Michael A. Rolenz
10 February 2003
 "Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies"

Reply comments in support and amplification to Comment #24 of Christopher Lewis

2. Descriptive Name Class : Data archival mechanisms

Summary:

The files created by computer programs are becoming increasingly complex. Most software allows users to create their own copyrighted works that are contained in those files. While every file generated by a computer is an ordered sequence of bytes of data, the ordering of those bytes does not correspond to the same order of information. Files are increasing containing scripted languages or microcode programs. Reconstruction of the information can be impossible unless the program that writes that information is known. The computer program that writes that information is protected by access controls in the form of end user agreements that prohibit reverse engineering, copy protection schemes, registration and access keys. The ability to circumvent these for the legal access of copyrighted works created by software programs is problematic and an exemption is needed to clarify this.

Facts and evidence:

The file formats that are generated by computer programs are becoming increasing complex. No longer are they merely a sequence of text records. Files use an extended non-textual character set. The sections in a file need not correspond to the order that they are presented. For example, word processing files may store the earlier sections of a book after the later ones if they are written after the latter[1]. Files often contain embedded formatting commands, scripts, macro instructions or microcode the tell the program how to interpret the file. Often these file formats are proprietary, undocumented, and too complex for reverse engineering[2]. Least the Librarian believe this is theoretical, here are some of the problems regarding data archival I have had to deal with for several years.

Recovery of Emailbox after file damage: Earthlink refused to provide file formats to me after my mail box got "corrupted" by some damage to some file. After a year, Earthlink provided a utility to recover corrupted mail boxes. I have yet to figure out how to

---

[1] As was done with the popular program Multimate in the 1980s. Three years ago I had to convert a coworkers work to Microsoft Word, which did NOT have a conversion program for that format. Fortunatly, a copy of Wordperfect 4.0 that was being discarded did have a conversion program that converted it to Wordperfect 4.0 and Microsoft Word97 was able to read that.
[2] Much of the Earth Sensor data taken by satellites by NASA in the 1960s and 1970s is unrecoverable. Even after tape readers were created at great expense, the information on the tapes could not be recovered because the formats were not documented and the programs that wrote the tapes were lost.

integrate my Earthlink mailbox with the program I now use. I would prefer to not have to cut and paste email messages but the exact formats of the mail boxes still eludes me.

Multimate Conversion: See footnote.

Chiwriter: A 300 page document I created in the late 1980s using that scientific word processor is inaccessible. Later versions of Chiwriter do not read files created by version #1. Now no software reads any Chiwriter files.

Email conversions. In the 1990s, I converted between 6 different email programs in less than two years. The last was Lotus Notes. The latter has a complicated proprietary file structure with their own scripting language that is intolerant of any damage to the file (even a bit error will render a Lotus Notes mailbox unreadable). Lotus mail boxes are also encrypted on the server.

Microsoft Word incompatibilities: I have documents I created with Microsoft Word (Version 7 Office 95) that cause Word Version 8 and higher to "crash". Commercial Conversion programs do exist but these do not do a complete conversion. In particular, mathematical equations, which are time consuming to typeset, are converted to graphics files that cannot be altered or edited again.

Argument:

Computer file formats have become increasingly complex. Files contain what is clearly data but also instructions for the computer in the form of scripts and microcode. The latter is covered by copyright. Without access to the program that generates those files, it can be impossible to interpret that file[3]. The embedded scripts and microcode are copyright material. The program that generates the files for the users is also. When that program contains access controls, then the DMCA applies.

Yet where one owns the files, either by virtue or creation or other means, one has a legal right to access those files. But by obsolescence, unavailability, or capriciousness on the part of the creator of the program, the copyrighted works created by users become inaccessible and still the work would appear to be protected by circumvention by the DMCA. Clearly, a pro forma exemption is needed for users to access works created by them under the above circumstances.

---

[3] This is not conjecture based upon the NASA data in the previous footnote. This is demonstrable using complexity theory known in mathematics and computer science.