

Host-based CD-ROM Data Recovery

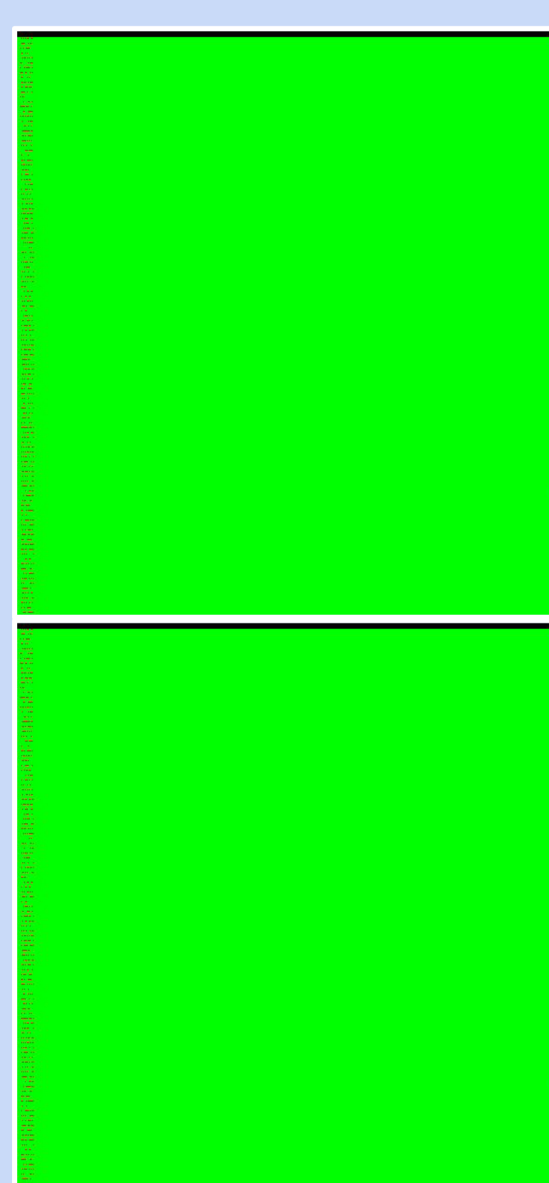
Danial Ahari, mentor Dr. Jacob Hauenstein
Department of Computer Science

Introduction

- Compact discs are typically very resistant to damage and corruption.
- Even still, the on-disc systems for error detection and correction (ED&C), driven by Reed Solomon Codes, can reach the limit of their usefulness.
- Results here indicate that it may be possible to recover additional data by bypassing the drive's ED&C and doing ED&C on the host computer.

Methodology

- Using open source resources, we designed an algorithm to correct errors that proceeded as follows:
 - a. Read a sector from the CD.
 - b. Process the sector and output scrambled data.
 - c. Descramble data and perform error correction.
- A direct replica of the methodology used in a drive, such as this, allows us to make direct comparisons between traditional images and our images.



Pictured above is a view of the data side of a CD containing the European release of Tropico, with the visualizations of dumps of the disc on the right. The top visualization is from the control method, and the bottom is from the experimental method. Note the identical nature, save for the sectors being offset by varying degrees, where green sectors are good, and red sectors contain errors.

References

1. Nguyen, H. T., et al. Adversarial Neural Networks for Error Correcting Codes. 2021, <https://doi.org/10.1109/GLOBECOM46510.2021.9685472>. Scopus.

Acknowledgements

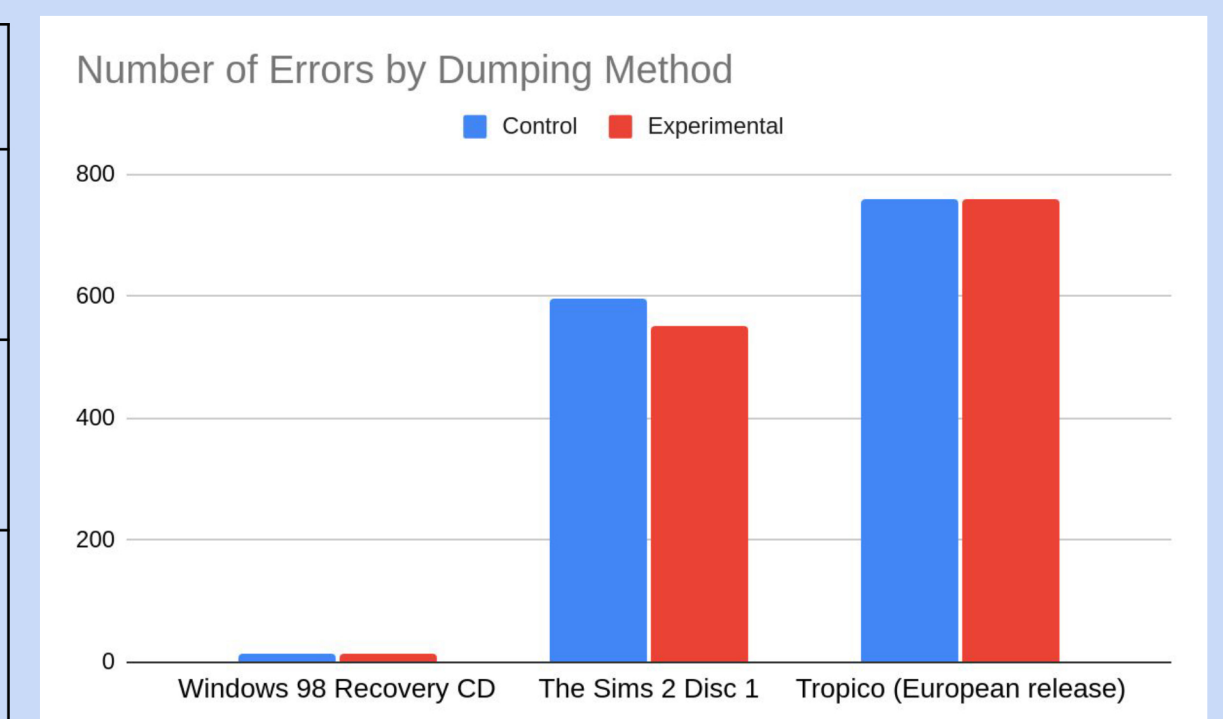
Many thanks to GitHub users clautia (Natalia Portillo), mersinvald (Mike Lubinets), as well as author of part of DisclImageCreator, Jonathan Gevaryahu for their contributions to the open source community, without which this study would not have been possible. Furthermore, thanks to byuu (David Kirk Ginder) and Tom Verbeure for their very enlightening write-ups on CD structure and Reed-Solomon Codes respectively, as well as to the authors of *Reed-Solomon codes for coders*. These writings helped substantially in the development of an understanding of Reed-Solomon codes as used on CDs. Further thanks to the general community of digital preservationists at the Redump forums, whose contributions to the community over the past two decades have made a significant difference in the understanding of how to recover data from CDs.

All RCEU projects were sponsored in part by the Alabama Space Grant Consortium, the UAH Office of the President, Office of the Provost, Office of the Vice President for Research and Economic Development, the Deans of the College of Science, the College of Engineering, the College of Arts, Humanities, and Social Sciences, the College of Education, and the College of Nursing.

Results

Our methodology for error correction was found to be correct equally as often or more often than the built-in decoders found in the CD drive.

Disc Name	Control Err.	Exp. Err.
Windows 98 Recovery CD	15	14
The Sims 2 Disc 1	595	550
Tropico (European)	758	758



Graph and table showing the number of errors in images produced by control and experimental dumping methods, by CD dumped.

Notes:

- Some insignificant results (no errors with either method) were not included in this graph.
- Certain methods of copy protection offset the number of errors making reads of damaged discs with copy protection difficult and open to interpretation (Sims 2; Tropico)

Impacts

Though our findings show slim margins between the two methods, they do shown that it might be possible to recover additional data from CDs by performing ED&C on the host computer. This has a significant impact on the computational aspect of archival science due to the potential for archiving a perfect image of a disc that is no longer in perfect condition.

In the Future

Further research might look at:

- using neural networks to search for valid data, as Nguyen et. al. [1] have already done for general Reed Solomon codes.
- implementing the use of non-coherent light to read discs in a user-friendly way.

View the GitHub!

