Systems, Networking, and Cybersecurity Ph.D. Qualifier Exam

Spring 2024

The following questions relate to the papers in the reading list on the Spring 2024 qualifier webpage (https://people.cs.vt.edu/thanghoang/qual24/). For full citations, please see that reading list. Before starting, read and understand the following guide provided by Virginia Tech: Avoiding Plagiarism: A Guide For Graduate Students at Virginia Tech (https://graduateschool.vt.edu/content/dam/graduateschool_vt_edu/graduate-honor- system/avoiding-plagiarismshort-guide.pdf). In your answers, you must avoid unattributed direct quotations and paraphrases and use proper documentation of all sources you use. This requires that you include a bibliography in your response. Failure to follow these guidelines Represents a violation of Virginia Tech's Honor Code and will result in a score of 0.

1. The following questions relate to the paper "Dynamo: Amazon's Highly Available Key-value Store"
   - ACID is the acronym for the set of properties that must be satisfied to guarantee that a database transaction is processed reliably. Discuss what each of the alphabets of the acronym mean and their implications for transaction processing.
   - Dynamo is a database system that sacrifices a property (of ACID) to enhance another property. Discuss which is sacrificed and which is enhanced and why this is possible with Dynamo.
   - In your own words, discuss Dynamo's partitioning and replication algorithms.
   - Describe the Merkle tree and how it is used to handle failures.

2. The following questions relate to the paper "The Design and Implementation of a Log-Structured File System"
   - Describe the motivation behind the design of LFS from the perspective of 1) a cache and 2) the workload. (LFS: Log-structured File System)
   - Based on your understanding of LFS, what is the key limitation in deploying LFS in a real-world setting? Explain the reasoning behind your answer.
   - Explain why cleaning is necessary in LFS. Explain how the Greedy and Cost-Benefit segment selection policies work, and present scenarios in which each of the policies will work best.
   - While LFS is described in the context of HDDs, today, SSDs are prevalent. SSDs are different from HDDs in that access performance to segments (as described in this paper), irrespective of location, is uniform. Propose an optimization or optimizations to the paper's LFS design such that one may exploit this uniform performance characteristic to improve LFS performance. Justify your proposal.

3. The following questions relate to the paper “PinK: High-speed In-storage Key-value Store with Bounded Tails”

   - Describe the workings of an LSM-Tree.
   - PinK is a KV-SSD that is based on an LSM-Tree. Describe the key observations that allow the LSM-Tree to be used as the main data structure within an SSD, which, unlike a server, has only limited resources.
   - Describe how tail performance may be affected in a typical LSM-Tree. Describe how this is overcome in PinK.
   - Performance evaluation is an important part of a systems paper. Point out any limitations that you see in the performance evaluations section of this paper and describe how you would have improved this section. Justify your description.

4. The following questions relate to the paper “Exokernel: An Operating System Architecture for Application-Level Resource Management”
   - Q1: How does Exokernel balance the need for code communication flexibility with the demand for reliability and security in operating system design?
   - Q2: How does Exokernel handle address translation and memory reclamation, including page replacement?
   - Q3: What performance optimizations does Exokernel offer? Please provide an explanation.

5. The following questions relate to the paper “Multiparty Computation from Somewhat Homomorphic Encryption”
   - What is Multiparty Computation and Homomorphic Encryption this paper is referring to? Give an example of the homomorphic encryption protocol used in this paper.
   - Describe how the secure multiplication proposed in this paper works.
   - What is the active adversary this paper is referring to? Describe the technique this paper proposes to verify the integrity of secure computation against an active adversary.

6. The following questions relate to the paper “Virtual Memory, Processes,and Sharing in MULTICS”

   The MULTICS operating system enabled the sharing of procedures and data in segments across multiple processes. Imagine the scenario where a process (named α) executing a procedure P wishes to reference data that is external to the procedure P; the first time the process makes this reference from P, it will be using the symbolic address for the data (in MULTICS notation, the symbolic address is <D> | [x] with the symbolic segment name <D> and the symbolic external address [x]). On the first reference to this symbolic name, the name is “made known” and translated into a generalized address (in the MULTICS notation, the generalized address is of the format d#$\alpha|$x).

- Why start with a symbolic name and then convert it into a generalized address in MULTICS? Why is using a generalized address directly within procedure P not feasible?
- What structures support this address translation in MULTICS, and what are their characteristics?
- What types of addressing mechanisms does MULTICS employ to facilitate these processes?

7. The following questions relate to the paper “Scheduler Activations: Effective Kernel Support for the User-Level Management of Parallelism”
   - What’s the main issue that the paper tackles?
   - What are the drawbacks of kernel threads and user threads respectively?
   - How does scheduler activation overcome that?

8. The following questions relate to the paper “Zerocash: Decentralized Anonymous Payments from Bitcoin”
   - What are the security issues in bitcoin this paper is trying to address?
   - Describe the main technique this paper uses to achieve privacy-preserving and verifiable cryptocurrency.
   - Explain the role of the Merkle tree in this paper.

9. The following questions relate to the paper “ZeroTrace: Oblivious Memory Primitives from Intel SGX”
   - What is the security problem this paper is trying to solve? Describe the vulnerability with concrete examples.
   - Describe the main cryptographic building blocks this paper uses to enable oblivious memory.
   - What is the role of the CMOV instruction used in this paper? Give two examples (one with CMOV and another without CMOV) to demonstrate the security advantages of the CMOV instruction.