C25 Forensics for Floppy Disks and Recordable Compact Discs

Jin Xie, PhD, Ahmet Kaya, PhD, B.V.K. Vijaya Kumar, PhD, and James Bain, PhD, Carnegie Mellon University, Data Storage Systems Center, ECE Department, CMU, 5000 Forbes Avenue, Pittsburgh, PA 15213*

Can it be determined whether a floppy disk or a recordable compact disc (CD) is recorded in a particular drive or not? Do the physical systems in these drives leave traceable "fingerprints" on the disk? How is the information extracted to determine which drive recorded a particular disk? How can these methods be applied to a recordable CD (CD-R)? The goal of this presentation is to address these questions. The authors will show that traceable information does exist on the disk and that image and signal processing methods can be used to extract these fingerprints from floppy disk and CD-R.

This presentation will impact the forensics community by introducing the authors' investigation of the problem of determining which drive recorded a given floppy disk or CD-R disc. Image and signal processing methods will be described for floppy disks and CD-R discs. Experimental results will be presented to show that identical floppy drives can be discriminated and different brand CD-R drives can be discriminated.

Content: A 3.5-inch, 1.2MB floppy disk has 80 tracks in which information bits are recorded, and it has 79 gaps between the tracks. Track widths reflect width of the magnetic head of the drive creating this disk, and gap widths reflect movement of the magnetic head. Therefore, track and gap widths of a disk can be considered as the "fingerprint" that a drive leaves. In this research, the authors took images of disk surface under a microscope, and used image processing methods to extract track and gap widths from the pictures. The 159 estimated widths (80 tracks, 79 gaps) can be considered as a feature vector in the 159-dimensional space characterizing that drive.

A 5-inch, 700MB CD-R has also been studied. When an optical head writes bit "1" onto the CD-R, it burns a "pit" on the medium. When it writes a "0", it leaves the medium untouched, which is called "land". Pit and land represent the information bits. Different CD-R drives may have subtle difference in their laser powers and optical efficiencies, and lead to pits with different lengths, which is called "bloom." The amount of bloom reflects the laser power, therefore can be considered as a fingerprint of CD-R drives. When the CD-R disc is being read, the readback signal (i.e., the output of the optical pick-up) can be used to estimate the amount of the bloom. In this research, the analog readback signal is captured by an oscilloscope, from which the information bits are extracted. From the information bits and the analog readback signal, a signal-dependent auto-regressive (AR) model is trained. Coefficients of the AR model can be used as these features reflect the amount of the bloom. The authors' experiments used 26 coefficients, so a CD-R drive is characterized by a point in 26-dimensional space.

Results: In the experiments, 10 floppy disks from drive A, and 9 from drive B were formatted. The two drives are of the same brand (TEAC), and the same model (model # FD235-HF-A429). One "test disk" was selected, and used the remaining 18 disks to train a classifier using Support Vector Machine (SVM) or Fisher discriminant function, and tested the classifier using the "test disk". Nineteen trials were performed, each time with a different one of the 19 disks as the "test disk." All 19 "test disks" were correctly classified. The same experiments were performed on the CD-R drives, 18 discs written from drive C, and 10 from drive D. Drives C and D are of different brands (C: RICOH, D: Dell). All the 28 CD-R discs were correctly classified. However, discriminating same brand, same model CD-R drives turned out to be difficult.

Conclusions: Drives do leave traceable "fingerprints" in the disks recorded. The authors have developed image processing method to extract features from floppy disks, and signal processing method to extract features from CD-R discs. In experiments, two identical floppy disk drives were successfully discriminated, and two CD-R drives of different brand were successfully discriminated.

This research is supported in part by The Technical Support Working Group (TSWG). TSWG originated the idea for this Link Analysis of Computer and Media research project and provided support for it.

Floppy, CD-R, Classification

 ** Presenting Author*