OITP Technology Policy Brief

Digital Rights Management: A Guide for Librarians version 1

Prepared by Michael Godwin

Copyright and DRM

Not long ago, digital technologies were regarded as being entirely beneficial to the work of librarians, because such technologies were already enabling greater access to collected materials, greater ease and searching or organizing such materials, and greater ability to reproduce and archive creative works, historical documents, scholarly research, and other important resources. *At its heart, this early perception of the usefulness of digital tools remains essentially correct*. Nevertheless, the digital revolution has also inspired the development of a range of technological tools and strategies aimed at restricting the ease with which the resources collected and maintained by libraries can be used, circulated, excerpted, and reproduced.

These technological tools and strategies are generally referred to as "**digital rights management**"-- a term commonly reduced to the acronym "DRM."[1] To put the matter another way: "digital rights management" is a collective name for technologies that prevent you from using a copyrighted digital work beyond the degree to which the copyright owner (or a publisher who may not actually hold a copyright) wishes to allow you to use it. The *primary* purpose of this paper is to familiarize librarians, archivists, and others with DRM and how it works. *Secondarily*, this paper will outline certain legal and policy issues that are raised by DRM -- issues that will continue to have an increasing impact on the ways in which librarians and libraries perform their functions. *To put the matter bluntly -- understanding the basics of DRM is becoming a necessary part of the work of librarians.*

DRM also, of course, has an effect on how copyright law functions in our society. Librarians often have relied on the provisions of copyright law to ensure that libraries can perform their proper functions. But DRM technologies can be used to "trump" copyright law by depriving librarians --

[1] In international policy-making circles, such as the World Intellectual Property Organization (a division of the United Nations), another term is as common as -- or perhaps more common than -- "digital rights management." That term is "technological protection measures" (frequently abbreviated as "TPM"). The meanings of "DRM" and "TPM" are not precisely the same; the latter term also encompasses non-digital protection technologies such as those used to prevent analog videotapes from being copied. Nevertheless, the terms are often used interchangeably in policy circles. In general, this paper will stick to the terms "digital rights management" and "DRM."

and citizens in general -- of rights they are granted by copyright law. Such an overbroad use of DRM technologies may raise questions about whether we should ask legislators or the courts to bring the uses of DRM into line with the rights we as citizens have come to expect under copyright law. Whether legislative or judicial intervention occurs may depend on how careful technologists are not to overstep the boundaries of the rights they are "managing" with "digital rights management." Where might they overstep those rights? Some examples could include:

(1) **Copyright terms**. DRM may be used to prevent copying of works beyond the terms of their copyrights, or may be used to impose copy protection for works otherwise in the public domain, or works that have been licensed for less restricted general uses.[2]

(2) **Preservation and archiving**. DRM may be used to limit librarians' ability to preserve or archive a work, e.g., by preventing a work from being moved from more perishable to more permanent storage media.

(3) **Artistic creation**. It has long been understood that the creation of new artistic works may require the excerpting or transformation of older ones; DRM may be used in ways that prevent such excerpting or transformation from happening.

(4) **Historiography**. Historical research fundamentally depends on being able to access and quote older documents and other kinds of works -- DRM can be implemented in ways that make historiography far more difficult, if not impossible, in many contexts.

(5) **Fair use**. The ability of anyone to make unlicensed use of the protected works of others -- within limits established by law and precedent -- can be curtailed by broad implementation of DRM.

(6) **The use of shared materials in learning environments**. DRM may make it more difficult, or even impossible, for works to be used in otherwise lawful ways in both real classrooms and "virtual" ones (e.g., distance learning).

This is not an exhaustive list of the ways in which the use of DRM may create tensions undercutting the balance of rights created by copyright law. These tensions are fundamental and unavoidable ones, although not everyone acknowledges this. What makes them unavoidable is that DRM tends to be precise and immutable, while our copyright law policy tends to be general

[2] See, for example, works licensed under Creative Commons licenses. You can learn more about Creative Commons licenses at http://creativecommons.org/.

and dynamic. What makes them fundamental is that democratic societies need libraries and librarians to support and promote public access to the broadest range of creative and scholarly work. In practice, this means that, for the foreseeable future, there will be ongoing tensions between libraries (and librarians), on the one hand, and DRM-using publishers and rights-holders, on the other.

At first glance, the increasing use of DRM by movie companies, by consumer-electronics makers, and by computer makers may not bother the working librarian very much. You may look over the increasing use of DRM in, say, DVD movie releases and ask yourself why you should care. After all, isn't DRM just a technical means of giving copyright owners new ways to protect the legal rights they already have?

But librarians, and citizens at large, should care about DRM because they have much at stake both in the balances built into our copyright law, and in the technologies, such as personal computers and the Internet, that might be restricted or controlled in order to protect copyright interests. The choices our society makes now about how we may use copyrighted works and about the technological protections for such works will affect us for a long time to come. This is why, as we work through our understanding of DRM, we need to make sure we understand the traditions and principles of copyright as well.[3] Although there is a tendency on the part of some people to equate copyright interests with other kinds of ownership and property interests, under our legal system copyright is actually significantly different.

Basic Considerations of Copyright Law

Copyright law creates a set of legal rights that are different from other rights. Here's one important difference, of particular importance to librarians: copyright law frequently allows other people (sometimes teachers, reporters, or scholarly researchers) to quote a copyrighted work without the copyright owner's permission. In our society, we don't normally make the same kinds of exceptions for unauthorized uses of other kinds of property — for example, you don't get to use your neighbor's car just because he's not using it this afternoon, and you happen to have an urgent need for transportation.[4]

[3] Please note, however, that this paper is not designed to function as a general guide to copyright issues for librarians. It aims, instead, to complement existing copyright-law resources for librarians with a general guide to digital-rights-management and copy-protection technologies and the policy issues specific to those technologies.

[4] When I say we don't "normally" make exceptions for unauthorized uses of other people's property, I am not saying there are no exceptions. The law allows us to make certain uses of other people's property in some kinds of emergencies, for example, and it may allow a government to invoke its power of eminent domain or to authorize a utility company to tear up your front yard in service of some larger public good. Copyright law is nevertheless different because it allows for some routine and regular uses of others' property without their authorization.

Copyrighted works are different in several other ways. They remain the owner's (or her heirs') "property" only for a limited time period, unlike other, older kinds of property. Your house may have belonged to your parents or your grandparents, but the copyright interests of your grandfather may, at some long period after your grandfather's death, cease to be anybody's property. We often refer to this final stage of the legal protection for a copyrighted work by saying it has "become part of the public domain" — a kind of "property" still, but one that belongs to everyone and can be used by and copied by anyone without restriction.[5]

Copyright protection of certain kinds of creative works is built into the Constitution. The language of the Constitution (in Article I) makes clear that the Framers saw value in granting artists and authors (or the people to whom the artists and authors gave their rights in their creations) a kind of "exclusive" right in the created work. This means that copyright law allows the copyright owner to "exclude" other people from copying it, at least so long as the legal term of protection of the work lasts, and so long as their copying or other use of your work doesn't fall within one of the specific exceptions, such as “fair use,” that are allowed for in our copyright scheme.[6]

The Focus on the Act of “Copying”

Why is there such a focus in our legal system on "copying?"[7] The answer is rooted in the history of technology. The technological reasons are that, for almost all of human history, the human capability to make copies of a creative work was very limited. Monks used to spend their lifetimes in the monasteries making copies of pre-existing works, sometimes adding valuable commentary or illustrations. (Then as now, it often helps to break up a block of text with a pretty illustration or a helpful diagram or an italicized headline.) Their copies had to be perfect, without error — and asking human beings to make perfect copies of anything someone else said or wrote is a very demanding thing.

[5] Sometimes such shared “property” is called a “commons,” based on the tradition that there may be property in a township whose use and ownership is shared by everybody in the town.

[6] The American copyright doctrine called “fair use” has no precise counterpart in other countries, but there are some similar protections in other countries for unauthorized copying -- these protections derive from copyright doctrines relating to “fair dealing” and “private copying.” This paper focuses more on the American doctrine of fair use than on these analogous doctrines in other countries, but its reasoning may be applied without much modification in most other countries that, like the United States, are signatories to the Berne Convention (an international copyright treaty).

[7] The term “copyright,” unlike many legal terms, is almost self-explanatory; the term literally focuses on the right to make copies, although over the course of time copyright law in the United States and around the world has increasingly encompassed some related rights as well.

This was true until a number of technological changes made copying easier. The first major change was the invention of the printing press — a huge device that required lots of expertise and maintenance to operate, but that enabled printers to make (and sell) many copies of books that were essentially identical to one another.

The printing press created a great opportunity for authors. Instead of begging for patronage from a rich man, a nobleman, or a church to create something in writing, an author could make a deal with a printer, so that the books or other materials he wrote could be sold — perhaps many copies of the work — and both the author and the printer could reap a direct financial reward. The printing press led to a slow but ultimately pervasive expansion of the number and availability of books in Europe and elsewhere that even ordinary citizens might buy (or, through libraries, might have access to).

All this was because the printing press made copying a certain kind of creative work — written works, perhaps supplemented with engraved illustrations — technologically possible in a fraction of the time it used to take the monks to make a single copy.

When Printing of Books Became Commonplace

Over the course of a couple of centuries, the printing of books became commonplace in Europe. But the process remained expensive, so anyone who wanted to publish a book usually had to find a printer or publisher to sponsor the making of copies for sale. Until very recently in our history, if an author had a book that you wanted to get to a wide audience, he or she had to have the support of a patron to get it published.

More important than the expense and difficulty of printing books, however, was that printing changed the way our culture thought of books (and later, of other creative works that could be copied). We began to think the creative effort of the author was the fundamental thing of value, or at least far more valuable than a single copy of the book might be. (In general, that is; sometimes books are themselves rare, collectable items, but that's not normally the case.) I might buy a copy of a Stephen King novel, and I might even say it's "my book," but on another level I reflexively realize that in some other sense, my book (and everyone else's copy of it) belongs to the author, Stephen King — he has more rights to use its text than I do. When I own a particular copy of a book, I own ordinary "tangible" property; the author, at least until the period of legal protection of the copyrighted work ends, owns the "intellectual property."[8] That is, the author, or the person or group or company he gives his rights to, "owns the copyright."

[8] The term “intellectual property” is a fairly new term in the history of copyright law — only a few decades old, according to law professor Mark Lemley. See Lemley, “Romantic Authorship and the Rhetoric of Property,” 75 Tex. L. Rev. 873 (1997) at footnote 123.

This whole system of "copyright law" — which began in Europe and then was made part of the Constitution of the United States, has worked reasonably well, at least with regard to textual works, for the few centuries it has existed. It has since been extended to other creative works, including paintings, photography, and even architectural design, not always without generating controversy. Overall, however, the copyright system has been seen as a force that makes our culture richer. For one thing, this new publishing-industry system meant that more authors got paid (not so many had to beg for a subsidy from a king or a pope). For another, more artists and authors were inspired to create, because the time and effort it takes to create new works could end up with the artists and authors getting paid, which made the creative effort even more worthwhile, and occasionally even paid the rent. And a larger socially beneficial result of these rewards to publishers and authors has been an increase in access to information and knowledge for the public. Libraries and librarians have played a key role here -- even as copyright holders have distributed their works to the public in order to reap financial benefits, librarians have bought copies of the copyright protected works to share with the public. As a result, citizens don't need to buy every copyrighted work in order to (potentially) have access to every copyrighted work.

The Growth of Copyright Law Raises Questions

Of course, this system has not been without its problems. Before there were comprehensive international copyright treaties, publishers in some countries might print works (sometimes in translation) of authors from other countries right next door, and the original authors might not even learn of this act. Then as now, advocates of international copyright systems have preferred to equate such unauthorized publication as a kind of "theft." In general, however, the word "theft" is a misnomer -- there's a special legal term for unauthorized use or copying of copyrighted works: "infringement". The reason for the difference in terminology is that, generally speaking, copyright infringement imposes a different kind of loss or damage on the copyright holder than the theft of physical property imposes on a traditional property owner. In addition, "infringement" is defined in terms of specific statutorily granted rights belonging to the copyright holder, while "theft" generally is understood in terms of older, common-law property interests that don't require any statutory creation or grant.

Even though it has always been possible for infringers to copy textual works (the copying of non-textual works posed different technological problems), the relative difficulty and expense of making copies of works — even with the invention of better and faster printing presses — kept unauthorized copying of legally protected works at a low level compared to what became possible in the modern era. And international legal agreements made it increasingly difficult for publishers in other countries to pull the old infringement-across-the-border dodge.

All of this began to change, however, in the 20th century, and has accelerated to a surprising degree in the 21st. This is because, when the lawyers and legislators and judges first framed the idea of protecting an author's interest in his or her creative work by focusing on the *copying* of the work, they did so at a time (just a two or three centuries ago) when copying was expensive, and illegal copying was hard to conceal. So it seemed only natural to take the *difficult* part of making unauthorized use of an author's work — the making of a copy — to build a framework of legal protections to protect authors and publishers from illicit copy makers.

With the advent of cameras, photocopying machines, tape recorders, and other consumer-operated copy-capable machinery and tools , however, the idea that the making of a *copy* is the easiest place to enforce a creator's or publisher's right might reasonably be called into question.[9]

Computers and computer networks, as well as other digital tools, have made this question even more acute. So it is no surprise that computer companies and software vendors discovered an aspect of this problem in the late 1970s and early 1980s, because the software that ran on personal computers is inherently copyable. Their experience led to the first efforts at DRM, then known only as “copy protection.” [10]

Many of us are aware how inexpensive it is to make computerized copies of *digital* creative works — the cost probably is too small to be measured. What is less well-recognized (but just as true) is that other digital tools make it possible to take *analog* works (original paintings, say, or printed books) and digitize them and distribute the perfect copies — sometimes as part of an illicit "copyright piracy" enterprise, for commercial gain, and sometimes just for free.[11] In effect, digital tools make the copying of any content, regardless of its form, far easier than it used to be.

[9] These tools have grown both remarkably less expensive and remarkably more powerful over the course of the 20th century.

[10] Throughout this paper, the term “copy protection” is used interchangeably with “DRM.”

[11] It is a common myth that works in digital form are more copyable than works in analog form. The reality is that digital copying technologies make it easier for anyone to copy works in either digital form or analog form. A more complete discussion of the fact that analog content can be digitally copied appears in section II of this paper.

DRM: A Response to Cheap Copying

The problem posed by digital technology, at least as many artists, authors, and publishers see it, is how to make copying of creative works more difficult or at least more controllable. If it is possible to put the breaks on easy copying, then it becomes less necessary to revise the whole system of law we've built around the notion of "copyright."[12]

For those who want to make copying difficult, as it once was, the digital revolution has been both a curse and a blessing. On the one hand, when a work is in digital form, it can easily be copied by digital mechanisms such as the “copy” functions in computer operating systems. But when the work is in digital form, it turns out, there are also a number of technological options that can be employed to limit one's ability to copy all or part of a work. When Stephen King published a novella in 2000 called "Riding the Bullet" and sold it over the Internet, the book was placed in a digital format that limited what you could do with it — there were restrictions on whether you could print any of it at all, or print it out on your printer just because you prefer the feel of paper.

King's experiment taught all of us at least two things: (a) that there's a market for works in digital form that can be downloaded and used on your computer and other digital devices, and (b) some folks who had never read a book online before discovered that, even apart from the difficulties of reading a book on a computer screen, the limitations on a *digital* creative work seemed to be even greater than those on a paperback copy. (At least with the latter you can take the paperback to a photocopier and produce a few page copies that way.)

King's experiment underscored the burgeoning movement by publishers, record and movie companies, and other enterprises that either are creative themselves or that work with creative people to develop and market new works — a movement to find digital tools that put limits on what individual citizens can do with the copies of creative works that they buy. And this new movement brings us back to the set of issues with which librarians and ordinary citizens should be increasingly concerned. Because, as Law Professor Julie Cohen has put it, the traditional copyright system has been helpfully "leaky" (i.e., general and dynamic versus particular and rigid) our culture has benefited from creative works even when those works are still protected by copyright law. For example, many of us know pop song lyrics even though

[12] Other artists and authors and publishers approach the problem differently, however, by suggesting that perhaps compensation of artists and authors should not focus so much on the making of copies of the creative work. There are other models for compensating artists, which include but are not limited to compulsory licensing and levies on playback or recording equipment. This paper does not advocate any of these alternatives; it assumes that the making and selling of copies of creative works will remain at the heart of the compensation system for some time to come.

we may never have bought a sheet of music — the fact that we know them by heart, and even can sometimes sing them to each other, is something that makes our lives a little richer, without really making the creator or publisher any poorer. Indeed, if we sing well enough (or badly enough), we may inspire someone to go out and buy the original recording.

More generally, creative works that can easily be quoted and shown to other people and reused in creative ways help enrich our culture by replenishing it with the latest and best creative works. To put it bluntly, our lives are richer because we get to share so much of our culture even as we continue to maintain a system that encourages artists and authors to create more cultural works.

A Cultural Downside to DRM?

A potential problem with DRM is that, when it's done in the wrong way, it may end up walling off parts of our culture from one another. Worse, the perceived need to use DRM to protect every digital work may cause undesirable changes in the very technologies that have revolutionized our daily lives over the past two and a half decades.

The questions we have to ask now are these:

- What does DRM look like? How do we recognize it?
- Should DRM be administered by the government or developed solely in the marketplace? What limits, if any, should be placed upon it?
- What harms can it do to the balances built into our copyright law? What other harms might DRM cause?
- Are there good forms of DRM that benefit citizens? If not, could there be? What would such forms of DRM look like?

We focus in this essay both on technical issues and proposals and on the legal and policy proposals these technical issues and proposals have generated. Ideally, consideration of these issues will result in conclusions that are applicable across a broad range of proposed copyright-protection schemes, and not just the particular technology-mandate proposals being considered in Congress and elsewhere.

The Forms of Digital Rights Management

Since the personal-computer revolution began, more than a quarter of a century ago, a number of approaches and technologies have been developed to prevent unauthorized copying of, or to otherwise control digital content. There are three broad classes of approaches that are currently used, or that have been proposed in various standards-setting or legislative proceedings. Sometimes these approaches are used by themselves, and sometimes in combination with one another.

Encrypting Content

The first and, still, the most common copy protection approach is encryption — the use of a mathematical/computational process to scramble information so that only those who have the right key or keys can obtain access to it. This, for example, is how DVD movies work — their content is scrambled so that only DVD players that have the right keys can decode the content so that someone can watch the DVD movie. Similarly, if you receive cable or satellite television, your TV service provider normally scrambles content in ways that prevent most unauthorized people (that is, nonsubscribers) from getting access to it.

The basic approach for encryption is to encrypt the digital content so that only a player with both the decryption device or software and the proper key can play the content. The content owner can broadcast the content to everyone but unless the recipient has valid decryption keys he or she cannot play the content. Scrambling is a similar copy-protection approach, but without a user-applied key; instead, the key that includes the unscrambling algorithm resides in the player device (which may be hardware, or software, or both).

There are several varieties to encryption-based copy protection. In the simplest type, all content is encrypted under a single master key. Technologists consider this approach a comparatively fragile scheme because, once the single key is compromised, the entire system is considered compromised or "broken."[13]

Other encryption-based approaches avoid the pitfalls of such easily broken schemes by using schemes that require many different keys. For example, some encryption-based systems encrypt each piece of content with its own individual key. As a result, the loss of a single key only means the compromising of a single piece of content, and not the entire system. In the most complex variants of encryption-based systems, the encryption keys are unique not only for the content but for the player as well. Thus, someone who received both the encrypted content and the encryption key associated with another person could not play the copied content on his own player. (Such schemes have obvious potential to restrict the ability of libraries to lend out protected works; if the DRM scheme limits use of a given copy of a work to one person, the ability of a library to offer that copy to many people -- a routine function of libraries -- is hindered.)

An example of this more secure encryption-based approach can be found in Europe, where pay-television satellite-TV providers commonly use

[13] Cryptographers generally disfavor such "Break Once Break Everywhere" (a.k.a. "BOBE") encryption schemes. This is the approach used by cable-television set-top boxes. Once the scrambling or decryption technology of a set-top box is understood, the box itself can be "pirated" and duplicated, making it possible for the creation of an aftermarket in illegal, cloned cable descramblers. And in fact a number of such descrambler-cloning enterprises exist, in spite of their illegality.

encryption-based copy protection. Subscribers have a hardware set-top box that unscrambles the satellite transmissions into video signals that are then displayed on a standard television. These subscribers also have a so-called "smart card": a personalized device that plugs into the set-top box and controls which television programs can be decrypted and displayed. The satellite television providers transmit their programming in scrambled form, and set-top boxes with authorized smart cards are able to unscramble programs for viewing. The TV distributors can also send instructions to individual smart cards, authorizing them to unscramble certain television shows or prohibiting them from unscrambling others.

A similar scheme, called Content Scrambling System (also known as "Content Scramble System" or "CSS"), has been used by the DVD divisions of the movie and television industries. Under this approach commercial DVDs are encrypted with a series of keys. These keys are embedded in different video players, whose manufacturers are licensed to build them into their products. As a general result, only "authorized viewers" (those with legitimate, authorized DVD players that use authorized keys) can watch the DVDs. The Content Scramble System was designed so that the administrators of the system are able to "turn off" certain keys if they are compromised — so that some DVD players with now-deauthorized keys could be shut out of playing new DVDs without causing the entire system to fail.

The CSS scheme has already been demonstrated, however, to be vulnerable. In its current form, CSS has already been compromised by a Norwegian citizen named Jon Johansen, who came up with a generalizable workaround for the DVD scrambling system. That workaround computer program, called "DeCSS," makes it possible for sufficiently sophisticated users to sidestep DVD scrambling and render DVD movie content unscrambled (a.k.a., "in the clear"), so that it can be viewed on any player. More important in the context of copyright policy, the unscrambled work can be copied on the Internet and elsewhere.

In spite of this breach of the CSS system, however, movie companies continue to produce DVD movies using the existing scrambling system. As a practical matter this breach has not hurt the DVD market, which has continued to see remarkable growth in the period since Johansen published DeCSS.[14]

While it is not yet evident that Johansen's "crack" of the CSS system will cause long-term harm to the DVD market, the failure of CSS to survive a deliberate attempt to circumvent it has spurred both the content companies and the computer and consumer-electronics companies that produce

[14] See, e.g., Hernandez, "DVD sales up 57% in 1st half of 2003," Los Angeles Daily News, Aug. 4, 2003, republished in The Arizona Republic's online edition at <http://www.azcentral.com/arizonarepublic/business/articles/0804dvds04.html>.

players for DVDs to explore alternatives in the delivery of content that may be more secure. The current push by the media companies towards "high-definition" DVDs is widely seen as a significant opportunity to enhance CSS or replace it with a superior DRM technology.

Personal Computers and "Trusted Computing"

It is widely accepted that personal-computing devices will increasingly be a platform for the delivery of copyrighted content. This prospect scares content companies, because personal computers and other digital devices have historically been far easier for individual users to investigate and reprogram.[15] As a result of this widely shared perception, major information-technology companies, including Intel and Microsoft, have invested significant resources in an approach that would deliver content in encrypted form, and allow it to be decrypted only within a tamper-resistant environment within a computer or other device. This approach relies on a new design feature for Intel-based personal computers — a design feature that goes by various names, most commonly "trusted computing" or "the Next Generation Secure Computing Base" (NGSCB).[16] This approach has its advantages — notably, it doesn't have the flaw of being "Break Once Break Everywhere" — but it also restricts users of this content more than they are restricted by analog versions of the content, or by previous digital versions of it.[17]

Unlike variations of the encryption approach that require that all content be encrypted under a single key, content-protection approaches that rely trusted computing, which use a separate key for each computer or other playback device, can be used to "tether" content to that particular device. The notion that content may be "tethered" to a particular computer or other device, or to particular individual users is comforting to many in the content industry, because doing so could drastically limit the extent to which content can be copied and redistributed on the Internet (or by any other means). It is troubling, however, to many copyright scholars, who believe that an individual's ability to give away or sell the copy of a work that he or she buys, or a

[15] The terms "hacker" and "hacking" used to refer to such investigation and reprogramming, although in recent years they have been more commonly used to refer to "computer intrusion" and other presumptively antisocial acts. It is not uncommon nowadays to see these terms, with their current negative connotations, applied to those who merely investigate and/or modify tools and content they already own.

[16] NGSCB was formerly known as "Palladium," its in-house codename at Microsoft Corp.

[17] Advocates of "trusted computing" make a point of stressing that trusted computing is not itself DRM, but merely may be used as the basis of a DRM scheme. That observation is certainly true, but the available public reports concerning trusted computing suggest that content protection and control (by the content provider) were among the initial motivations for the development of this technology. It also seems likely that content-protection schemes will be among the first and foremost uses of trusted-computing technologies.

library's ability to lend books that it purchases — this right is referred to by copyright scholars as the First Sale Doctrine — is central to the balance of rights built into our copyright laws. It is also troubling to many consumers, consumer advocates, and businesses, because they expect to have the right to give away or resell the copies of copyrighted works that they buy. Used bookstores, for example, depend on this right for their very existence. And many libraries routinely sell off lesser-used parts of their collections in order to make room for new works. (Obviously, if the works in question are purely digital in form, making room for new works is less of a problem, but many digital works -- *e.g.*, some CD-ROMs and movie DVDs -- still take up space in libraries, even though their content is essentially digital.)

"Marking" Digital Content for Copy Protection

A second major approach to copy protection is something we can call "marking;" it depends on adding a mark in some way to the digital content. The mark may be used to indicate that the content is copyrighted, and in some cases it also carries instructions about what uses of the content are authorized. For example, in theory a mark may label some content as "do not copy" and another mark may label some other content as "copy once but don't re-copy."

There are three general forms a mark may take in the digital world. First, it may take the form of a **simple label** that is sent along with the content. Second, it may be a **"watermark"** - an arrangement of digital bits hidden in the background of the digital content. Or, third, it may be a **"fingerprint"** — a unique identifier that is derived from the characteristics of the content itself.

Marking is typically used for one of three reasons. First, it is used when an encryption-based method, for whatever reason, is not viable (or is not perceived to be viable). For example, if the Federal Communications Commission requires that broadcast television signals not be encrypted — that they be broadcast "in the clear" — any DRM for broadcast television signals must be based on marking.[18] Second, marking is used in systems that attempt to detect copying after the fact rather than preventing it — such use is among the so-called "forensic" uses of marking. Putting a mark on a piece of digital music, for example, allows one to create a search engine that can find a marked clip on the Internet, which the searcher might then assume is an unauthorized clip (on the theory that *authorized* marked clips aren't available at all via the Internet). Moreover, if the mark is sufficiently sophisticated, it may carry information that can be used to determine where the unauthorized content originated (*e.g.*, from a movie-studio employee).

[18] Ironically, however, the typical protection scheme offered for marked TV content that is broadcast "in the clear" is to encrypt it after it has been " demodulated" (that is, received by a TV receiver). Under this scheme, marking the content is not really an alternative to encryption-based protection; it simply requires encryption at a different point in the transmission chain from broadcaster to audience. Why policymakers might consider this a better alternative than simple end-to-end encryption, especially given that the content is "in the clear" and unprotected for most of the distance it travels, is unclear.

The third major use of marking is as a response to the so-called "analog hole." The term "analog hole" refers to the ability of a would-be infringer to capture content as it is being played (or just before it is played). One obvious example of this would be playing of a DVD and capturing the DVD's content by using a camera with a microphone, or by replacing an output device such as a television set with a recording device, or by connecting to the digital player through analog connectors.

Encryption-based methods by themselves do not address the analog hole, because content must be decrypted in order to play it. At the point of playing, decrypted content is, at least for the moment, "in the clear" and can be captured in a number of ways and redigitized.[19] By contrast, some kinds of marks may remain attached to the content even as the content is being played, which is the basis for some models of content-protection schemes.

Simple Marking Schemes

The simplest type of mark consists of a straightforward label that is sent along with the content. Think of it this way: in effect, a "simple mark" approach is one under which a copyright notice is paper-clipped to a document. The "simple mark" approach is especially cheap and simple to implement — although only at the content-production level — because the mark is easily located and is separate from the content. The costs associated with marking content are not, however, necessarily cheap and simple to implement on the hardware side — there are many costs associated with upgrading computers and other digital devices to recognize the mark. (Librarians should be aware, furthermore, that when they buy new digital equipment for library-goers to use, that equipment may come with built-in mark-recognition technologies that restrict how individuals can view or otherwise use digital works.) Moreover, a simple-mark scheme has other weaknesses -- a simple mark may not carry enough information to indicate the precise origin of the unauthorized content, and since even if it does carry that information the mark may nonetheless be easy to remove.

The most prominent example of a simple-mark or simple-label scheme is the "broadcast flag" scheme recently adopted by the Federal Communications Commission (although later struck down in a federal court challenge). The "broadcast flag" is a simple mark that can be attached to a digital-television broadcast. (Note that the term "broadcast flag" is often used, inaccurately, to refer to a much broader and more complex broadcast protection scheme of which the flag itself is only one small part; the "broadcast flag" scheme is

[19] The most obvious way to digitize (or redigitize) analog content is to point a digital movie camera at a movie or television screen. A less obvious way, perhaps, is the use of a personal video recorder, such as TiVo or Replay TV, to capture analog television content in digital form. Many digital television receivers have been produced with analog outputs so that they can be integrated into existing home-entertainment systems. This means such receivers can be used to "route" in-the-clear digital content through analog interfaces, after which the content can be redigitized without any protections associated with it.

outlined in greater detail later in this paper.) Here we use the term “broadcast flag” to mean the digital mark by itself — “broadcast flag scheme” is the term used for any larger framework that is based on systematic mechanical recognition of the broadcast-flag mark.

Broadly described, a broadcast-flag scheme for digital television works this way: A digital television broadcast transmits a sequence of discrete packets of data to its recipients. Each packet contains a part of video that is to be displayed, preceded by a brief “header” that conveys such information as where the packet fits into the overall sequence. The header may also contain a short digital sequence or “broadcast flag,” which labels the broadcast as copyrighted and which additionally conveys that the copyright owner grants only certain limited privileges to the broadcast’s recipients. Alternatively, a broadcast flag may not be present in every packet of digital TV content — if so, hardware must be redesigned to capture and “hold in detention” the packets of televised content until enough of packets can be examined to determine whether a broadcast flag is present.

In such a scheme the mark is not part of the content itself; instead the mark merely accompanies the content. This is both the strength and the weakness of this approach. It is a strength because it allows the mark to be found and interpreted easily, and because the mark can be applied to virtually any type of content. It is a weakness because anyone who receives the content outside of a secure personal computer or consumer-electronics device (or a home-entertainment environment constructed of such secure devices) can easily separate the mark from the content by editing it out.[20]

“Watermarking” Digital Content

Most people think of the term “watermark” in reference to paper products — hold a watermarked piece of paper up to the light, and you can see where the manufacturer has marked it, perhaps with his or her company logo. On paper, a watermark is not part of the content, but part of the medium (paper) on which the content has been placed.

In the digital world, however, a watermark is a subtle mark that is added to the digital content itself. For example, if the content is a recorded song, the watermark might be a faint sound that is added as background noise to the song. Digital watermarks are so named because they serve the same purpose as watermarks on paper — the idea is to embed a subtle mark deeply into the fabric of the content, without interfering too much with the content itself.

[20] Once one captures digital content, one can use a computer to remove, modify, or forge a label. While the content may be encrypted while in transition to display, it must ultimately be displayed “in the clear” (that is, decrypted) in order for an audience to use it. Obviously, cryptography alone cannot prevent the complete removal of a label of content that is displayed or otherwise made available in unencrypted form.

A successful watermark must have three characteristics. It must be:

- **Imperceptible to the user of the content**: Adding the watermark must not affect the user's experience in viewing the content.
- **Detectable by machines**: An authorized player or other digital tool must be able reliably to detect the watermark.
- **Difficult or impossible to remove**: It must be difficult or impossible for an unauthorized party to remove the watermark or to render it undetectable, except by unacceptably damaging the perceptual quality of the content.

Several companies offer products that claim to meet these requirements. Only a few of these products, however, have undergone independent scientific scrutiny, and those few that have undergone such scrutiny have not stood up well. As a technical matter no one knows for certain whether it is even possible to meet all three requirements simultaneously.Unlike a simple label, a successful watermark would be embedded in the content itself, and if the watermark were adequately persistent then nobody would be able to separate it from the content — or at least not without a great deal of trouble. [21]

"Fingerprinting" Digital Content

A fingerprint is a type of mark that is not added to the content, but is extracted from the preexisting characteristics of the content.[22] For example, if the content is a recorded song, then the fingerprint may be derived from the song's tempo, its rhythms, the length of its verses or movements, and mix of instruments used, and/or other features.

[21] The watermarks discussed in this paper are "robust" watermarks, which are designed to survive common operations such as data compression. Other types of watermarks are "fragile," meaning that they are deliberately designed so that they do not survive such operations. Fragile watermarks are useful only in conjunction with robust watermarks, and fragile watermarks are by definition easily removed from content, so their use cannot increase a system's resistance to hostile attack. The idea is that if a machine detector finds the presence of a robust watermark in the absence of a fragile watermark, it follows that the content has been inappropriately manipulated. But since the theory relies on the proposition that there is such a thing as a robust watermark, we need to focus only on whether such robust watermarks are possible.

[22] "Fingerprint" is a term of art that unfortunately has different meanings in different subareas of computer science (as well as other areas of science, of course). Here we use the meaning common in discussions of digital copy protection.

To be effective, a fingerprinting method must be:

- **Unique or At Least Precise**: Two pieces of content that look or sound different to a person should almost always have different fingerprints.
- **Difficult or Impossible to Remove**: It must be difficult or impossible for an unauthorized party to alter the content in a way that changes its fingerprint, except by unacceptably damaging the perceptual quality of the content.

To be successful, a fingerprinting method must meet both of these requirements. A number of companies offer fingerprinting technologies that claim to satisfy these requirements, but none of these claims has undergone independent scientific scrutiny. As a technical matter, it has not been independently established whether it is even possible to meet both requirements simultaneously.

Since the fingerprint is derived from the preexisting content, it cannot be used to store information about the content, such as an enumeration of authorized uses. (By way of analogy, your actual fingerprints may be unique to you, and may serve to identify you, while telling us nothing at all about your legal status.) Instead, the fingerprint acts as a unique identifier for each piece of content, and this identifier can be used to access an external database containing information about each piece of content. Assuming that such a database could be built, a fingerprint could serve roughly the same function as a watermark.[23]

A Deeper Understanding of the "Persistence" Requirement

To be persistent, a watermark or fingerprint must be able to survive any of the digital transformations that a would-be infringer might attempt to perform on the digital content. A wide range of such transformations exists. These include (but are not limited to):

- **Playing** the content, then using a recording device such as a microphone or a camera to recapture the played content,
- **Compressing** the content using a method such as MP3 that

[23] One possible use of watermarks that has been discussed is the use of "serial watermarks" as a form of what might be called "externally imposed fingerprinting." For example, when a user buys a recording from a website, the downloaded file could be uniquely marked with a subtle mark reflecting not the characteristics of the music but the identity of the buyer. This measure might enable content companies to track the source of an infringing digital copy of the recording found on the Internet. No scheme currently proposed entails the use of "serial watermarks," so I do not critique such a scheme here, although I note in passing that users may be wary of a scheme aimed at branding the content they experience with identifying information.

makes some modifications in the content in order to facilitate compression,

- **Adding** certain kinds of random noise to the content, and
- **Altering** the content by making subtle changes in the tempo, timing, pitch, or coloration of the content.

Many of these changes are often made for legitimate reasons, and there are many useful (and lawful) signal-processing and image-processing tools that allow an even broader range of possible transformations. Experts agree that devising a mark or label capable of surviving the full range of these transformations is much more difficult than a non-expert might initially expect.

How "Marking" Functions in a Copy Protection Scheme

By itself, no mark can function as a copy-protection scheme. Instead, a mark is a building block that is used in designing a copy-protection scheme. Though the details of such schemes differ, they share certain important characteristics.

First, marking schemes rely on widespread marking of copyrighted content, since they cannot hope to protect content that is not marked. If a "simple marking" or "watermarking" approach is being used, then of course there is no way to mark content that was distributed before the copy protection scheme was adopted.[24] Unmarked unauthorized copies of content could continue to be copied on the Internet and elsewhere, and could continue to be experienced and manipulated by users, so long as players and other devices that inspect content for marks, but do not find them, continue to be capable of playing or processing unmarked content. This is why some critics of marking-based schemes argue that the only way for marking-based schemes to work is if players and other devices read and play only marked content, and refuse to read or play unmarked content.

Second, marking schemes rely on all devices that read the content to check for the mark and, if the mark is found, to obey any corresponding restrictions on use of the content. Of course, devices that were sold before the copy-protection scheme was adopted will not be able to satisfy this requirement. This gives rise to what may be characterized as "the backward-compatibility problem," which may undermine attempts to implement industry wide copy protection schemes.

The Backward-Compatability Problem

When a new copy protection scheme is launched, it generally isn't implemented in pre-existing devices. For example, a new scheme for copy protecting recorded music may not be supported by the existing CD players already in use. This fact poses serious problems for the advocates of

[24] In practical terms, this means that all content is traded in unprotected, unmarked form on the Internet today may continue to be traded, absent the sort of regime described below.

copy protection. There are three ways for proponents or implementers of this scheme to deal with this backward compatibility problem, but all three have serious costs and other flaws.

The **first approach** is to ignore the problem. This makes the owners of existing devices happy, but the existing devices become a loophole in the system, a loophole that is widely available to would-be infringers. This approach is precisely what is asked for by proponents of the broadcast-flag approach to DRM for digital television broadcasts — existing digital television receivers will continue to function regardless of the presence of the broadcast-flag bit, which means they can be used to sidestep attempts to limit copying of television programs.

The **second approach** is to require all consumers to upgrade immediately to new players that support the new copy protection scheme. This seems likely to anger consumers, and understandably so, as they would be forced to throw away perfectly good equipment and replace it with expensive new equipment. It is hard to imagine such an approach being viable for established media such as television,[25] movies, or music.[26]

The **third approach** is to accept the existing-equipment loophole for existing content, but to release new content in a fashion that allows it to be played only on new players. This is essentially a slow-motion version of the preceding approach. Consumers would be forced to choose either to buy an expensive (and perhaps redundant) new player, or to forgo all new content. This approach too seems likely to provoke a high level of consumer anger and expense, albeit perhaps less than the backlash that might be triggered by an abrupt cut-off of existing home-entertainment equipment.

There seems to be no clearly unproblematic way to address the backward compatibility problem, except in cases where a truly new medium (rather than a new format or new distribution method for an existing medium) is

[25] Nevertheless, such an approach has been suggested by some advocates of copy protection in the context of the United States' planned transition to digital broadcast television, which is seen, incorrectly, as posing a unique threat to copyright holders.

[26] The consumer markets for movies and (especially) for music have of course endured a number of format changes. What made the transitions — e.g., from VHS movies to DVD movies, and from music LPs and cassettes to CDS — tolerable for consumers were that they were unforced; consumers got to choose when they would move to a new format, and could set up their home entertainment systems to accommodate multiple formats. This will not be the case during the transition to digital television in the United States because the U.S. government plans to reclaim the spectrum loaned to broadcasters during the digital transition so that broadcasters could transmit both digital and analog television. Once the transition is complete, analog broadcasts will be shut off, and consumers will have to pay to upgrade their equipment in order to receive broadcast television content.

being created. The advent of DVD movies was an instance of a new media format[27] that gave moviemakers and DVD player builders an opportunity to build in a type of DRM. But such helpful transitions to new media, new media formats, and new technologies are not predictable as a general rule.

Moreover, waiting for a new medium or media format is no help for those who hold copyrights in existing media such as music and movies, and who may be heavily invested in business models based on media formats such as CD audio recordings and DVD video. It has been argued that improvements in older media forms — *e.g.*, high-definition television — may be compelling enough to ease the consumer transition to new players and other devices, but it has yet to be seen whether these claims for HDTV will be borne out in the marketplace. Historically, changes or improvements that provide opportunities for new protection schemes have come about because of market demand rather than government mandate. In the HDTV context, however, the transition to digital television has largely been driven by the government.

A Closer Look at the Broadcast Flag Scheme

Because digital television, including HDTV, is commonly (if incorrectly[28]) perceived to be more easily copied and transmitted over networks like the Internet, there has been a push by content companies to protect over-the-air broadcasting with a marking scheme called, generally, "the broadcast flag" scheme or sometimes just "the broadcast flag." ("The broadcast flag" is actually the term for the mark that is used in the scheme.) A version of that scheme was adopted by the Federal Communications Commission in November 2003. (The FCC's regulation was later struck down by the DC Circuit Court of Appeals.) The goal of a broadcast flag scheme is to label the digital broadcast content, then somehow ensure that it cannot be captured at all, or that if it can be captured (e.g., by consumer personal video recorders like TiVo or Replay TV) it cannot be duplicated without limit or redistributed to the Internet.

[27] I distinguish here between media forms (such as music recordings and television) and media formats (such as music CD and cassette formats, or the analog and digital television formats). Media formats may change frequently, and at a faster rate as technological advance accelerates, but new media forms arise less frequently.

[28] This threat model for digital broadcast television has been significantly criticized by a number of participants in the Federal Communications Commission's proceedings regarding broadcast-flag regulation. It has been noted, for example, that data files of digital-television content are generally too large to be easily transmitted over the Internet, absent file compression that reduces the quality of such content. This remains true even if we assume that broadband Internet transmission capability will continue to grow significantly over the next decade. As has been pointed out earlier in this paper, analog television content is more vulnerable to Internet redistribution because, once digitized, its data-file sizes are significant smaller than those for HDTV program, and thus easier to transmit and receive. See also the discussion of "Analog versus Digital" in Appendix I.

The broadcast-flag scheme is a hybrid copy protection scheme that uses different methods to protect the content at different points in the distribution chain. In the first stage of distribution, the content is broadcast over the air-waves in digital form but is otherwise "in the clear" (unencrypted, unscrambled). In this stage, a broadcast flag is invisibly attached to each field or frame of digital television content. The broadcast flag, when present, denotes the copyright owner's statement that the recipient is not authorized to redistribute the content.

To state the flag scheme in somewhat oversimplified form, the proposal requires that when a device containing a "demodulator" (*e.g.*, a set-top box that receives a signal from an antenna or from a cable-TV feed) has received the content, the demodulator must check for the existence of the broadcast flag, and if the broadcast flag is present, the decoder may not pass on the content to another downstream device (such as a television, a video recorder, or a computer) unless the device containing the demodulator first re-encodes the content using some other copy-protection technology, or else passes it through "robust" (user-inaccessible) channels to another device that can be relied upon to do the re-encoding.[29]

If the decoder does re-protect the content, it must do so using one of the approved copy protection technologies that are listed in "Table A" of the broadcast-flag framework.[30]

The scheme additionally requires that if a downstream device is capable of understanding content that is encoded using one of the Table A technologies, then that downstream device must itself implement that Table A technology.

Given the relative simplicity of the broadcast flag itself, the mandates relating to other technologies are in fact the main effect of the broadcast-flag approach. In this sense, the broadcast-flag scheme can be considered primarily to be a meta-standard whose purpose is to mandate the use of other standards. Thus, to analyze the real effect of the broadcast-flag scheme we must consider the effect of mandating all of these other technologies. A full

[29] There is one exception to this rule: the demodulating device can only pass on content without checking for the broadcast flag, provided that it does not erase any broadcast flag that might be present, and provided that it passes the content through robust channels on only to devices which will themselves check the broadcast flag. Since ultimately the content is streamed or transmitted only to devices that check for the flag, we will discuss only the flag-checking component of the broadcast-flag scheme here.

[30] "Table A" is the name of the appendix to the broadcast-flag proposal that lists technologies acceptable for protecting flagged content. The Federal Communications Commission adopted the "Table A" approach, and a number of technologies were admitted to Table A, but with the striking down of the Commission's proposed regulatory framework by the DC Circuit Court of Appeals in 2005, the status of Table A is now unclear.

assessment of that effect is beyond the scope of this paper, but we may be sure that costs associated with a broadcast-flag mandate spread out far beyond the costs of building flag-detectors into TV-receiver hardware.

To understand why this is so, we need to focus on the two main lessons to be drawn from our examination of the broadcast-flag proposal. First, consider **the ever-expanding nature of broadcast-flag technology mandate**. We start with a simple broadcast-flag label on digital television broadcasts. To protect the effectiveness of this, we need a mandate on all demodulator-containing devices. But this is pointless unless we impose additional mandates on all of the devices that might be "downstream"[31] from the demodulator. Because there are so many types of downstream devices, **we must incorporate by reference a set of other copy-protection technologies**. What started out as a "simple" broadcast flag scheme ends up including a range of copy-protection technologies, and what started out applying only to digital television demodulators must, to have any hope of even being effective at all, up applying to virtually all digital video equipment, personal computers, and personal-computer software.[32]

This expanding-mandate phenomenon is to be expected with "marking" approaches generally. Any technology mandate covers a limited set of devices and situations, and the devices at the edge of this coverage tend to become loopholes through which the content can escape. The natural response is to widen the coverage area to address the loopholes - but this tends only to move the boundary rather than eliminating it.

Arguably, then, the only mandate that might claim to be truly effective is one that expands to reach the entire universe of digital devices—in effect, it requires a massive universal redesign of digital technologies that might be used to capture, copy, and redistribute content labeled by the "broadcast flag." Efforts to "cabin" the effect of the broadcast-flag scheme by limiting it to certain classes of digital devices (digital TV receivers, set-top boxes, and personal video recorders, for example) may limit the extent to which IT companies and others must comply with such mandates, but at the price of

[31] A "downstream" device is, essentially, any device that does not itself contain a television signal demodulator but that can connect to a device that either demodulates broadcast signals, or to a device that itself has received demodulated content. In short, a downstream device is any device that receives demodulated content as the result of a digital connection — the hard disk in a TiVo personal video recorder, for example.

[32] The fact that the scope of devices affected by a broadcast-flag scheme is indeterminate is a key reason that there has never been an inter-industry consensus favoring the scheme.

increasing the risk that the marking scheme will be sidestepped, either by current or future digital tools that aren't covered by the scheme.[33] This development would render a "cabined" mandate and related expenses a relatively useless and costly exercise.

Premature Deployment of a Watermark Scheme: The Risks

As discussed in Appendix II, watermarking remains an unresolved area of scientific research and debate, with many fundamental open problems. No completely satisfactory watermarking techniques have yet been developed for the audio, video, or text domains, nor is it certain that a sufficiently secure, robust and invisible marking technology could be developed in the foreseeable future. It would be very risky, at present, to deploy systems (or to base regulatory structures) that depend for their security or viability on the highly speculative assumption that a practical watermarking scheme will be able to be developed. Should an adequate watermarking technique be invented, however, it would likely play a role in several aspects of copy protection and enforcement.

At least two applications of digital watermarking technology relate to DRM. The first is **content labeling**, in which the content owner aims to identify protected material and specify permissible uses and copying restrictions. The second is **serialization**, which aims to mark material with a unique serial number or message that identifies the authorized end user and thereby provides evidence of the source of illegal copying.

Neither content labeling nor serialization is sufficient by itself to prevent illegal copying, however. Both approaches require that various parts of the content distribution process be "trusted" and secured against unauthorized access. It is theoretically possible that a practical system might be based on either, or both, approaches, but such a practical system would impose certain requirements on device-makers and other industries that enable the playback of digital content.

Systems based exclusively on content labeling require that all devices that use restricted material will read the label and refuse to act in a manner that is contrary to the restrictions encoded in the label. Furthermore, each system component must contain all the necessary keys to access the protected content. In addition, the required trusted system components include essentially any end-user equipment that must process labeled content. This last requirement is an ambitious one, since it would, for audio and video, include the entire range of consumer electronic devices, potentially including general-purpose computers.

[33] The FCC's broadcast-flag report and order in November 2003 takes just this approach — a "narrowed" order that leaves many devices untouched, and that also does not address alternative means by which digital TV content can be captured, such as the "analog hole."

Systems based exclusively on serialization, on the other hand, do not place any special requirements on end-user devices, since they would depend for their security on the fact that the perpetrators of illegal copying risk exposure by having their identity encoded in every copy they produce. However, such a system still entails considerable security infrastructure, with significant costs and risks. In a serialization scheme, the entire distribution chain, up to the point at which material is serialized, must be secure against unauthorized use; a single unauthorized, un-serialized copy has the potential to compromise the entire system. Furthermore, any effective serialization scheme requires that consumers be positively identified at the point of sale and associated with their serialized copy, a difficult administrative task at best, and one with serious privacy implications.[34]

Hybrid schemes may also be possible. For example, a system could employ both labeling and serialization (in which case both the distribution channel and the end-user devices must be secured). It might also be possible to perform part of a user-serialization process in the end-user devices, although that would still require a trusted distribution channel as well as trusted end-user hardware. The recent developments in "trusted computing" championed by Intel and by Microsoft may facilitate such a trusted distribution channel. But since "trusted computing" depends primarily on encryption and on the creation of secure environments ***within*** computing platforms, a marking scheme may be superfluous in a true trusted-computing environment.

Outside the trusted-computing context, systems based on watermarking, whether for labeling or for serialization, are often quite vulnerable to single points of failure. In particular, currently proposed "watermarking" systems all have the property that anyone with enough information to read a watermark can easily derive the information needed to remove it. In the case of labeling systems, this means that if any user device is compromised and the watermarking parameters discovered, not only can that user device make unlimited copies, but also labels can be removed or altered from content to be played on unmodified devices. In the case of serialization schemes, this means that if a single user is prosecuted in open court, the very same evidence that identifies and convicts him will provide a primer for future illicit copiers to escape detection.

[34] It may be argued that consumers already know that they are routinely surrendering private information in other contexts, such as credit-card purchases, video-on-demand orders, and so forth, so that their surrender of private information in the content-purchase context is at most incremental and unlikely to trouble consumers. Nevertheless, it seems likely that at least some consumers will be troubled by the transition from (a) a world in which content such as books, music albums, and DVDs can be purchased anonymously with cash, enjoyed, and then resold into (b) a world in which the content a consumer purchases is "tethered" to that individual, *known* to be associated with that individual, and *can't* be resold.

Other Approaches

Faced with the difficulties associated with each of the major types of copy protection discussed above, content owners have begun to explore other options. Among them are ***selective incompatibility*** and ***DRM hybrids***.

Selective incompability is an approach we've already seen in the marketplace for music CDs. Here the notion has been that a music CD manufacturer will add deliberate "errors" into encoding of music content on CDs, with the result that the CDs will be readable by some CD players (typically consumer-electronics single-purpose devices) and not by others (typically computer CD drives). Music companies' initial efforts in this direction suggest that this approach is not a particularly viable one (one protection scheme for CDs could be defeated by using a felt-tip marker to cover up the "errors" around the edge of the CD[35]), and in the long term the risk of too many "false positives" (CDs that are judged to be illicit copies by protected players) and "false negatives" (CDs that are judged to be unprotected when in fact the manufacturer meant for it to be judged as protected) is significant with the selective-incompatibility approach. Moreover, both device makers and consumers are likely to react negatively to CDs that do not reliably play on the platforms that consumers customarily use. (This negative reaction can no doubt be diminished by clearly marking such CDs as protected in this manner, but this may also result in diminished sales, at least in some markets.) The issue of selective incompatibility may also arise in the near-term with the deployment of DVD-movie products in "higher-definition" or other "higher-quality" formats that cannot be read by existing DVD players.

An example of a **DRM Hybrid** includes some variants of the broadcast-flag approach for protecting television content, combined with some use of encryption. Under the "hybrid" version of the scheme, the first step is, as with the broadcast-flag scheme generally, to insert the "broadcast flag" into the digital-television signal. If that signal is not itself encrypted, there are no technical barriers to removal of this flag.

For this reason, the DRM Hybrid version of the broadcast-flag scheme would require a legal or regulatory mandate that receivers check for the broadcast flag, and apply an approved DRM method (such as encryption or "tethering" the received content to a particular home entertainment system or user) to the content if it is marked with the flag.

[35] The federal Digital Millennium Copyright Act has been interpreted in at least one case to forbid even the dissemination of information that can be used to defeat a DRM scheme. See Universal City Studios, Inc. v. Reimerdes, 111 F. Supp. 2d 294 (S.D.N.Y. 2000). In theory, then, this discussion of using felt-tip markers to defeat CD copy protection might be deemed illegal. A discussion of the DMCA follows below.

The DRM Hybrid version of the broadcast flag approach illustrates a problem with marking systems generally — whether one is using a "pure" marking scheme or a DRM hybrid version of the scheme, there is a general requirement that the scheme be buttressed by government regulation of some sort. This brings us to our next policy question.

DRM: What Role Should the Government Play?

No treatment of the DRM-mandate issue (which hereafter we'll refer to as "technology mandates" or simply "mandates") can proceed without recognizing that a confluence of several factors has brought copyright issues to the center of the public-policy arena. Although not quite commanding the same attention as, say, counterterrorism measures or foreign-policy concerns, copyright issues have risen to the top of the discussion of broadband-Internet policy and digital-television policy in addition to commanding attention on their own.

The reason for this increasingly evident intersection of copyright policy and technology policy lies in the fundamental nature of the personal computer itself. Computers are designed to copy and manipulate data with ease and with accuracy, (*e.g.*, from hard disk to RAM, or from your e-mail program to a friend's). This copying is part of the essential functionality of computers. For this reason, the increasing ubiquity of powerful but inexpensive general-purpose computers poses a particular challenge for copyright holders whose interests lie in digital works.

The Lessons of Software Copy Protection

For some copyright holders, this challenge has long been apparent. Software makers in the 1970s and 1980s were fully aware that computers could be used to make unlicensed copies of their products. Many and perhaps most software makers attempted to use various "copy protection" technologies to prevent users from making copies of commercial software; these efforts were less than completely successful in large part because general-purpose computers could be programmed to edit or alter the very copy-protection measure that was designed to prevent copying. As a result of this aspect of the nature of computers, and in response to the inconvenience of copy-protection schemes of the period, an aftermarket in utility programs that enabled the defeat of such measures quickly appeared. For several years there was an ongoing "arms race" between commercial software vendors, who developed ever more powerful and arcane copy-protection strategies, and the makers of copy-protection-defeating utilities. Generally speaking, however, digital information is inherently copyable and inherently alterable, which poses special problems for the copyright holder who seeks to prevent his or her digital work from being copied or altered.

In the same period, the general-purpose nature of personal computers was what created the greatest number of business opportunities for software makers. Because the leading personal computers at the beginning of the microcomputer revolution — most notably the Apple II (1978) and the IBM PC (1981) — were designed to be "open platforms," this meant that third-party software vendors were able to freely develop new applications for those platforms. (One of those vendors, Microsoft, has been so successful at this that its revenues currently exceed those of any single computer maker.) At the same time, efforts to meet consumer needs by promoting dedicated word processors and other dedicated digital tools failed in the marketplace, with the notable exception of game consoles. In general, when it comes to computing, the public prefers general-purpose, unconstrained, "open-platform" tools to special-purpose, limited ones. We may reasonably infer, therefore, that any government mandate that focuses on limiting the functionality of general-purpose computing tools may have the unintended effect of diminishing the market for the resulting products. A mandate might also prime an aftermarket for pre-mandate "open architecture" devices, perhaps via auction websites such as eBay.

How Digital Content is Different From Software

Software makers in the 1970s and 1980s had certain advantages when attempting to block software copyright infringement. The first and most important advantage was the fact that most commercial software that is capable of performing complex tasks requires a significant degree of documentation and support; bona-fide purchasers of software were able to receive such benefits, whereas those who made unlicensed copies typically had to do without. This led to another aftermarket, this one in third-party manuals and workbooks for commercial software products. That aftermarket continues to this day, somewhat to the chagrin of software vendors. Nevertheless, software makers have generally abandoned the harsher varieties of copy-protection schemes to prevent unauthorized software copying, largely because of negative consumer response. A fundamental principle of the personal-computer business — the principle of High Volume/Low Cost — drove vendors away from the more rigid and restrictive types of copy protection; the vendors manage the problem of "leakage" partly by seeking legal remedies against the more egregious infringers, partly by making software more affordable and thus easier to acquire legally, and partly by employing measures such as registration schemes that make illicit copying somewhat more difficult for ordinary users.

But the other advantages for software makers were also considerable: modem speeds were comparatively slow, computer storage was comparatively expensive, and relatively few consumers had access to the Internet. All of these advantages evaporated in the late 1980s, in the 1990s, as well as in the current era, as modem speeds increased by orders of magnitude, computer storage became increasingly cheap, and access to the Internet became ubiquitous. Indeed, current telephone modems, fast as they are, are likely to be supplanted by high-speed broadband connections to the Internet, which are becoming increasingly available to most businesses and homes in America.

These factors compounded the problems of digital copyright holders in a number of ways. First of all, just as the "open platform" of the general purpose computer was designed to make the copying of digital information easy and reliable, the "open platform" of the Internet was designed to make the copying of bits over long distances reliable. In addition, the Internet makes it possible for individual computer users to copy works to a multiplicity of recipients—to effectively "broadcast" unlicensed copies of copyrighted works.

Secondly, greater computer capacity, advances in compression technology, and greater bandwidth have made it possible to copy significant numbers of copyrighted works (most notably, songs in the form of MP3 files) and copyrighted works of significant size (*e.g*., television shows and feature films). Third, digital copyrighted content, unlike copyrighted software, does not typically require documentation or support to be used.

Content Producers Respond: The DMCA

As we have seen, in the computer era one approach of content owners has been to enclose or protect digital copyrighted works with what we once called "copy-protection technologies" but which we are now commonly referred to as Digital Rights Management technologies or "DRM." These technologies are commonly (but not always) based on encryption. The use of encryption-based copy-protection technologies has been the foremost of the two major approaches to preventing the unauthorized copying of copyrighted works in the digital world. The other approach — the reliance on government regulations (often called "technology mandates") to constrain the capabilities of consumer technologies — has been much less widespread, although this may be changing.

The Digital Millennium Copyright Act (DMCA), enacted in 1998, reflects the Content companies' assumption at the time that it would rely primarily on DRM technologies to protect its copyrighted works in the digital age. Most of the broad prohibitions of the DMCA are aimed at preventing circumvention of these technologies, which, once again, are usually based on encryption. At the same time, the DMCA expressly codified a technological mandate concerning videocassette recorders, but also expressly sidestepped the issue of technology mandates generally.[36]

The DMCA's provisions in 1998 were aimed at making DRM approaches more secure by broadly prohibiting circumvention (thus preventing the kind of "arms race" between copy-protection developers and copy-protection circumventers that occurred in the 1980s in the software industry). At the time, content-company advocates, as well as advocates from the information-technology sectors, believed that the DMCA was in itself a long-term solution to the copyright-infringement problems they feared. The DMCA, which did not provide significant exceptions for circumvention or circumvention tools — even when the underlying goal of the circumventer or the toolmaker was a legal one — was enough, they believed to forestall the kind of widespread infringement and unlicensed copying to which the Internet might give rise.

[36] See 17 U.S.C., Sec. 1201(c)(3).

The Problem of Peer-to-Peer Content Sharing

Only a year later, however, peer-to-peer file sharing became a visible and well-publicized phenomenon (most famously, through the use of Napster). While particular peer-to-peer applications and services may be hindered or neutralized through litigation, the essentially decentralized nature of the Internet, together with the ubiquity and increasing cheapness of computers, has made it possible for file-sharing, particularly of music, to continue.

Music companies have been particularly vulnerable to file sharing, for a number of reasons. First and foremost, MP3 compression makes most song-files remarkably compact, which makes them far easier to transmit and receive, even over today's relatively limited consumer broadband networks. Moreover, most music companies' catalogs are available on CD in unprotected formats. This is because music companies, when adopting the CD format for distribution of music, did not anticipate the ubiquity and ease of use of "ripping"[37] and "burning"[38] applications and technologies, nor did they anticipate the quickness with which music hobbyists would begin to share "ripped" music files on the Internet via peer-to-peer mechanisms.

Movie and television studios and networks fear that increasing adoption of broadband Internet services, together with the migration of video content to digital formats (such as HDTV) will result in the same "Napsterization" of their offerings that the music companies have endured. There is less justification for their fears, at least in the short term. Partly this is because the most common form of commercial video distribution, other than television, is DVD sales, and DVD content is scrambled to prevent easy copying. Another factor that has slowed or even prevented true "Napsterization" of television content has been the sheer size of video files; in general, a digital file created from an hour of standard television is two or more orders of magnitude larger than an MP3 file. (When the digital file is HDTV television, the disparity in file sizes is even larger.)

Nevertheless, TV and movie offerings that are distributed via broadcast and cable channels are either unprotected ("in the clear") or, if protected by DRM, they are descrambled at the player/receiver end (which they must be in order to be viewed), whereupon they can be captured by users through a variety of means. These users can then digitize, alter, or disseminate the works to the Internet and elsewhere. The fact that some users do this (albeit after reducing significantly the resolution and quality of the captured video

[37] "Ripping" is the reduction of music from its native digital format on CD to reduced-in-size formats such as MP3. The term may also be applied to the reduction to MP3 of music in analog formats, such as the tracks on LP records.

[38] "Burning" is the reproduction of music in compressed formats, such as MP3, to a writable CD or some other writable optical medium. "Burning" of music files may or may not include re-expansion of the files into a non-compressed format.

content) is adduced by movie and TV companies as evidence that the threat to their business models posed by peer-to-peer file sharing is either just around the corner, or is already here.[39]

Neither the issue of peer-to-peer file sharing nor the increasing capability of computers to copy and transmit content was fully foreseen by the drafters of the DMCA. Specifically, these issues are not addressed by the anti-circumvention provisions of the DMCA[40] (since they rarely if ever involve actual circumvention of copy-protection technology) and only tangentially addressed by the Notice-and-Takedown provisions of the DMCA (since peer-to-peer file sharing does not typically require the use of an Internet Service Provider as a site for making illicit content available). Of course, to the extent that file sharing constitutes copyright infringement, it is squarely addressed by the substantive provisions of the Copyright Act, Title 17 of the U.S. Code, but the prospect of having to file thousands or even millions of infringement actions against American citizens is a daunting one for even the most assiduously protective content company.[41]

As a result of the onset of peer-to-peer file sharing, a number of content companies have increasingly asserted that the solution to peer-to-peer infringement lies in constraining what those tools and mechanisms can do,

[39] Ironically enough, the unauthorized copying of current television shows invariably originates as digitized copies of ***analog*** transmissions. Actual HDTV content, which is digital, cannot be significantly compressed without loss of the very video quality that makes it special, which makes digital TV content safer from this kind of infringement than is ordinary analog television. Nevertheless, it is a widely accepted myth that digital television, merely by virtue of being digital, is more subject to peer-to-peer infringement than is analog TV content. At some level, however, content companies see past this myth; hence their efforts to develop marking schemes that survive digital-to-analog/analog-to-digital conversions, or, in the alternative, to pressure device makers to "retire" analog technologies altogether.

[40] The DMCA's notice-and-takedown provisions create a liability "safe harbor" for Internet service providers and others who, when given notice by a copyright holder that their service or site contains infringing copies of copyrighted works, immediately take down the content in question. The statute also includes an appeals process for those who believe the works on their service or site are not instances of copyright infringement.

[41] Of course, it is well documented that the recording industry has started to bring infringement actions against parties it considers to be the most egregious file-trading infringers of music copyrights. Whether this campaign has successfully deterred illegal file trading is unclear. A recent Pew Research Center poll <http://www.pewinternet.org/reports/toc.asp?Report=109> concluded that such deterrence may have occurred, but critics of the poll's methodology have argued that the data are ambiguous at best, since accurate data would depend on respondents' admitting to illegal activity. See, *e.g*., Schwartz, John, "In Survey, Fewer Are Sharing Files (Or Admitting It)," *The New York Times*, Jan. 5, 2004. Section C, Page 1.

through technological measures in tandem with government mandates (either regulatory or legislative or both). This perception has been at the root of recent legislative and regulatory initiatives that are designed to limit or prevent peer-to-peer file sharing, but that may, if enacted, have unintended consequences, including a significant chilling effect on innovation. That such measures may be worse than ineffective — that they may even be counterproductive — becomes apparent when we dig deeper into our analysis of what the real "threat" to content companies is.

But before we discuss this threat model, we need to consider one final aspect of the issue of government mandates of DRM — that, even considered purely in terms of content-rights-holder interests, a government-imposed technology mandate might actually exacerbate rather than diminish infringement problems. Consider that when it comes to "solving" infringement problems DRM is aimed at a moving target. As circumvention tools evolve, and as new technologies pose new infringement problems, the locking of industrial sectors into a particular "standard" scheme, mediated and supervised by government, actually ***slows the ability of the Content sector to respond to new problems***. There are fewer incentives to develop solutions that lie outside the standard. And solutions that are developed within the framework of an "open" yet government-administered standard will take longer to be approved and longer to find their way into market offerings. While it is unclear that truly effective long-term DRM solutions can be developed in the digital environment, certainly they are less likely to be developed if development is constrained both by an increasingly outdated standard and by a government-approval process.

The Threat of Universal Infringement, and the Threat of Ill-Considered DRM

If there is anything that those of us who live in the age of the Internet, and who have access to the Internet, know for certain, it's that our use of computers and Internet enables us to engage in the broad sharing of any information. Since any content can, in principle, be broadly shared or disseminated over the Internet, it follows that the illegal distribution of copyrighted content is possible as well. In effect, the same aspects of computers and the Internet that empower us to be global publishers also empower us to be global copyright infringers.

In a nutshell: ***content owners fear that once an unprotected copy of a copyrighted digital work becomes available, it can and will be distributed universally on the Internet, and its distribution will destroy, or at least severely diminish, its ability to generate revenue***.

This fear often results in self-contradictory statements from content companies that seek, in various forums, legal or regulatory mandates for copy protection. On the one hand, such proposals are defended as "speed bumps" that merely "keep honest people honest" and that are not meant to be unduly burdensome to ordinary users of the content. On the other, when objections to certain kinds of mandates are raised, the advocates of the mandate frequently invoke the specter of the "one perfect copy" of the content escaping the secure system and then being distributed universally on the Internet. Policy discussions of DRM frequently oscillate between the advocacy of limited (and therefore ineffective) proposals and broad (and therefore less politically palatable) proposals. Sometimes the very same proposal may be described at one point as "limited" and at another point as "necessary to prevent Internet distribution." As a practical matter, no "limited" proposal can prevent Internet distribution of the copyrighted work.

Of course, not all unauthorized distribution of copyrighted content over the Internet is necessarily infringement. For example, we have built into our copyright-law framework the important principle of "fair use." Without discussing "fair use" and other exceptions to copyright protection in detail — a project that all by itself would take an essay far longer than this one — we can say generally that the Framers of the Constitution and subsequent interpreters of the Constitution's Copyright Clause and of the Copyright Act that springs from that clause believed that some degree of unlicensed or unauthorized use of another person's content is lawful. To the extent that such distribution is large scale, however, and to the extent that this large-scale distribution undermines the commercial value of the work being distributed, it is more likely to be found by a court to be infringement.

Theoretically, this notion of lawful though unauthorized use of copyrighted works is well-established and uncontroversial, and even the content industries can be heard to say they agree with the general principle that fair use is important. In practice, however, there is wide disagreement among stakeholders as to what the contours of "fair use" or other lawful but unauthorized uses might be. In the digital world, however, it is theoretically possible for content owners to use DRM to foreclose most or all unauthorized uses of even parts of copyrighted works. This foreclosure of uses does not merely affect our ability to make unauthorized copies; it also affects our ability to ***own*** a copy of a copyrighted work, since the rights associated with ownership of a traditional book or record album or movie may be significantly reduced in a DRM-mediated environment. This raises the question of whether the very existence and use of DRM alters the balances of our system of copyright in one direction, even as computers and peer-to-peer file sharing may alter them in another direction.[42] If so, then we may face the

[42] Increasingly, there are efforts to design what is called "fair use" into DRM schemes — here the term "fair use" does not carry the same meaning it carries in the Copyright Act. Instead, it signifies some degree of individual copying, in line with what is currently considered to be fair use under American copyright law. Legally speaking, such design efforts cannot be said to add up to "fair use," since in effect they allow instances of authorized copying — authorized, in this case, by the copyright holder and the designers of the DRM scheme — rather than the kind of unauthorized copying that is dealt with in Section 107 of the Copyright Act.

challenge of developing ways to ensure that DRM does not skew the fundamental structure of rights in copyright. The last part of this paper addresses some choices we may make in that regard.

Can "Peer-to-Peer" be Stopped on the Internet Itself?

Foremost among the perceived threats faced by the copyright industries in the digital age is peer-to-peer file sharing, which arrived as a widespread mass consumer phenomenon. Although peer-to-peer file sharing is commonly regarded in the content community as a new, and pernicious, technological development, it actually derives from the "architectural" design of the Internet itself. The Internet was designed to be a simple, robust, reliable, and (most important) decentralized computer-based communications medium. Because the "peer-to-peer" aspect of the Internet is central to its design, it is difficult to imagine any "solution" to peer-to-peer-based copyright infringement that does not require, at minimum, a fundamental redesign of the Internet.

At bottom, peer-to-peer file sharing can be understood simply as the use of multiple computers connected to the Internet as both "servers" (storing specified files that other computers can retrieve) and as "clients" (able to retrieve specified files from other computers that are storing them). Because computers engaging in such reciprocal file sharing and retrieval are acting both as "servers" and as "clients," they are, in effect, "peers"—hence the term "peer-to-peer." It is generally believed that peer-to-peer file-sharing has greatly increased the volume of unlicensed copying, although reliable statistics as to the actual extent of such copying or as to the economic impact of such copying are currently unavailable, partly due to the problem with tracking such copying. Moreover, although peer-to-peer applications are perceived by many to be primarily tools of copyright infringement, it is important to stress that such applications have both infringing and non-infringing uses.

What this all adds up to is that the aspect of the Internet that most bothers content companies is an aspect that is central to its design. In effect, it is exceedingly difficult to craft a law or regulation that categorically outlaws peer-to-peer file sharing without, in doing so, outlawing the Internet itself.

This may be counterintuitive to you if you think of peer-to-peer file sharing as a relatively recent phenomenon. One of the reasons peer-to-peer file-sharing may seem to be a new phenomenon is that, during the explosion of commercial activity on the Internet in the 1990s, it was common to have larger, more powerful computers function primarily as servers, which then could be accessed by personal computers and other devices that would retrieve files as necessary. The term "web server" in the mid-1990s typically denoted, or at least suggested, the use of a larger, more powerful machine to "serve" web content to users who were surfing the Web with their personal computers.

As a factual matter, however, any computer capable of connecting to the Internet in a manner that relies on the standard Internet Protocol (sometimes referred to as "IP" or even redundantly as "the IP protocol") can potentially act either as a server, or as a client, or as both. (Increasingly, consumer operating systems, from Windows to the Mac OS to GNU/Linux, include software designed to enable the use of the computer running the operating system as a Web server.)

To understand how intertwined peer-to-peer file sharing is with the Internet itself, it helps to consider what the Internet actually is. One way to understand the Internet is to say that the Internet consists of special computers called "routers" interconnected with fiber optic lines, cable and dialup modems, and the like. General-purpose "host" computers, including your personal computer, connect to the routers to use the Internet to communicate, and in effect they become part of the Internet as well.

The Internet's End-to-End Design Principle

Routers are functionally similar to postal sorting machines. But instead of sorting paper envelopes, they handle small electronic "packets" of data from and to the host computers. By design, routers provide only a minimum set of services. All other processing is done on an "end-to-end" basis by the hosts. This "end-to-end principle" was and continues to be extremely influential in the development of the Internet. The principle serves two vital purposes: it simplifies and reduces the cost of the Internet infrastructure, and it facilitates the development of new and innovative Internet applications on the host computers.

This end-to-end architecture has a profound effect on the viability of any mandated technological copy-protection scheme implemented within the Internet itself (i.e., in the routers), as opposed to the host computers connected to it. Just as a postal sorting machine looks only at the addresses on the outsides of envelopes without opening them, Internet routers only need look at the Internet protocol "header" on each packet. The content of each packet is wholly arbitrary, and is not necessarily meaningful to anyone but the host to which it is addressed. This is especially true when the packet has been encrypted with a key possessed by only the source and destination hosts. Several security (encryption) protocols are already widely used on the Internet, including SSL, SSH, TLS and IPSEC.

Since end-to-end encryption can completely hide the meaning of each packet, use of encryption would make it completely impossible for Internet routers to scan encrypted packets for "broadcast flags" or any other copyright information so that the transfer of such packets could be blocked.
Making things worse for would-be infringement detectives, even when Internet Protocol (IP) communications traffic is not encrypted, IP "sniffing" (sampling of packets) is not particularly effective at detecting infringement. A single IP packet can carry only a limited amount of data (1500 characters is the usual maximum) and it is both permitted and fairly common for the

different IP packets that make up a transfer to follow different paths to their destination. To determine by "sniffing" packets whether content is being infringed would require the gathering, buffering, and examination of whole files, or least large chunks of them, even if they were transmitted "in the clear."

What this discussion suggests is that any technology mandate that requires core Internet components (routers, transmission links, and so on) to implement schemes to thwart the transfer of copyrighted material is likely unworkable and easily circumventable. This means that copy-protection mandates, if they were to work at all, would require mechanisms and measures to be implemented in the Internet hosts.

Who's a Host on Today's Internet?

This is a broader prescription than it may first appear to be because nowadays virtually every computer on the Internet can function as a host.

This was not always the case, however. In the years before the personal computer revolution (which can be said to have begun approximately in 1976), Internet hosts were usually physically large, continuously operating computers that acted as both servers and clients, depending on how you used them. They communicated as equal peers.

With the personal computer driving the Internet's explosive growth in the mid to late 1990s, clients and servers began to differentiate. Today, many and perhaps most personal computers function only as clients; they rely on services provided by relatively few dedicated "server" hosts. Many Internet service providers (ISPs) provide server-type services to their customers so that they may publish web pages and other information. One consequence of this design is good news for content owners: Specifically, when a user publishes infringing material on his ISP's server, it is relatively easy to identify and contact the ISP staff to request removal of the infringing materials.

But this aspect of Internet use is changing, and in a way that harkens back to the original structure of the Internet. With the availability of high speed, "always-on" Internet access by cable modem and DSL, of advanced operating systems such as GNU/Linux and BSD and later versions of Windows, and of continuing price/performance improvements in computer processors, memory and disk storage, individuals can now run their own servers, accessible to anyone on the Internet. Users no longer necessarily rely on ISP-provided servers; they need only basic Internet connectivity. Individual users become, in effect, "hosts." This is why, while some perceive user-run peer-to-peer servers as a novel development, they are actually nothing more than a return to the Internet's original model as a network of computers as equal peers, each acting as both client and server.

The implication of this development is both clear and disturbing for those who wish to outlaw peer-to-peer services as such — one probably cannot build effective DRM at the router level[43], and building it in at the host level probably requires building DRM into every personal computer that can connect to the Internet.

This is a tall order, but at least some representatives of content companies hope to approach this goal. One way to do so is on a step-by-step basis. For example, content companies could seek regulations and other measures that affect the design of computers that receive television content (as an increasing number of personal computers are able to do), or that affect the design of devices that can be connected to TV receivers (this would cover a broad range of digital and consumer-electronics devices). Indeed, many observers regard the content companies' push for a broadcast-flag regulation to be evidence of such a step-by-step strategy.

The step-by-step approach can be used in more than one arena. For example, content companies can seek DRM-based design changes through private contracts (*e.g.*, by refusing to license content to cable or satellite-TV companies that don't incorporate certain DRM measures into their equipment, including the equipment they license consumers to use). At they same time they can seek to advance the ubiquity of DRM by public regulation such as the broadcast-flag proposal, which the FCC adopted, although the Commission's proposed regulation was adopted in significantly altered form, and even so was ultimately struck down in court.

The content companies' efforts to make DRM more pervasive in the digital world are complemented by efforts in the computer industry, some members of which hope to establish through "trusted computing" and similar initiatives a kind of DRM-based secure space inside your next computer — secure in ways that benefit you, perhaps, but also secure in ways that prevent you from having full control over your computer, especially when it is being used as a channel for delivery of commercial content.

Collectively, these efforts may hurt citizens in at least three ways:

First, they may swing the balance of rights in copyright so much further in the direction of the copyright owners that, in effect, they make the "fair use" and other balancing provisions of the Copyright Act unusable and thus irrelevant in practical terms.

[43] Nevertheless, at least one router company has offered to build DRM in at the router level. Outside experts remain skeptical that such a scheme can be implemented credibly, however, and they also note that, in order to work, such a scheme would require the replacement of most or all Internet routers currently being used. This of course would be a boon for router manufacturers, albeit a cost to nearly everyone else.

Secondly, to the extent that these efforts result in new limitations on personal computers and consumer electronics, citizens may soon find themselves in a world in which these tools empower them much less than they once did.

A **third**, related point is this: the computer revolution and the remarkable advances we've seen in computer technology over the last quarter century have been dependent largely on so-called "open architectures." Personal computers are said to have "open architectures" because you can buy or build new devices that the computers can use in new ways, and because you can program them to do things that their designers never thought of. Moreover, the Internet itself, through its "end-to-end" principle, is another example of an open architecture — because its underlying principles are decentralized, simple, and robust, it's possible for inventors to come up with new uses for it. A notable example of the latter is the World Wide Web itself, which originated two decades after the Internet was invented.

In sum, then, it's not just copyright-law interests that are at stake — or even citizens' relationship to copyrighted works. DRM, if too broadly and indiscriminately applied, may throttle the advance of personal-computer technology itself. What this means is that, in addition to the problems that DRM may create for libraries and librarians in limiting the use of content, it also might limit the creation and use of more refined and advanced information-retrieval tools.

DRM and Public Policy: Where Do We Go From Here?

As this essay demonstrates, the balance of rights and public policies we have grown accustomed to in our copyright-law framework is being pulled in more than one direction. On one side, digital technologies seem to have the potential to undermine and perhaps even destroy the incentive system we have constructed as part of society's "copyright bargain" with artists and authors. On the other side, DRM may have the potential to destroy society's part of that bargain, by enabling copyright owners to prevent even those unauthorized uses of copyrighted works that we recognize to be lawful, all in the name of stopping Internet-based infringement.

On still another side, we have the technology companies, who are torn between their desire to provide new platforms for copyrighted works (and who themselves value copyright) and their desire to prevent the open platforms of digital tools and the Internet from becoming "closed" in a way that hinders or halts innovations we haven't thought of yet.

No one can yet claim convincingly to have an obvious solution to the tensions created by the collision of DRM, copyright law, and the informational needs of an open, democratic society. Still, this should not prevent us from beginning to talk about what a balanced approach to DRM and related issues might look like, assuming for the sake of discussion that such a form of DRM is possible. [44]

The kind of DRM framework I imagine for the sake of this discussion meets, at least, the following set of criteria:

A. For Copyright Rights-Holders: It must limit (or, ideally, prevent) large-scale unauthorized redistribution of copyrighted works over the Internet or any similar medium. In addition, it must allow a range of business models for distributing content, within the constraints of copyright law.

B. For Technology Makers: It must maintain technology companies' ability to create a wide range of innovative non-infringing products, and to design, build, and maintain those products efficiently. It must maintain the ability to choose between open-source and closed-source development models. It must enable technology makers to come up with robust, interoperable, relatively simple technologies that are fault-tolerant and easy to maintain.

C. For Citizens, Ordinary Users, and the Communities Libraries Serve: It must maintain access to a wide variety of creative works, both past and present, including both public-domain works and works still protected by copyright. It must maintain access to advancing consumer technology for uses not related to copyright. It must continue to allow for maintain fair use and other lawful unlicensed uses of copyrighted works (including time-shifting, space-shifting, archiving, format translation, excerpting, and so on) and also must be flexible enough to allow for new, innovative fair uses (*e.g.*, uses of home networking and other kinds of beneficial uses we haven't yet imagined or discovered).

[44] Some readers may ask at this point whether it is appropriate to allow DRM to exist at all, given the remedies that copyright owners already possess under our copyright law. I understand and sympathize with their point. I also note, however, that our legal system allows us to take steps in other areas to prevent harm from coming to ourselves and to our interests. For example, we have the right to physically defend ourselves from assault, even when such defenses might themselves be considered criminal if unprovoked, and we have the right to lock up our tangible goods in our houses, even though we also have legal redress should we be stolen from or burglarized. Neither of these examples should be taken as analogous to the copyright bargain, whose built-in balances are unique, implicate free-speech considerations, and do not easily map to other areas of law. Copyright, as the Supreme Court has said, is "no ordinary chattel." But I think most citizens' intuitions about the copyright-law balance is that it should include at least some measure of self-help. Assuming those intuitions are correct, I infer that at least some degree of DRM may be considered acceptable, so long as it is implemented in a manner consistent with longstanding copyright policy and with the First Amendment.

It will be difficult for any policy to meet all three of the above criteria in a way that fully satisfies all stakeholders. A truly diverse free market in DRM-protected works might get us at least closer to these combined goals, but such a market will occur only if publishers commit to experiment with all variations of DRM implementations -- including experimenting with little or no copy protection at all. Currently, however, publishers tend to favor DRM-encumbered digital-media formats that are as limited as, or even more limited than, their analog predecessors.

Take e-books, for example. When reading an e-book, one often is blocked by DRM from cutting and pasting a passage from an e-book into a word-processing document -- this limitation forces students and scholars to retype literary passages that they're analyzing, and the e-book format also may restrict the kinds of scholarly analysis that could be done by doing computerized searches of the text. The practical outcome of the restrictions on proprietary e-book formats is that these formats are dead in the marketplace — e-books are unpopular, considered clunky and burdensome, and, at best, an idea whose time has not yet come. Designers of e-book platforms have been cogitating about the "right" combination of content protection and flexibility, but few publishers nowadays are considering less DRM-intensive options. If, however, some enterprising company set out to make some e-book editions of works available with most or all the DRM copy protection turned off, the market might learn whether the money to be made by publishing works in more convenient digital forms compensated for the (presumed) risk of widespread digital infringement. No such experiment appears to be on the horizon, however.

If the market is unlikely to experiment with less DRM-encumbered formats for digital media, does this mean there's a role for government in setting DRM standards and policies? Perhaps, but we must remain aware that, if history is any guide, the copyright industries will field countless lobbyists and spend significant amounts of money to promote their interests with policymakers.

We can compensate for that factor in part by educating consumers that it's not that e-book or digital-media formats are inherently limited — instead, it's that the limitations have been insisted upon by particular publishers or artists. This additional information enables consumers and libraries to make better-informed choices — they might choose one work over another because its DRM-enabled player has been set to be more flexible or less restrictive. They might forgo buying books from a particular publisher if that publisher insisted on too many restrictions. A more educated market for digital works is likely over time to become a more rational market, making better-informed choices about what kinds of access to a work they are willing to pay for.

Most of us believe that artists and authors deserve to be compensated, and even that publishers deserve compensation for bringing them to us. At the same time, it is a natural human impulse to share the creative works we love. In the absence of a more humane variety of DRM, these interests may be at odds with one another. The temptation will be to ask government to set DRM standards, but that process, as we see in the case of the broadcast-flag regulation, may lead to results that are potentially worse -- because they lock in restrictive DRM schemes -- than if the government did nothing. A better path to more humane DRM, a path along which all stakeholder interests may converge, may be for government to avoid mandating DRM standards, but at the same time to speak clearly (both in the legislative branch and in the courts) in stating that DRM should not be implemented or enforced in ways that contract the dynamic range of lawful uses of copyrighted works that our copyright law historically has allowed, and on which our open, democratic society continues to depend. But our government can't be expected to speak clearly on these issues unless it *hears* clearly from librarians, and from the public in general, how the traditional balance of rights under our copyright law must be preserved even as the landscape of copyright is changed by our transition into a digital world.

APPENDIX I: Analog Versus Digital

It's important to understand the fundamental differences between digital technologies and analog technologies. "Digital" generally refers to representation of information, including content, as ones and zeros (or "bits"). There are a number of advantages to the use of digital technologies — the major one is that it is possible for the receiver of digital content to determine whether there has been an error in transmission, and to correct the error (by seeking retransmission of the altered or lost bits). This is why the word "digital" has a certain appeal for both consumers and vendors — the word connotes quality, because it suggests (not always accurately) that the content has been perfectly copied or transmitted for consumer use. Moreover, the fact that it can be subjected to such error-checking is what makes it possible for the content to be subjected to digital encryption and decryption techniques.

For most of the history of consumer electronics, however, analog technologies, which directly reproduce the waveforms of auditory and visual information but do not translate them to "bits," have been at the heart of home-entertainment systems. Even where digital technologies and content formats have become commonplace (as music CDs and movie DVDs are), they are most commonly used on systems with analog components (such as stereo systems that use analog connectors to connect CD players to speakers). Similarly, in the United States, most TV watchers view television content through analog TV displays, even when the actual signal carrying that content (a cable or satellite signal, for example) may have been digital when it first arrived in the home.

APPENDIX II: Technical Issues with "Watermark" Schemes

As we have noted, a watermark must meet three technical criteria; it must be:

- Imperceptible to the user of the content
- Detectable by machines
- Difficult or impossible to remove.

Is it possible to meet these three criteria simultaneously? We find ourselves asking this question because the criteria seem to be in conflict with one another. For example, it must be possible to add to digital content a mark that is an imperceptible mark, yet it also must not be possible to subtract out that mark imperceptibly. Similarly, it must be relatively easy and cheap for any player to find a watermark; but it must be impossible for anyone to find (and then presumably remove) the watermark.

Watermarking also appears to conflict with popular data compression methods such as MPEG4 and MP3.[45] These methods reduce the size of a content file, and thus allow that file to be stored more compactly or transmitted more quickly, by discarding any aspect of the content that human eyes cannot see (or that human ears cannot hear), and that therefore is unnecessary to one's enjoyment of the content. This poses a problem for watermarks, as an imperceptible watermark consists of exactly the kind of information that such a compression method is trying to remove. If compression methods are imperfect, as today's are, then watermarks can be "hidden in the margins" by building them out of imperceptible elements that the compression methods do not yet know how to remove. But as compression methods get better, these "margins" will shrink, and it will become harder and harder to create imperceptible watermarks that are persistent in the face of compression.

As we have no solid evidence that watermarking is possible, and we have reason to doubt whether the requirements for a watermarking scheme can be met, we have every reason to doubt that a successful watermarking method will be discovered any time soon. These general reasons for skepticism are supported by the history of watermarking research, which has repeatedly shown the weakness of proposed watermarks.

[45] MPEG stands for "Moving Pictures Experts Group," which is the name of family of standards used for coding audio-visual information (e.g., movies, video, music) in a digital compressed format. MPEG-4 is a one of the more recent standards of audiovisual content compression, and is more efficient than earlier standards, such as MPEG-1. MP3 is the term for the audio layer of the MPEG-1 standard, and is the part of the MPEG-1 standard that is used for encoding soundtracks. More recently, it has become the most common format for digital music that is traded or distributed online.

One prominent example of the problems inherent in attempting to develop a standard watermarking technology for content can be found in the experience of the Secure Digital Music Initiative (SDMI), a consortium of companies from the music, consumer electronics, and software industries. SDMI sought to design a marking-based DRM system for recorded music. After choosing a set of proposed DRM technologies, including four watermarking methods, SDMI announced a public challenge, inviting the public to try to defeat the proposed technologies.

A team of researchers from Princeton University and Rice University studied the four watermarking methods, and was able to defeat all of them - that is, to remove each watermark without unacceptably damaging the audio quality of the content—in less than three weeks of work.[46] During this time the researchers had access to less information than a real would-be copyright infringer would have had.

The researchers were able to defeat each SDMI watermark technology by first pinning down the nature of the watermark and where in the content it was hidden, and then by devising a modification to the content that would target the watermark's location and thereby either remove or mask the watermark. As an example, in one of the SDMI technologies, the watermark consisted of a small amount of noise added within a certain narrow range of musical tones. Having identified this range of tones, the researchers found it easy to isolate and suppress this noise, thus defeating the watermark.

None of the SDMI watermarks required highly advanced technology to defeat. The Princeton and Rice researchers concluded that an attacker of even moderate technical sophistication can defeat current watermarking technology. Although we cannot rule out the possibility that a major advance in watermarking technology will occur, history suggests that any purported advance would have to be subjected to substantial public scrutiny and testing before it could be deemed reliable.

[46] See Scott A. Craver, Min Wu, Bede Liu, Adam Stubblefield, Ben Swartzlander, Dan S. Wallach, Drew Dean, and Edward W. Felten. "Reading Between the Lines: Lessons from the SDMI Challenge," Proceedings of 10th Annual USENIX Security Symposium, Washington, DC, August 2001.

APPENDIX III: Technical Issues with "Fingerprinting" Schemes

As discussed in the main body of this paper, a fingerprint is a "mark" that is extracted from the preexisting characteristics of the content. For example, if the content is a recorded song, then the fingerprint may be derived from the song's tempo, its rhythms, the length of its verses or movements, the mix of instruments used, and similar features.

To be effective, a fingerprinting method must be:

- Unique or At Least Precise
- Difficult or Impossible to Remove

As is the case with digital watermarks, it is not established whether it is even possible to meet these two criteria simultaneously.

Unlike a watermark, which can carry instructions about how content is to be treated, a fingerprint carries no descriptive data about the content but can only to serve as a unique identifier for a particular content file. Information about the copyright status and permissions associated with the content cannot be stored in the fingerprint, but must be obtained from a database somewhere. It follows that in a DRM system based on fingerprinting, every player must be connected to the Internet (or some similar system) so that it can contact the database to check the status of each content file before playing that file. This fact rules out the use of fingerprinting in many DRM scenarios.

There has been no generally recognized public scientific research on the question of whether a fingerprinting method can be both precise and persistent. This is not to say that there may be no use for fingerprinting. As it happens, fingerprinting has uses other than DRM, and the evidence indicates that it has promise for those other uses. The key unanswered question is whether a fingerprint can be persistent — whether an attacker can find a way to modify the content so that the fingerprint changes, without damaging the perceptual quality of the content. In the absence of evidence suggesting that fingerprints are persistent, it is appropriate for us to be skeptical about them.

Mike Godwin (mnemonic@well.com) is an attorney, a fellow at the Yale Information Society Project, and a research scientist affiliated with the PORTIA Project at Yale University

(NSF Information-Technology- Research Program grant 0331548)