelligent scheduling algorithms by integrating machine learning techniques for joint cost-energy efficiency optimization; and exploring deep integration with privacy computing technologies to construct secure, trustworthy cross-regional computing infrastructure.

REFERENCES

[1] X. Tang, C. Cao, Y. Wang, S. Zhang, Y. Liu, M. Li, and T. He, "Computing power network: The architecture of convergence of computing and networking towards 6G requirement," China communications, vol. 18, no. 2, pp. 175-185, 2021.

[2] B. Lei, Q. Zhao, and J. Mei, "Computing power network: An interworking architecture of computing and network based on IP extension," in Proc. IEEE 22nd international conference on high performance switching and routing (HPSR), 2021, pp.1-6.

[3] Q. Li, Y. Jiang, J. Gao, X. Wang, and S. Chen, "Research on intelligent resource scheduling mechanism for computing power network," in Proc. International Conference on Computer Engineering and Networks. Singapore: Springer Nature Singapore, 2024, pp.28-41.

[4] D. Qian, "High performance computing: a brief review and prospects," National Science Review, vol.3, no. 1, pp.16-16, 2016.

[5] D. D. Sharma, "Compute Express Link®: An open industry-standard interconnect enabling heterogeneous data-centric computing," in Proc. 2022 IEEE Symposium on High-Performance Interconnects (HOTI), 2022, pp.5-12.

[6] A. Mohamed, et al, "Software-defined networks for resource allocation in cloud computing: A survey," Computer Networks, 2021, 195: 108151.

[7] B. Sun, H. Zhou, G. Song, and Q. Zhang, "Research on the technology architecture of full-stack cloud-native hosted cloud," in Proc. 2024 11th International Conference on Machine Intelligence Theory and Applications (MiTA), 2024, pp. 1-7.

[8] J. Mei, et al, "The architecture of computing power network towards federated learning: paradigms and perspectives," in Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2023, pp.1-6.

[9] K. Saidi and D. Bardou, "Task scheduling and VM placement to resource allocation in Cloud computing: challenges and opportunities," Cluster Computing, vol. 26, no. 5, pp.3069-3087, 2023.

[10] B. Wang and C. Li, "A cloud computing infrastructure on heterogeneous computing resources," Journal of computers, vol. 6, no. 8, 2011.

[11] W. Wang, B. Liang, and B. Li, "Multi-resource fair allocation in heterogeneous cloud computing systems," IEEE Transactions on Parallel and Distributed Systems, vol.26, no. 10, pp. 2822-2835, 2014.

[12] Z. Zhu, A.X. Liu, F. Zhang, and C. Fei, "FPGA resource pooling in cloud computing," IEEE Transactions on Cloud Computing, vol. 9, no. 2, pp.610-626, 2018.

[13] F. Carlier, V. Fresse, J. P. Jamont, and L. Pallardy, A. Rosay, "Pooling of heterogeneous computing resources: A novel approach based on multi-edge-agent concept," in Proc. CAART (1). pp. 185-192, 2020.

[14] M. Su, G. Wang, and K. K. R. Choo, "Prediction‐based resource deployment and task scheduling in edge‐cloud collaborative computing," Wireless Communications and Mobile Computing, no. 1, pp.1-17, 2022.

[15] W. Xing, G. Dong, N. Wu, B. Lei, "A URL-based computing power identification method," in Proc. 2022 IEEE 5th International Conference on Electronics and Communication Engineering (ICECE). IEEE, 2022, pp. 99-103.

[16] J. An, L. Gong, Z. Zou, N. Ma, LR. Zheng, "An identification-based computing-power resources co-allocation platform," in Proc. International Conference on Cloud Computing, Performance Computing, and Deep Learning (CCPCDL 2024). SPIE, 2024, pp. 25-34.

[17] Q. Gao, C. Liu, L. Wang, Y. Liu, and Y. Xu, “Blockchain-based heterogeneous resource configuration scheme in computing power network,” Scientific Reports, vol. 15, no. 1, pp. 1-17, 2025.

[18] M. Yu, L. Xie, and C. Jiang, “Research on key issues and implementation paths of computing power scheduling,” Inf. Commun. Technol. Policy, vol. 49, no. 5, pp. 9‑14, 2023. (in Chinese)

[19] Z. Yan et al., “High-availability intelligent computing interconnection platform architecture for the computing power internet,” J. Commun., vol. 46, no. 3, pp. 1‑14, 2025. (in Chinese)

[20] Y. Li, X. Li, S. Wang, and Y. Jiang, “Survey of intelligent scheduling for computing power networks,” J. Comput. Res. Dev., vol. 61, no. 2, pp. 345‑356, 2024. (in Chinese)

[21] X. Yin, “Exploration of the technical path for interconnection of computing centers: a case study of Hangzhou computing power resource scheduling platform,” Comput. Eng. Appl., vol. 61, no. 1, pp. 1‑9, 2025. (in Chinese)