**Statement** **of** **Purpose**

Simply with some configurations and just one single command, thousands of remote machines distributed all over the world will be running your applications concurrently. Thanks to the distributed cloud infrastructures, all machines can run as if they were in the same rack, enabling us to easily access remote physical resources. I have always been fascinated by the theories underlying these systems, craving to learn more about their seemingly “magical” abilities. As a Computer Science undergraduate passionate about learning and building these systems, I have built standalone distributed systems, worked in a top-level cloud company, and collaborated on rewriting the entire code base of a renowned open-source cloud infrastructure. From these experiences, I realized that I need more advanced knowledge to achieve my future goal: to become an extinguished system engineer focusing on building large-scale cloud infrastructure for the general public. I am therefore applying for the well- famed Carnegie Mellon University’s MSCS program to equip myself with more advanced knowledge and further contribute my system insight to the world.

My interest in large-scale systems was sparked when I took the Computer Network course during my second year at Shanghai Jiao Tong University. As an additional class project, I built a peer-to-peer chunk network system with Distributed Hash Table and profiled the system’s performance. After I observed the dramatically increased efficiency, my curiosity to learn more about machine cooperation was awakened. Therefore, I audited MIT 6.824, which was an advanced graduate course on Distributed Systems. Throughout the course, I read 10+ papers on distributed system theory and real-world system designs. Learning about great theories like *Paxos* and *Chain* *Replication* and brilliant architectural designs like *Spanner*, I got inspired to build a distributed system on my own. To test out my newly acquired knowledge, I developed a fault-tolerant key-value storage that enforced strong consistency under machine failures and supported distributed transactions based on *Raft* and *Percolator*. This project not only helped me improve my programming and logical reasoning skills, but also fueled my passion for large-scale systems even further. However, I quickly realized that I wanted to do more than just build systems for learning purposed - I wanted to apply me knowledge to real-world systems and make a tangible impact.

My journey as a cloud system engineer started when I joined the Operating System Group of Al- ibaba Cloud as an R&D intern. During the internship, I was responsible for optimizing the cloud kernel’s memory footprint for our team’s secure container runtime. One of the most challenging prob- lems I faced was related to Linux Kernel Virtual Machine (KVM). I observed that the guest’s pages were mostly tagged inactive on the host side, forcing us to stop reclaiming guest’s anonymous page to avoid performance degradation brought by kswap. The first challenge was to identify its root cause, which was particularly difficult due to the complexity of the code. After carefully analyzing about thirty thousand lines of Linux Kernel and KVM source code, I discovered that the issue was rooted in the *Extended* *Page* *Table* *(EPT)*. The two-dimensional paging mechanism was causing the host ker- nel to evaluate the guest’s working set incorrectly. Once I had identified the root cause, I designed a kernel daemon, periodically synchronizing the access and dirty bits between the guest *EPT* and the host page table. To reduced the daemon’s overhead, I also proposed an optimization that took the advantage of the access patterns of guest memory. My solution significantly alleviated the problem and reduced the guest’s memory footprint by 12%. This experience was the first time when I realized that cloud systems could be optimized by incorporating knowledge from multiple fields of Computer Science, such as Operating Systems and Virtualization. After I realized the potential for incorporating cross-disciplinary approaches, I became interested in applying them to my future work in the field and exploring these connections further in my future graduate study.

After my previous experiences with cloud systems, I became eager to gain a comprehensive un- derstanding of mature cloud systems for real-world clusters. With this overarching goal in mind, I joined the renowned community of *Kata* *Containers*, a top-level open-source project under the Open- Infra Foundation that worked well with companies like Intel, AMD, and IBM. During my time in the community, I had the opportunity to design and build almost the entire cloud infrastructure. My team and I rebuilt the architecture of*Kata* *Containers* and migrated the project’s entire codebase to the Rust programming language for higher performance and maintainability. As one of the main contributors, I took up the responsibility to not only implement more than ten indispensable core features but also fre- quently assist other community members with their tasks, such as supporting QEMU. After adopting the Rust codebase and the new architecture, which utilized the built-in virtual machine, we eliminated the overhead of an extra process and the Golang Runtime. This significantly reduced the boot time from 700 ms to 340 ms and ensured consistent performance when scaling the system vertically and horizontally. Working on an open-source project with developers from diverse backgrounds and with various skill sets, I learned how to build a large project from scratch and collaborate effectively in diverse and interfocus settings. This became especially apparent to me when interacting with commu- nities like *Kubernetes* and *Containerd*. Throughout my time with the community, I also learned how the interactions between communities can strengthen mutual outcomes and brought knowledge from various fields to me. This encouraged me to seek environments like CMU, which offers a flexible curriculum and bring together talented individuals to collaborate with.

Seeing the impacts that my previous work had on companies’ workflows filled me with a sense of purpose and led to my goal of becoming a system engineer focusing on building cloud infrastruc- tures and related systems. I believe the cloud systems will shape the future diagram of computation and continue to change the way people access remote data and resources. To push the boundaries of cloud systems, I wish to incorporate knowledge from various fields of CS into the cloud systems. Therefore, I want to learn advanced designs of cloud systems that do so already, such as *Kubernetes* which utilizes Distributed Systems, and *Krustlet*, which utilizes WebAssembly. The cloud industry has unlimited possibilities, continuously being invigorated by emerging ideas from different areas of Computer Science. Over the next couple of years, I hope to acquire advanced knowledge from MSCS to help me achieve my career vision.

The MSCS program at Carnegie Mellon University is the perfect place for me to do exactly that - expand my knowledge, thus preparing me to enter and succeed in the cloud industry. The flexibility of the curriculum that MSCS provides is especially appealing to me. With the 96 units of the core courses, I can not only learn how to design and build large-scale systems through courses including *Advanced* *Cloud* *Computing* but also gain valuable interfocus insights needed to optimize complex cloud sys- tems by taking modules including *Intro* *to* *Machine* *Learning*. Working with excellent academics and engineers from CMU will inspire me to explore outside of my comfort zone, develop diverse skill sets, and develop excellent collaboration capabilities. What’s more, the MSCS-AS variant will allow me to formalize my academic coursework through a real-world application during the summer applied study experience and further help me make thoughtful decisions for my professional future.

With a robust quantitative skillset, significant experience, and a strong interest in expanding my knowledge of computer systems, I hold a strong belief that I am an excellent candidate for the MSCS program. I believe that after the journey at Carnegie Mellon University, I will be fully prepared to apply my insights into the cloud industry and finally contribute back to the community at CMU.