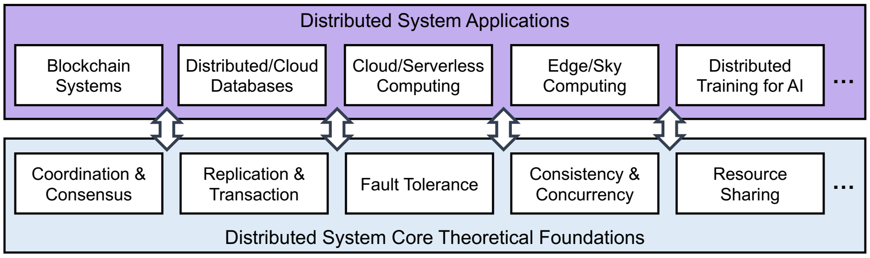
Gengrui Zhang
https://gengruizhang.github.io

Research Statement

My research is at the core of distributed systems. Recent years have seen significant technological advancements in distributed systems, with various applications ranging from localized to global-scale systems involving millions of nodes, such as serverless computing, cloud databases, and blockchains. As the demand for more scalable, available, and versatile distributed systems continues to grow, new challenges have emerged in terms of efficiency, robustness, and heterogeneity in system design and architecture.

The goal of my research is to create advanced algorithms and architectures that enable the development of high-performance, highly scalable, and highly available distributed systems. During my Ph.D., I have worked to bridge the gap between distributed system applications and their underlying theoretical foundations, utilizing real-world scenarios to inspire the development of new algorithms and system architectures.

I have conducted research on the following topics: ① efficient and robust consensus algorithms, ② fault-tolerant protocols for distributed system applications (especially blockchain systems), and ③ fast replication protocols for cloud and distributed databases. My research has been featured in highly regarded conferences focused on distributed systems, such as IEEE ICDCS and ACM Middleware [3, 4]. Additionally, some of my work has led to collaboration with industry partners and patented [2, 5].


Distributed System Applications
Blockchain Systems
Distributed/Cloud Databases
Cloud/Serverless Computing
Edge/Sky Computing
Distributed Training for AI
...
Coordination & Consensus
Replication & Transaction
Fault Tolerance
Consistency & Concurrency
Resource Sharing
...
Distributed System Core Theoretical Foundations

My long-term research goal is to develop efficient, reliable, and robust algorithms and architectures for large-scale distributed systems that provide computation, coordination, replication, and transaction services as a utility. I envision future research progressing along the following lines of exploration, as illustrated in the diagram above: ① developing more secure and efficient consensus services, particularly for blockchain systems; ② enabling ubiquitous coordination among computational resources in cloud, edge, and sky computing; and ③ facilitating large-scale distributed training for AI.

Current Research

CR1: Reputation-based consensus algorithms.

Consensus algorithms conduct state machine replication (SMR) among servers despite failures. However, under Byzantine failures, the traditional specification of SMR describes the process of replication but has no language to describe the correctness of participants, leaving the system vulnerable to repeated attacks that target leader servers.

To address this vulnerability, my research developed reputation-based consensus algorithms, `Prosecutor` [3] and `PrestigeBFT` [8]. `Prosecutor` imposes Proof-of-Work computation on suspected faulty servers during view changes, suppressing Byzantine servers from becoming new leaders. Moreover, `PrestigeBFT` establishes a reputation engine that discredits misbehaving servers and rewards protocol-obedient servers, using worsening and improving reputations, respectively. In addition, `PrestigeBFT` enables active view changes where servers proactively campaign for leadership, resulting in a more efficient view change process compared to state-of-the-art BFT algorithms and blockchain platforms [7].

Impact: My reputation-based consensus algorithms have made a significant impact on distributed computing theory and system architectures. Firstly, they extend traditional state machine replication properties to a reputation state, which opens up new discussions on Byzantine fault tolerance. Secondly, they are the first consensus algorithms that not only achieve high performance but also suppress intentional faults, binding efficiency and robustness together. Specifically, `PrestigeBFT` achieves a 5.4× and 4.2× higher throughput than HotStuff in peak performance under normal operation and sustained Byzantine attacks, respectively.

CR2: Blockchains for vehicle-to-everything (V2X) networks.

The integration of blockchain technology into the automotive industry has emerged as a pressing research challenge, as automobile manufacturers increasingly exert centralized control over vehicle data. This raises significant concerns with respect to the allocation of legal responsibility between vehicle and driver in the event of accidents, thereby necessitating the development of blockchain-based solutions for V2X networks. However, traditional BFT algorithms and permissioned blockchains operate in stable networks characterized by a static set of servers. In contrast, V2X networks are highly dynamic and susceptible to frequent disruptions, as vehicles may enter or exit the network at will. This problem poses a significant technical challenge for developing robust and scalable blockchain solutions for the automotive industry.

My research proposed a novel permissioned blockchain with a new consensus algorithm for V2X networks, namely `V-Guard` [6], which aims to address the issue of intermittently connected vehicles in V2X networks. V-Guard establishes a membership management unit that allows transactions to be ordered and committed under different memberships (sets of vehicles). It achieves consensus seamlessly under changing members (e.g., with joining or leaving vehicles) and produces an immutable ledger recording traceable data transactions with their corresponding membership profiles. This project has become open source on GitHub at https://github.com/vguardbc/vguardbft.

Impact: ① `V-Guard` is the first blockchain architecture that allows consensus to be achieved in a dynamic environment with high performance. ② This project has filed a US patent [5] and is being used by an industry collaborator. ③ `V-Guard`'s general-purpose architecture makes it suitable for applications operating in unstable networks, such as Internet-of-Things, supply chain, and retail applications. We have observed several projects that utilize `V-Guard` to build their applications [1].

CR3: Fast leader election protocols.

Leader election protocols are essential for large-scale systems that have a single cluster leader, such as GFS and HDFS, as they facilitate the election of new leaders through voting-based mechanisms. However, such mechanisms often lead to competition among candidates for leadership when votes are split, resulting in a prolonged and undesirable leader election process.

In my research, I proposed a leader election protocol called `Escape` [4] that addresses the issue of split votes in voting-based mechanisms. Specifically, the protocol examines Raft's leader election mechanism and offers solutions that fundamentally resolve split votes. `Escape` assigns priorities to servers based on their log responsiveness, keeping track of their logs and assigning configurations that favor more up-to-date servers, thereby creating a pool of prioritized candidates. During leader election, `Escape` can terminate the process in a single messaging round without being affected by split votes.

Impact: The `Escape` leader election protocol provides a generalized framework that resolves the split-vote problem in leader election algorithms. This solution can be adopted by other leader-based systems like Zookeeper, Redis, and Azure election protocols to improve their performance and efficiency.

Future Research

FR1: Towards software-defined consistency services for distributed systems.

Traditionally, distributed applications only provide hard-coded consistency services based on pre-defined failure assumptions, limiting their versatility. Future distributed applications need to support consistency services that can be defined at an application level to better address various consistency requirements.

Drawing on my previous research in developing consensus algorithms and architectures under various failure models [3, 4, 8, 6, 9], my future research will explore one-size-fits-all architectures that integrate a collection of consistency services with invariants of linearizable, sequential, causal, and FIFO orderings. By enabling software-defined consistency services, we can revolutionize the design of blockchain applications and distributed databases, allowing consistency services to vary based on the request at hand. My research will focus on the following aspects:

① Coordination as a utility. My research will build up fine-grained consistency service components, including communication, quorum construction, storage, and cryptography, and allow for multiplexing among different consistency services.

② Software-defined consistency. My research will investigate the possibility of defining consistency guarantees differently across applications, according to varying requests. Consistency components will be assembled as the "Legos" of the invariants of the defined consistency model, which will enable the development of software-defined consistency services tailored to a wide range of applications.

FR2: Coordination as a service in cloud, edge, and sky computing.

Computing has become a ubiquitous and essential aspect of modern society, with various computation models such as cloud computing (computation-centric) and edge computing (data-centric) gaining prominence. Cloud computing, in particular, has witnessed a significant shift towards serverless computing, offering a cost-effective pay-as-you-go model and automatic, rapid, and unlimited scaling resources up and down per demand. However, auto-scaling in cloud computing presents challenges that require further research and investigation.

① My future research will endeavor to investigate innovative approaches in serverless computing to devise efficient consensus protocols for horizontally scaling added nodes (containers). This approach is expected to outperform traditional vertical scaling methods, especially in scenarios with diverse workloads, and prove highly beneficial for short-term applications experiencing bursts of requests.

In contrast to the centralized model of cloud computing, edge computing leverages computation resources located near the physical location of the user or data source. However, due to the limited computational capacity of edge devices, coordination between cloud and edge devices becomes necessary for processing large datasets.

② (Edge computing.) My research will investigate coordination algorithms and replication protocols between cloud and edge devices under varying computation topologies, with the goal of achieving efficient and effective data processing and analysis in edge computing systems.

Sky computing is a concept that emerges from the utilization of multicloud in a heterogeneous architecture, where computing and storage services come from different vendors. Its primary objective is to facilitate interoperability among these clouds, allowing for the seamless transfer of data based on user-defined criteria, such as moving data from AWS to Google Cloud.

③ (Sky computing.) My research will aim to develop a secure and efficient peering layer that facilitates the interconnection and interoperability of clouds from different vendors in a multicloud environment. This peering layer will establish agreements between clouds on how to quickly and securely exchange services, enabling seamless data transfer based on user-defined criteria.

FR3: More efficient distributed computing and training systems for machine learning.

Recent years have seen explosive growth in the scale and complexity of machine learning applications. Due to the ever-increasing demand for computational resources, even specialized processors have become insufficient, thus necessitating the distribution of computations. For instance, Google's TPU v3 Pods are capable of connecting up to 1,000 TPUs using a high-speed mesh network. As distributed training is becoming more prevalent, my future research will focus on investigating the following areas:

① Communication-efficient distributed training algorithms: As the scale of machine learning models and the number of machines involved in training grows, communication can become a bottleneck in the training process. My research will focus on developing algorithms that optimize communication patterns to minimize the amount of data that needs to be transmitted while maintaining the quality of the trained model.

② Fault-tolerant distributed training: In distributed systems, failures are inevitable, and machine learning training is no exception. My research will explore fault-tolerant algorithms that can recover from node failures and ensure that the training process continues with minimal disruption.

③ Privacy-preserving distributed training: training data is sensitive in many machine learning applications, and privacy is a concern. My research will investigate methods for conducting distributed training while preserving the privacy of the training data, such as secure multi-party computation and federated learning.

References

[1] **Gengrui Zhang**. Projects using V-Guard, 2023. https://github.com/vguardbc/vguardbft#projects-using-v-guard.

[2] **Gengrui Zhang**, Tongxin Bai, and Chengzhong Xu. A kind of Second-hand Vehicle Transaction method, apparatus and system based on block chain technology, 2017. CN Patent 106897887 A[P].

[3] **Gengrui Zhang** and Hans-Arno Jacobsen. Prosecutor: An Efficient BFT Consensus Algorithm with Behavior-aware Penalization against Byzantine Attacks. In *Proceedings of the 22nd International Middleware Conference*, 2021.

[4] **Gengrui Zhang** and Hans-Arno Jacobsen. ESCAPE to Precaution against Leader Failures. In *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, 2022.

[5] **Gengrui Zhang**, Hans-Arno Jacobsen, and Sheng Sun. Method and System for Creating a Distributed Ledger of Verified Vehicle Transactions, 2022. US Patent (Invention Disclosure ID: 10004394).

[6] **Gengrui Zhang**, Yunhao Mao, Shiquan Zhang, Shashank Motepalli, and Hans-Arno Jacobsen. V-Guard: A Fast, Dynamic, and Versatile Permissioned Blockchain Framework for V2X Networks. In *Under view*, 2022.

[7] **Gengrui Zhang**, Fei Pan, Michael Dang'ana, Yunhao Mao, Shashank Motepalli, Shiquan Zhang, and Hans-Arno Jacobsen. Reaching Consensus in the Byzantine Empire: A Comprehensive Review of BFT Consensus Algorithms. *arXiv preprint arXiv:2204.03181*, 2022.

[8] **Gengrui Zhang**, Fei Pan, Sofia Tijanic, and Hans-Arno Jacobsen. Prestige BFT: Making Decentralization Efficient in Distributed Ledgers using Reputation-based Byzantine Fault-Tolerant Consensus Algorithms. In *Under view*, 2022.

[9] **Gengrui Zhang** and Chengzhong Xu. An Efficient Consensus Protocol for Real-time Permissioned Blockchains under Non-Byzantine Conditions. In *International Conference on Green, Pervasive, and Cloud Computing*, pages 298–311. Springer, 2018.