Using the Internet of Everything for High-Performance Computing

Zihan Zhang[1], Juan Chen[1](✉), Sitan Liu[1], and John Impagliazzo[2]

[1] College of Computer Science and Technology, National University of Defense Technology,
Changsha 410073, Hunan, China
{zihanzhang,juanchen,liusitan20316}@nudt.edu.cn
[2] Hofstra University, Hempstead, NY, USA
john.impagliazzo@hofstra.edu

Abstract. In today's world, the demand for big data and computing power has made High-Performance Computing (HPC) popular among various fields. Supercomputing as one of the representatives of HPC applications, the combination with the Internet of Everything (IoE) in its usage cycle cannot be ignored. For HPC, the Internet is the fundamental prerequisite for interconnecting multiple computers. The IoE enables more devices to be connected and extends how HPCs can be used. The Internet of Everything will affect the applications between HPC and various fields more deeply by connecting people, data, and machines. This paper provides an overview of the history of the Internet of Everything and HPC. It also shows the interaction issues between humans and HPC for the Internet of Everything. The Internet of Everything is widely used and has been shown to impact the HPC field profoundly. There is great potential for developing the Internet of Everything and HPC in the coming years and combining the two to drive multi-domain development.

Keywords: Internet of Everything · HPC · supercomputer

1 Introduction

The emergence of the Internet has brought people closer to each other, and the emergence of the Internet of Things (IoT) has made it possible to connect all things together and realize the interconnection of all things. The Internet of Everything has penetrated every aspect of people's lives. The Internet of Everything is a vast network that combines people, processes, data, and things, interacting and exchanging real-time data between components, an extension and expansion of the Internet of Things [1]. In terms of supercomputing, people (users and managers), data, and supercomputing form a system that can interact in two ways. Raw data is more valuable and influential due to the interconnection of everything. Supercomputers with many computer nodes cannot be more effective if they cannot communicate with each other, between computer nodes, between supercomputers and supercomputers, or between supercomputers and users, and only perform stand-alone operations.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2024
Published by Springer Nature Switzerland AG 2024. All Rights Reserved
T. Pereira et al. (Eds.): IOECON 2023, LNICST 551, pp. 3–10, 2024.
https://doi.org/10.1007/978-3-031-51572-9_1

From accessing only sporadic devices to the Internet of everything, the intelligent application of network technology brings great convenience. HPC can serve as the intersection of data collection and utilization and provide more intelligent services for human beings according to the needs of users [2].

This paper provides a brief overview of the inextricable relationship between the Internet of Everything and HPC. The article introduces the background of the development of the Internet of Everything and HPC, outlines the impact of the Internet of Everything on HPC, and lists the relationship between existing supercomputing and the Internet of Everything. It also outlines the importance of the Internet of Everything in HPC and an outlook on future development. Finally, the article discusses the combination of IoT and HPC to drive growth in multiple fields.

2 Background of IoE

The foundation of the Internet of Everything is the Internet. The Internet originated in the 1960s when the U.S. Department of Defense commissioned the development of ARPANET to research the Internet [3]. The earliest network formed was the ARPANET, with only four nodes, based on a report by Larry Roberts. As the number of computing nodes increased, the NCP protocol initially used needed to be revised to support the accurate location of information between different kinds of computers. The TCP/IP protocol proposed by Wynton Cerf and Robert Kahn solved this problem and significantly advanced the development of the Internet. In the 1980s, to share computer and network resources, the National Science Foundation established the NSFNET wide area network to make the Internet available to the entire community. Academic groups, corporate R&D groups, and individuals could share the computing power of NSF's giant computers and communicate with each other.

The Internet of Things (IoT) [4] is the third wave of information industry development after the development of the Internet. IoT is a more extensive, complex network formed by connecting Internet-connected devices that can sense and share data and securely interact with each other based on Internet protocols. IoT extends the user side to communicate at the item level. The world-famous "Troy Coffee Pot" is a portable camera installed next to the coffee pot to transmit coffee status information to a personal computer so that people can check it at any time [5].

The Internet of Everything is seen as the next phase of the Internet of Things and will be the future of society. The Internet of Everything (IoE) combines people, processes, data, and things that make network connections more relevant and valuable [1]. The Internet of Everything is driving the market for new applications in the communications industry, profoundly affecting how people travel, work, educate, and more.

3 Background of HPC

High-performance computing (HPC) [6] is widely used to solve performance-intensive and complex problems in defense, science, and finance by aggregating computing power using multiple computers and storage devices, with the advantage of high speed and low cost. HPC can run locally, in the cloud, or a hybrid mode [7].

Table 1. The Development of HPC

Machine	Year	Phase
Cray-1 (LANL)	1976	Vector
Cray X-MP (Digital Productions)	1982	Vector
Cray Y-MP (NASA Ames)	1988	Vector
Thinking Machines CM-5 (LLNL)	1993	MPP/SMP
Hitachi SR2201 & CP-PACS (Tokyo and Tsukuba)	1996	MPP/SMP
Intel ASCI Red (SNL-1 teraflop)	1997	MPP/SMP
IBM ASCI White (LLNL)	2000	MPP/SMP
NEC Earth Simulator (JAMSTEC)	2002	Clusters
IBM Blue Gene/L (LLNL)	2004	Clusters
IBM Roadrunner (LANL-1 pataflop)	2008	Clusters
Cray Titan (ORNL)	2012	Clusters
NUDT Tianhe-2A (NCSS Guangzhou)	2013	Clusters
Sunway TaihuLight (NCSS Wuxi)	2016	Clusters
IBM Summit HPE (ORNL)	2018	Clusters
HPE Acquires Cray	2019	Clusters
Fujitsu Fugaku (Japan RIKEN)	2020	Clusters

The development of HPC can be divided into four phases: vector machines, SMP, MPP, and clusters [8]. The first generation of high-performance computers was the vector computer named the Cray machine which appeared in the 1970s [9]. Vector machines add vector flow components to computers to improve computer performance, such as CDC series and CRAY series. Due to the high cost of vector machines, SMP (Symmetric Multiprocessor) was born, an architecture where a relatively small number of processors are installed inside the computer. All processors share memory and data bus, so SMP is high-performance, cost-effective, and compatible with standard hardware and software devices of the time. The MPP (massively parallel processing) [10] architecture was proposed to overcome the poor scalability of SMP. MPP consists of multiple independent microprocessors and uses a dedicated high-speed Internet to collaborate on tasks through software. The fourth stage of HPC development is clustering. Cluster system integrates various IT technologies in one, and at the same time, has low cost and high scalability, which is now the mainstream of HPC development. There are many areas where cluster architecture is applied, such as cloud computing and supercomputing [6]. Representative computers of different stages are listed in Table 1.

Unlike standard computing systems that primarily use serial computing to solve problems, HPC performs massively parallel computing and has high-performance computing components that guarantee the computing power of the cluster.

Due to the increasing demand for processing large amounts of data and the decreasing cost of usage, HPC technology has gained wide popularity in various fields. Generally, users can enjoy HPC services using locally deployed infrastructure or cloud resources. In addition, we can use HPC for weather forecasting, climate modeling, energy research, and intelligence processing, involving the processing of large amounts of data and millions of changes in the associated data points [11].

4 Impact of the IoE on HPC

In the late 1980s, the emergence of high-performance computing pushed supercomputing from a small number of professionals to a broader range of users. At the same time, the pursuit of metrics changed from a requirement for supercomputing arithmetic power alone to a comprehensive plan of high performance.

Jupyter [12] serves as an entry point for users to use HPC. Jupiter is a rich user interface that facilitates user interaction with supercomputers. It can be the entry point for users on HPC. Jupyter serves as an interface to the HPC hub and can be used as a platform for innovation. Users can systematically extend access to more shared nodes on Cori. A modular extension framework for JupyterLub that addresses some of the challenges users face on their laptops. For example, some new integrations are generated: file system navigation system, batch queues, and shared template notebooks.

Jupyter implements a new model of interactive supercomputing integration that allows code, analysis, and data to be brought together under a single visual interface with seamless access to powerful hardware resources. Jupiter developers work at the intersection of project and HPC centers and have built networks of contacts to collaborate and learn from each other.

5 Impact of the IoE on Supercomputers

A supercomputer is a high-performance computing cluster in which a dedicated or high-speed IB network connects numerous computers. Supercomputers include state-of-the-art hardware and software systems, testing tools, and algorithms for solving complex computations. Supercomputers seek high computational speed, and their benchmarks are focused on testing floating-point speed.

Typically, IoE is understood as the bridge that connects spaces such as people's homes with commercial and mobile environments (see Fig. 1) [1]. During the study of supercomputing, as shown in Fig. 2, the Internet of Everything connects users, data, and things into a whole, which includes person-to-person (P2P), machine-to-machine (M2M), and person-to-machine (P2M) systems. The person-to-person (P2P) system includes four interactions: administrator-to-administrator (A2A), administrator-to-user (A2U), user-to-administrator (U2A), and user-to-user (U2U). The machine-to-machine (M2M) system contains two parts: different supercomputing centers cooperating and the HPC interconnection network collecting data from each computer node. The human-to-machine (P2M) system results from the different requirements of experts in different application areas, and the supercomputing centers provide powerful computing power to domain-oriented experts. While advancing science, it also drives the development of more powerful supercomputers.

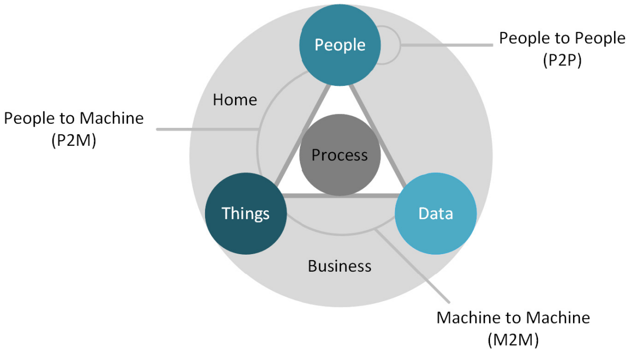


Fig. 1. Internet of Everything (See Fig. 1 in [1])

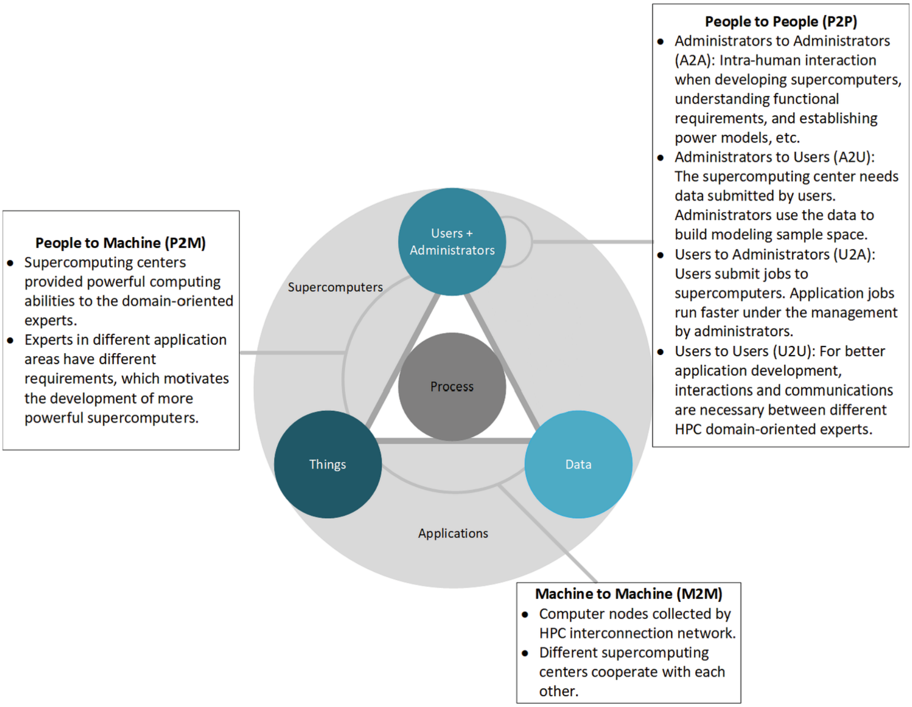
6 The Future of the IoE and HPC

The development of HPC is at a crossroads [6]. Semiconductor technology limitations and market forces are affecting the growth of HPC innovation. Denard scaling and the end of Moore's Law have continued to drive up the cost of manufacturing HPC systems. Cloud services companies should invest more deeply in the HPC market, and HPC with a leading edge is rarely seen as a business opportunity. Cutting-edge HPC systems need to rethink and re-build hardware and software configurations, revolutionize historical approaches, and face new challenges in the market [13].

The rapid development of the Internet of Everything puts new demands on HPC, and the centralized computing and processing model can no longer meet the needs of multiple business models. A new technology has emerged - the small distributed edge-based HPC, which provides end users with cloud computing resources and storage capabilities [14]. The dramatic growth in connectivity and increased data traffic of IoT devices has increased the need for edge-enabled HPC with low latency. Compared to traditional HPC, edge-enabled HPC can reduce processing time, provide data processing and management services, and take up much less space.

Edge-enabled HPC and emerging IT technologies are combined and developed, and new business forms are born and come into people's lives. For example, Telematics, smart classrooms, and AR/VR technology [15–19]. The future HPC industry will see bipolar development: edge-based HPC that conducts latency-sensitive services at the edge and traditional HPC systems that require centralized storage, deep processing, and analysis of large amounts of data. Edge-enabled HPC is also the first entry point for a wide range of data sources, forming a vast network. In addition, HPC will also converge with big data to analyze big data and run simulations and other HPC loads through the same large-scale computer clusters, more powerfully advancing groundbreaking research and innovation.

Cloud computing, edge computing, and end devices can be used as an infrastructure for IoT to bring computing closer to the end devices. The rapid growth in the computational demand for HPC systems has led to edge-level and cloud-level tiers being considered together for energy optimization and scheduling. Li et al. [20] comprehensively quantify the carbon footprint of HPC systems, considering both the hardware



People to People (P2P)
- Administrators to Administrators (A2A): Intra-human interaction when developing supercomputers, understanding functional requirements, and establishing power models, etc.
- Administrators to Users (A2U): The supercomputing center needs data submitted by users. Administrators use the data to build modeling sample space.
- Users to Administrators (U2A): Users submit jobs to supercomputers. Application jobs run faster under the management by administrators.
- Users to Users (U2U): For better application development, interactions and communications are necessary between different HPC domain-oriented experts.

People to Machine (P2M)
- Supercomputing centers provided powerful computing abilities to the domain-oriented experts.
- Experts in different application areas have different requirements, which motivates the development of more powerful supercomputers.

Machine to Machine (M2M)
- Computer nodes collected by HPC interconnection network.
- Different supercomputing centers cooperate with each other.

Fig. 2. P2P, P2M, and M2M in the Supercomputer Power Modeling and Management

production and system operation phases. Lin et al. [21] propose a multi-layer federation learning protocol called HybridFL that implements different aggregation strategies at the edge and cloud levels. Experiments demonstrate that HybridFL significantly speeds up the FL process and reduces the system energy consumption by reducing the length of short federation rounds and accelerating the convergence of the global model. Hou et al. [22] designed a hardware/software co-design solution AMG for multimodal artificial intelligence systems. AMG features a novel decoupled modal sensor structure and supports a new intelligent power management strategy that reduces the energy consumption of the modal sensors while ensuring system accuracy.

Artificial Intelligence and Edge Computing combine to solve the problems posed by the limits of Moore's Law. NVIDIA has created a performance-oriented Streaming Reactive Framework (SRF). The framework standardizes HPC streaming data pipelines to build modular and reusable pipelines for sensor data [23]. APS is a machine that generates photon beams to study materials, physics, and biological structures. Most of the inversion process can be avoided by using artificial intelligence when generating material images with nanoscale resolution via photonography, and deploying PtychoNN models trained in data centers on edge devices can increase image processing speed by more than 300 times.

7 Discussion

The combination of the Internet of Things and high-performance computing drives growth in multiple areas.

Within the traditional field of HPC, applications focus on more than just areas dominated by the solution of partial differential equations. Still, they are also expanding into new areas of development in conjunction with new technologies. HPC machines are not "general purpose" platforms, which poses several challenges for various applications in an HPC environment. Weßner et al. [24] have developed a framework, Geneva, for the parallel optimization of large-scale problems with a highly nonlinear quality surface. They created a new network component, MPI Consumer, that conforms to the standard paradigm of HPC programs. The powerful computing power of HPC contributes significantly to the simulation of physical systems. Galen et al. [25] provides a detailed quantitative analysis of representative simulations. He argues that to maximize the value of HPC computing power, emphasis should be placed on the deep synergistic development of hardware and software.

Based on edge computing, more and more devices have become mobile or non-mobile computing platforms. Along with the development of machine learning, sensors, and other technologies, IoT devices are equipped with intelligent and interconnected computing systems while maintaining their original functions. IoT devices have three main functions: sensing, communication, and computing. Telematics is to use the vehicle as an independent computing platform, sensing the environment through devices such as cameras and sensors, performing data analysis as well as computation of in-vehicle applications within the forum, and accomplishing the communication between vehicle and vehicle, and vehicle and cloud, which solves the problem of the transmission delay of the data in the large-scale Internet. Zhang et al. [26] proposed to build an open in-vehicle data analytics platform for automatic driving OpenVDAP, with an operating system that realizes the computation, communication, and security and privacy protection of the vehicle and provides optimal strategies based on real-time information of vehicle dynamic detection. Smart home pays more attention to the ubiquitous sensing ability of edge nodes, studies the fusion design of wireless sensing and wireless communication, and combines machine learning to build personalized models to ensure the privacy and security of user data. Huang et al. [27] explored the potential of smart homes using non-contact respiratory monitoring, pervasive sensing of devices through sound and electromagnetic signals, co-design of sensing and communication signals, and finally optimization of computation and communication overhead.

References

1. Kiesler, N., Impagliazzo, J.: Perspectives on the internet of everything. In: IoECon2022 (2022)
2. Higginbotham, S.: Network included - [Internet of Everything]. IEEE Spectr. **57**(11), 22–23 (2020). https://doi.org/10.1109/MSPEC.2020.9262153
3. Leiner, B., et al.: A brief history of the internet. Comput. Commun. Rev. **39**, 22–31 (2009). https://doi.org/10.1145/1629607.1629613
4. Khvoynitskaya, S.: The IoT history and future. Online (2019). https://www.itransition.com/blog/iot-history

5. https://bbs.huaweicloud.com/blogs/308804
6. Reed, D., Gannon, D., Dongarra, J.: HPC forecast: cloudy and uncertain. CACM **66**(2), 82–90 (2023)
7. Chen*, J., et al.: More bang for your buck: boosting performance with capped power consumption. Tsinghua Sci. Technol. **26**(3), 370–383 (2021). ISSN 1007-021. https://doi.org/10.26599/TST.2020.9010012
8. Wang, R., et al.: Brief introduction of TianHe exascale prototype system. Tsinghua Sci. Technol. **26**(3), 361–369 (2021). ISSN 1007–0214. https://doi.org/10.26599/TST.2020.9010009
9. https://www.cnblogs.com/laxcus/p/14666901.html
10. https://blog.csdn.net/ieee2007/article/details/5974472
11. Chen, J., Ghafoor, S., Impagliazzo, J.: Producing competent HPC graduates. Commun. ACM **65**(12), 56–65 (2022). https://doi.org/10.1145/3538878
12. Thomas, R., Cholla, S.: Interactive supercomputing with jupyter. Comput. Sci. Eng. **23**(2), 93–98 (2021)
13. Dong, Y., Chen*, J., Tang, Y., Wu, J., Wang, H., Zhou, E.: Lazy scheduling based disk energy optimization method. Tsinghua Sci. Technol. **25**(2), 203–216 (2020). https://doi.org/10.26599/TST.2018.9010140
14. Kong, X., Wu, Y., Wang, H., Xia, F.: Edge computing for internet of everything: a survey. IEEE Internet Things J. **9**(23), 23472–23485 (2022). https://doi.org/10.1109/JIOT.2022.3200431
15. Badia, R.M., Foster, I., Milojicic, D.: Future of HPC. IEEE Internet Comput. **27**(1), 5–6 (2023). https://doi.org/10.1109/MIC.2022.3228323
16. Bauer, M., Lee, W., Papadakis, M., Zalewski, M., Garland, M.: Supercomputing in python with legate. Comput. Sci. Eng. **23**(4), 73–79 (2021)
17. Kline, C.: Supercomputers on the internet: a case study. ACM SIGCOMM Comput. Commun. Rev. **17**(5), 27–33 (1987). https://doi.org/10.1145/55483.55487
18. Sato, M., Kodama, Y., Tsuji, M., Odajima, T.: Co-design and system for the supercomputer “Fugaku.” IEEE Micro **42**(2), 26–34 (2021). https://doi.org/10.1109/MM.2021.3136882
19. Biryal’tsev, E.V., Galimov, M.R., Elizarov, A.M.: Workflow-based internet platform for mass supercomputing. Lobachevskii J. Math. **39**, 647–654 (2018). https://doi.org/10.1134/S1995080218050050
20. Li, B., Samsi, S., Gadepally, V., Tiwari, D.: Sustainable HPC: modeling, characterization, and implications of carbon footprint in modern HPC systems (2023). arXiv:2306.13177
21. Wu, W., He*, L., Lin, W., Mao, R.: Accelerating federated learning over reliability-agnostic clients in mobile edge computing systems. IEEE Trans. Parallel Distrib. Syst. **32**(7), 1539–1551 (2021)
22. Hou, X., et al.: Architecting efficient multi-modal AIoT systems. In: Proceedings the 50th International Symposium on Computer Architecture (ISCA) (2023)
23. https://developer.nvidia.com/zh-cn/blog/facing-the-edge-data-challenge-with-hpc-ai/
24. Weßner, J., Berlich, R., Schwarz, K., et al.: Parametric optimization on HPC clusters with Geneva. Comput. Softw. Big Sci. **7**, 4 (2023). https://doi.org/10.1007/s41781-023-00098-6
25. Shipman, G.M., et al.: The future of HPC in nuclear security. IEEE Internet Comput. **27**(1), 16–23 (2023)
26. Zhang, Q., et al.: Openvdap: an open vehicular data analytics platform for Cavs. In: 2017 IEEE 38th International Conference on IEEE Distributed Computing Systems (ICDCS) (2018)
27. Huang, Q., Li, Z., Xie, W., et al.: Edge computing in smart homes. Comput. Res. Dev. (9) (2020). https://doi.org/10.7544/issn1000-1239.2020.20200253