

**CS E6998-10: Advanced Topics in Network Storage Systems**  
**Assignment #1**

Due date: Tuesday 2/24

1. **Multi-level caching in distributed file systems.** One of the key design principles in the Andrew File System (AFS) is the use of client caching to reduce server load and improve scalability. Previous work has explored the potential gains of employing *multiple* levels of caching.
  - a. Read and review the paper "*Multi-level Caching in Distributed File Systems*", USENIX Winter 1992, by Honeyman and Muntz. In reviewing this paper, imagine that you are a conference program committee member and focusing on your critique of the research methodology, recommend whether to accept or reject the paper and why.
  
2. **Server caching.** In most distributed file systems, file caching is also performed in the server's in-memory buffer cache.
  - a. What are the potential benefits of such caching, if any?
  - b. Given that the client cache is managed by an LRU replacement policy, what is a good choice for the server's cache replacement policy and why?
  
3. **Content-based addressing.** Some storage systems use the hash of a data block's contents as the block's "address" within the storage device. Two widely used, cryptographically-strong hash functions are MD5 and SHA-1.
  - a. Implement or use existing implementations of the SHA-1 algorithm and characterize its performance as a function of the data block size. In your experiments use blocks with randomly generated contents.
  - b. How does the cost of computing the SHA-1 hash of a data block compare to other operations, such as transmitting the block over a slow network link? Back-of-the-envelope calculations will do here.