Technological Access Control Interferes with Noninfringing Scholarship[*]

Andrew W. Appel
Professor of Computer Science
Princeton University[†]

35 Olden Street
Princeton NJ 08544
(609) 258-4627
fax: (609) 258-1771
appel@cs.princeton.edu

Edward W. Felten
Associate Professor of Computer Science
Princeton University

35 Olden Street
Princeton NJ 08544
(609) 258-5906
fax: (609) 258-1771
felten@cs.princeton.edu

February 17, 2000

When copyrighted works are distributed in digital form, publishers have sometimes attached "technical access control" – such as encryption – that controls the mode of use. For example, an electronic book could be viewed but not printed, or a movie could be shown on a screen but not copied. Technical access control prevents many kinds of uses of the material that would not be considered an infringement of the copyright, however: making a copy for personal use, searching the material, and so on. Therefore, until 1998 it was legal for the owner of a copy of a copyrighted work to circumvent the access control for a noninfringing use. (Circumvention for the purpose of infringing on the copyright – e.g., redistributing copies – has never been legal.)

The Digital Millennium Copyright Act of 1998 provides that "No person shall circumvent a technological measure that effectively controls access to a work protected under this title" and makes circumvention a federal felony. However, the Librarian of Congress is given regulatory power to exempt certain classes of work from the prohibition on circumvention of technical access restrictions. (It is interesting to note that there is a quadruple negative in the law!)

The Copyright Office of the Library of Congress has solicited public comment as part of the process of determining whether there are particular classes of copyrighted works that shall be exempted from the prohibition because persons who are users of those classes of works "are, or are likely to be in the succeeding 3-year period, adversely affected by virtue of the prohibition in their ability to make noninfringing uses of that particular class of works under this title."

We have submitted this document – not including the explanatory material in this box – to the Copyright Office as a public comment.

[*] This is a response to the Copyright Office's request for comments [CO99] on what classes of works should be exempted from the Digital Millennium Copyright Act's prohibition on circumventing technological measures that control access to copyrighted works.

[†] The views expressed in this document are those of the authors, not necessarily those of Princeton University. Affiliation is listed only to identify the authors.

Abstract. Modern scholarship increasingly relies on sophisticated computerized analyses of copyrighted works. Technological access control schemes that prevent novel computerized analyses of works prevent fair use and impede scholarship, and are therefore counter to the goals of copyright law.

Introduction. Scholarship would be impeded if scholars lost the ability to use computer programs of their own devising to analyze the full digitized versions of copyrighted works. We provide specific examples of scholarly projects that rely on this ability. The examples apply to works that are in the form of text documents, musical scores, audio, video, and computer programs.

These facts justify a finding that scholarship is impeded by the anti-circumvention prohibitions in the Digital Millennium Copyright Act, with respect to works in the form of text, musical scores, audio, video, and computer programs.

Simple search of books. Suppose that Alice, a scholar who owns a roomful of books, wants to search all of the books looking for references to Francis Bacon, accumulating a list of citations. Alice may employ an assistant to skim through the books and collect this information. Similarly, if Alice owns a collection of copyrighted books on digital media, she may want to perform similar searches electronically. Whether a human assistant or a computer program searches the books is legally immaterial; employing a computer program to search the books is fair use.

Computers offer many practical advantages for search applications. It might be prohibitively expensive to search a large collection by hand, but doing the same search on an inexpensive computer might provide an instant result. Thus manual searches cannot substitute for computerized searches.

Laws that prohibit scholars from using computerized "assistants" artificially impede the progress of scholarship and science. If the digital works are technically protected in such a way that they can be viewed on the screen but not electronically searched, then the technical protection interferes with noninfringing uses.

In this scenario, the publisher may meet Alice's needs by providing a generic text search facility. Alice could search for the words "Francis Bacon", or perhaps "Bacon" and sort through the results of the search manually. Although a generic publisher-provided search facility can satisfy Alice, we will see below that such a facility fails to meet the needs of many other scholars.

Thematic search of musical scores. Suppose that Bob, a scholar who owns a collection of musical scores, wants to search the collection looking for the occurrence of a particular musical theme. Copyright law permits Bob to do this; whether a human assistant or a computer program performs this search is legally immaterial. Technical protections on digital works that prevent computerized searches (on privately owned copies) interfere with noninfringing uses of copyrighted works.

Searches of this type have many research uses in musicology. Indeed, entire research centers, such as the Center for Computer Assisted Research in the Humanities (at Stanford University), focus on technological search and analysis of music. There is a great deal of active research on how to encode musical scores for computerized analysis and how to perform the analyses. (Selfridge-Field's book [SF97] summarizes research in this area and provides many citations to the research literature.)

Musicology researchers perform several kinds of operations on musical scores. They translate the scores into different electronic formats to facilitate analysis. They develop novel search and analysis criteria to represent abstract concepts such as "musical themes". They develop novel search techniques to efficiently find certain patterns in encoded musical scores.

These activities all require the ability to write computer programs that analyze a score directly. Unless the publisher of an electronic musical score provides scholars with the ability to write computer programs that directly access the score, scholars will lose the ability to perform these kinds of analyses.

Note that generic publisher-provided search facilities cannot possibly meet this need. Researchers are constantly developing new and better search methodologies. Confining scholars to any particular search facility will impede research on new search methods.

Thematic search of a musical work. Suppose that Claire, a scholar who owns a collection of musical recordings, wants to search the collection looking for a particular musical theme. Like Alice and Bob, Claire has the right under copyright law to do this, using either a human assistant or a computer program.

Claire finds herself facing a more difficult research problem than Bob faces. Effective searching through audio recordings of music is a very difficult research problem that has seen steady but slow progress over the last twenty years, for example in the research on "structured audio" [VGS98]. Active research groups in this area need access to a wide variety of recorded musical works in order to prototype, test, and improve their technology. Like Bob, Claire needs to write computer programs that access the original work directly.

Video. Suppose that David, a public-health researcher who owns a collection of recorded movies, wants to search the collection looking for depictions of cigarettes and related paraphernalia. David has the right under copyright law to do this.

The algorithms for doing this automatically are not yet mature, but an active and robust discipline of "video content analysis" [SZ99] or "object-based video coding" [PCK$^+$99] is seeking to provide tools for this kind of query. Research in these areas progresses by devising computer programs that take video content as input. The research would be severely inhibited if scientists cannot get access to the actual video content of the works they purchase, but are limited by restrictive interface mechanisms to on-screen viewing or specific kinds of searches.

Innovative Text Searches and Analysis Modern scholars of Shakespeare analyze the frequency of word usage in the different plays. Shakespeare is known to have acted the role of the ghost in Hamlet. Donald Foster of Vassar College used statistical computations to notice that specific words that the ghost speaks appear more frequently in the next play that Shakespeare wrote — it's as if they were on his mind while writing the next play. In each play, there seems to be one role whose words appear more frequently in all roles of the next play [Dol91].

This particular kind of statistical analysis could not be foreseen by a publisher of the texts of Shakespeare's plays. Almost any generic search-engine interface would be too limited to calculate the specific correlations necessary for this analysis. To efficiently perform a computerized test of this theory that Shakespeare acted in all his own plays, the full text of the plays must be readable by a computer program of the scholar's own devising.

Innovative Analysis of Computer Programs The same kind of analysis that Foster applied to Shakespeare can be applied to computer programs. A large computer program is typically written by many programmers, each contributing a part. An automatic analysis of the program might correlate this data with the engineering practices used to develop the program. Such data could be useful in developing codes of engineering practice. Technical mechanisms that restrict access to the computer software inhibit scholarly research on the program.

Computer programs are usually protected not just by copyright law, but by licenses. The license contracts may prohibit the kind of analysis we have described. We do not wish to address the legitimacy of such constructs. However, the law of copyright would naturally consider such analysis to be a noninfringing use, and therefore the regulations of copyright should not sanction technical protection mechanisms that interfere with this use.

Conclusion. Technical protection mechanisms that prevent computer programs from accessing the underlying content of copyrighted materials will hinder legitimate noninfringing uses that are vital to scholarship and science, specifically in the domains of natural language text (such as books), musical scores, musical performance, other audio material, video and movies, and computer programs.

References

[CO99] Library of Congress Copyright Office. Exemption to prohibition on circumvention of copyright protection systems for access control technologies. *Federal Register*, 64(226):66139–66143, November 1999. http://lcweb.loc.gov/copyright/fedreg/64fr66139.pdf.

[Dol91] Edward Dolnick. The ghost's vocabulary. *The Atlantic Monthly*, 268(4):82–86, October 1991. http://www.theatlantic.com/unbound/flashbks/shakes/dolnick.htm.

[PCK+99] Fernando Pereira, Shih-Fu Chang, Rob Koenen, Atul Puri, and Olivier Avaro. Introduction to the special issue on object-based video coding and description. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(8):1144–1145, December 1999.

[SF97] Eleanor Selfridge-Field, editor. *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, Cambridge, MA, 1997.

[SZ99] Emile Sahouria and Avideh Zakhor. Content analysis of video using principal components. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(8):1290–1298, December 1999.

[VGS98] Barry L. Vercoe, William G. Gardner, and Eric D. Scheirer. Structured audio: Creation, transmission, and rendering of parametric sound representations. *Proceedings of the IEEE*, 86(5):922–939, May 1998.