The topic of the research is on memory allocation algorithms that addresses demands for new computer systems. The focus will be on innovative memory allocation strategies that enhance performance and efficiency in various computing environments, particularly those involving complex server workloads and large-scale data processing. The following “Reference and current research” section will be used for this research project.These studies highlight the evolution of memory allocation techniques from traditional approaches to more sophisticated, context-aware strategies that leverage machine learning to optimize resource distribution.

Exploring from the base concepts presented by the textbook, this research will dive into specialized areas of memory allocation algorithms, and explore recent applications of said algorithms. It’s likely that more research papers will be added on top, and the final project will be focusing on only a few(3 or 4) in order to limit the scope of the research. Through a comparative analysis of these innovative approaches, the project aims to identify key principles and techniques that can significantly improve memory management practices.

References and current researches:

NextGen-Malloc: Giving Memory Allocator Its Own Room in the House
https://dl.acm.org/doi/pdf/10.1145/3593856.3595911

Learning-based Memory Allocation for C++ Server Workloads
https://dl.acm.org/doi/pdf/10.1145/3373376.3378525

Understanding and Optimizing Persistent Memory Allocation
https://dl.acm.org/doi/pdf/10.1145/3381898.3397212

Releasing Memory with Optimistic Access: A Hybrid Approach to Memory Reclamation and Allocation in Lock-Free Programs
https://dl.acm.org/doi/pdf/10.1145/3558481.3591089

NUMAlloc: A Faster NUMA Memory Allocator
https://dl.acm.org/doi/pdf/10.1145/3591195.3595276

Efficient Memory Management for Large Language Model Serving with PagedAttention
https://dl.acm.org/doi/pdf/10.1145/3600006.3613165