1-1-2023

Host-based CD-ROM Data Recovery

Jacob Hauenstein
University of Alabama in Huntsville

RCEU 2023 Project Proposal

Project Title

Host-based CD-ROM Data Recovery

Faculty Information

Name: Jacob Hauenstein

Status: Assistant Professor

Department/Program: Computer Science

College: Science

Phone: 256-824-6186

UAH Email: jacob.hauenstein@uah.edu

Proposal ID RCEU23-CS-JDH-01

RCEU 2023 Project Proposal

I. Project Description

There is wide interest in recovery of data stored on aging and/or damaged digital media. For example, forensic investigators are frequently tasked to recover data from damaged media for investigative purposes. As another example, archivists are interested in ensuring that software, video games, audio recordings, video recordings, etc. stored on such media are preserved for future generations. And, as digital media ages and degrades, the problem of data recovery from damaged media becomes an integral part of this data preservation process.

While the problem of data recovery from damaged digital media is an important one, existing methods for data recovery often perform poorly (in terms both run time performance and recovery success). One area where much progress stands to be made (and the focus of this work) is in the recovery of data from damaged CD-ROMs (henceforth, CDs).

CDs are an especially interesting area for data recovery research, because they include multiple interleaved layers of error correction (henceforth, EC) data that can be used to recover data from damaged discs. Thus, CD EC is an iterative process whereby EC performed at one layer may allow further EC at a prior layer, making CD data recovery a many-step, compute-intensive process whereby EC layers are processed repeatedly. In data recovery workflows, this EC is done by the optical drive itself. Thus, current data recovery efforts are limited by the relatively limited computing resources of the optical drive. In contrast, it is possible to perform error recovery using the computational resources of the much more powerful host PC, enabling more (and faster) EC iterations (thus recovering more data more quickly than existing approaches).

This project seeks to develop new methods for recovering data from damaged and aging CD-ROMs. These methods will work by bypassing the EC process within the optical drive and instead performing EC using the resources of the host machine (where processor and memory resources are much less limited). These tools and methods will potentially enable more effective data recovery than traditional CD recovery methods.

II. Student Duties, Contributions, and Outcomes

a. Specific Student Duties

The student will build upon an existing rudimentary codebase (written by the faculty mentor). Currently, as verified by the faculty mentor, the existing codebase can do the basic process of interacting with the optical drive and reading raw, uncorrected sectors. This code base will be expanded by the student to enable it to perform error checking and error correcting, do caching of valid data between re-reads, etc. As part of this process, the student will consult necessary architecture and standards references (e.g., the so called “Yellow Book” CD-ROM standard).

b. Tangible Contributions by the Student to the Project (10% of Review)

The primary contributions by the student will be: (1) implementation of several (potential) methods for data recovery of CDs (previously identified by the faculty mentor), (2) testing of

these methods (on actual damaged CDs), (3) expansion of these methods (based on the test results and discussions with the faculty mentor), and (4) writing a conference paper summarizing the methods and results (alongside the faculty mentor).

c. Specific Outcomes Provided by the Project to the Student (30% of Review)

It is expected that the student will carry away several outcomes by completing this project. The student selected for the project, Danial Ahari, has expressed that he has a personal interest in archival and data recovery. As such, it is expected that Danial will come out of the project with experience and knowledge that will help him continue to pursue one of his personal interests.

Academically, it is expected that Danial's experiences on this project will further his understanding of the computer science curriculum (especially algorithms, data structures, and architectures). Specifically, the code to be written will need to (1) use various data structures (for buffers, lookup tables, etc.), (2) require development of novel algorithms (e.g., for retaining valid data between reads) and implementation of standard algorithms, and (3) require some (limited) knowledge of architecture topics such as caching and I/O.

There will also be beneficial outcomes for Danial's research and career goals. In terms of research, he will gain experience in both research writing and general research workflow. For his career, Danial will have some experience interacting with I/O APIs, consulting standards documents, and performing other tasks that are potentially career relevant but often not a substantial component of the computer science curriculum.

III. Student Selection Criteria

The work has been constructed so that it will be manageable by any competent Sophomore (or higher) level computer science student. Danial was selected because he expressed great enthusiasm about the project and demonstrated great knowledge during our interview.

IV. Project Mentorship *(30% of Review)*

The faculty mentor has a personal interest in this topic and thus is prepared to ensure the student's success. First, as noted previously, the mentor has already implemented some rudimentary code upon which the student will build. Second, the mentor has identified, broadly, several approaches that may work toward solving the problem. The mentor will work with the student (in regular meetings) to further refine and fill in missing details for these methods. The details the student learns in these discussions will help the student understand the problem and, hopefully, identify his own potential solutions.

The mentor will provide one (or more) optical drives, various test discs, and lab space for the student to work in. The mentor has previously written multiple papers / done research work on a related topic and thus is comfortable and ready to help the student get going with the hardware and development environment.