Limiting Distribution: Personal DRM on the Web

Tara Whalen
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia
B3H 1W5 CANADA
whalen@cs.dal.ca

Carrie Gates
CERT NetSA
Carnegie Mellon University
Pittsburgh, PA
15213 USA
cgates@cert.org

Jason Rouse
Cylab
Carnegie Mellon University
Pittsburgh, PA
15213 USA
rouse@andrew.cmu.edu

ABSTRACT

Digital rights management (DRM) has traditionally been discussed in the context of limiting distribution due to economic reasons. We argue in this paper that the issue of limiting distribution of digital material is much broader, and encompasses privacy concerns as well as economic concerns. We further argue that people may wish to limit distribution across two dimensions: the number of people with access to the material, and the length of time material is available. We present this argument in the context of the world wide web, which is an ubiquitous medium for limited publication by individuals, and provide the legal context for this discussion, arguing that privacy is already a component of the legal system for the use of both published and unpublished works. We conclude by presenting the requirements of a DRM system in the context of the web, and discuss how existing and proposed systems meet these requirements.

Keywords

privacy, copyright, digital rights management

1. INTRODUCTION

It is widely accepted that individuals have a right to privacy, and this right has been recognized by laws such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in the United States [13], the Gramm-Leach-Bliley Act of 1999 in the United States [10], and the Personal Information Protection and Electronic Documents Act (PIPEDA) of 2000 in Canada [17]. However, these laws have focused on preventing companies from releasing private information, such as health or financial records. On the opposite end of the spectrum, it has also been acknowledged that individuals have the right to publish creative works, and copyright laws have been developed to protect the rights of authors. These laws stipulate that authors own the works they have produced, and that others may only use them in the context of fair dealing (Canada) or fair use (United States) restrictions.

We note that individuals publishing on the World Wide Web is not addressed by current legislation. Individuals still have the right to privacy, as is evidenced by both the legislation mentioned above, and by articles within copyright legislation (discussed in Section 3). However, laws such as HIPAA and PIPEDA do not apply as individuals have posted the information themselves, and it is likely of a personal (as opposed to "private") nature. It could, however, be argued that the information has been "published", and therefore copyright laws apply. We disagree with this too, based on articles by L. Jean Camp who notes that copyright systems require content integrity, document persistence and the binding of reputation to the document [5, 6]. Web-based information lacks both content integrity (as content can be easily and frequently changed) and document persistence (as documents can be moved, changed and deleted).

In fact, individuals often do not want to publish information in the strong sense implied by copyright. Rather, they are interested in limiting distribution of information. On the one side, they might be interested in limiting distribution to a small number of people (which would be considered unpublished works under current legislation). On the other, we have found that people are interested in limiting the length of time that information they have publicly posted in available. In the first instance, people may want to be assured that only those others that they have specified have access to particular material. For example, they may want their private blog or personal photographs to only be available for viewing by friends and family. In the second instance, people may want to make certain information publicly available, but either have it expire after a particular period of time, or be able to ensure that only the latest information is available. For example, they may want to have newsgroup postings be available for only a year or so, or they may want to ensure that any access to their resume always retrieves the most current version.

Limiting distribution of digital material has also been of keen interest to the entertainment industry, who have been a driving force behind digital rights management systems (DRM). While DRM has been presented as a method for enforcing copyright (and which is refuted in [5] and [6]), we view it as a method of limiting the distribution of electronic information. In particular, the entertainment industry is has used it to restrict access to digital material to those individuals who have paid for that access, preventing them from sharing this material freely with others who have not paid

for access. The motivation here for the entertainment industry is economics — artists should be fairly compensated for the works they produce, and the entertainment industry is interested in maintaining a system where their works do not become available for free.

However, we argue in this paper that limiting distribution of digital material has a much broader range than is currently considered in popular media, extending to individuals who might want to restrict access to items ranging from draft versions of papers or reports, to family photographs, to newsgroup postings. Thus there is not only an economics aspect to digital rights management, but also a privacy aspect. We demonstrate how this privacy aspect is supported by the legal systems in Canada and the United States in their definitions for copyright protection of both published and unpublished works.

This paper is structured as follows: we first discuss the new paradigm of considering privacy as a motivating factor for digital rights management of personal information, and of considering there to be a temporal aspect to "published" information. This paradigm is motivated by a previous user study performed by Whalen and Gates [20], where users expressed the desire to limit access to some information, and to have other information be removed from public view after some period of time, both in the context of web search engines. We then present the legal context for this argument in Section 3, showing how privacy is already a part of current law for copyright protection. In Section 4 we present the requirements for a personal digital rights management system (DRM) that protects web-based information, focusing on the security requirements of such a system, and where current DRM security requirements could be relaxed. We then describe current DRM and access control approaches, and discuss which requirements are met by each system. Concluding remarks are provided in Section 5.

2. A NEW PARADIGM

Creative works in digital form, such as music, written articles, and animated films, are readily available through online distribution methods that allow creators to easily share their work with the public. However, not all materials that people create are intended for widespread public consumption. Sometimes, a work is only meant to be shared with a select circle, which means that the author does not want it to be further distributed. This is tightly integrated with the notion of privacy: there may be many reasons why the material is not publicized (e.g., it may be an early unfinished draft), and there is a need to protect the work by limiting its exposure. From family photographs, to personal letters, to journal entries, there are several examples of personal creations that are never meant to reach a large audience. The effects of revealing this material could be quite detrimental, such as embarassment or the disruption of personal relationships. In a professional context, consider the case of drafts: if confidentiality cannot be relied upon, authors might refrain from circulating preliminary material for fear that it would be prematurely exposed, which may prevent them from receiving useful feedback during the revision process.

This need for limited exposure is part of the digital realm as well, but here it is very easy for a work to spread outside the author's desired boundaries. This may not happen deliberately, but may be entirely accidental: for example, mistakes in configuring a private website may lead to the text being exposed, or recipients may forget that the material was not intended for distribution and may pass it on to others. When such material is online, global exposure can happen quickly and easily. The purportedly private text may be indexed in a search engine, making it instantly accessible to anyone who types in the correct search terms.

In addition to limiting exposure of material based on the number of people who can view it, we have identified time as a second dimension along which people may wish to limit exposure. In particular, a previous study carried out by Whalen and Gates [20] identified that people did not believe that information should be available on the web for long periods of time. This study consisted of an on-line survey that was completed by 16 individuals from the Dalhousie University School of Library Sciences. Out of 15 questions, focusing on journal entries, newsgroup postings and photographs, not one respondent felt that any of this information should be available forever (out of 240 total responses). The majority of respondents felt that information should either never be available (74 responses), be available for less than three years (81 responses), or be available until the source changed (84 responses). The results of this study argue for the need to provide users with a method to revoke, update or expire published information, at least in the context of personal information that has been published on the web.

We wish to address this problem in this paper: how can we make it easier for authors to limit the distribution of their work, not due to issues of compensation, but due to privacy concerns? Traditionally, the digitial rights management (DRM) debate has remained within the sphere of consumer and creator rights. This is fundamentally an economic argument, which tries to set appropriate limits so that content creators (e.g., authors, musicians, filmmakers) will be fairly compensated and motivated to produce, while maximizing the degree to which society at large is able to benefit from these creations. It is understandable that this aspect of copyright debate has been at the forefront, given the magnitude of the problem and the impact on the economy (particularly within the entertainment industry). However, we feel that this discussion has been missing one vital aspect: when authors seek to limit publication, it may not always be to maximize revenue. Sometimes, it may be that the work is simply not intended for publication. There may be personal reasons for constraining the distribution of material, and the DRM debate could benefit from including this aspect when determining appropriate methods and levels of control. In fact, DRM may be the best way to provide technical solutions to protect unpublished digital works. Copyright law is beneficial and necessary in that it present legal regulations for appropriate use; however, for actual enforcement of author's rights, we must turn to DRM for guidance.

The second issue to be addressed in this paper is how we can make it easier for authors to limit the continued distribution of their work. That is, how can we allow authors to provide updates to published versions of their work, or to expire their work after some period of time, or even to revoke use of their work and remove it from the published domain?

3. LEGAL CONTEXT

In considering the problem of limiting distribution of the creative works of authors, we first present examples of materials that might require limited distribution (unpublished works), and discuss them in the context of related work on privacy. We then describe how unpublished works are considered under copyright law (in the US and Canada). In particular, we demonstrate that existing copyright law makes special provisions for unpublished works, in recognition of their distinct status from published works. In addition, we demonstrate how fair use provisions are strongly curtailed where unpublished works are concerned, which indicates that privacy rights are an established component of copyright law.

3.1 The Need for Restricted Publication

> I confess to a qualm about this potential immortality of digital information: The loss of second chances, fresh starts. In static societies people's reputations may have faded slowly, but in today's complex and dynamic world, the capability to begin over is useful. We have been able to recreate ourselves when we go to university, change jobs, or move, distancing ourselves from others' expectations and assumptions about us. This capability is disappearing... To me this is daunting. Starting over can be a way to grow, to experiment and optimize one's potential. [11]

Grudin [11] succiently presents a compelling reason for wishing to limit publication: creating a global online "dossier" about oneself makes it much harder to maintain a desirable persona. There is a need to be able to corral immature works, to keep personal materials within personal communities, and even to allow for some material to expire without being archived. Current copyright law does not grant mortality to digital creations, but it certainly does address issues of distribution and publication. Author control of copying permits some degree of "flow control" over those elements that for whatever reason — immaturity, personal elements — are to be excluded from one's public online presence.

Palen and Dourish [14] discuss some of the more complex issues with online privacy, and describe privacy concerns along three boundaries: disclosure, identity, and temporality. Disclosure — the choice of what elements we wish to reveal — requires some negotiation between public and private faces. Identity represents the boundary between self and other, which must take into account the nature of the "others" with whom one interacts. Temporality deals with the effects of past actions on future events, which is part of a continuum. Within the context of unpublished materials, we can see effects on all three boundaries: the need to restrict disclosure of some personal elements, the construction of self in relation to specific bounded audiences, and the wish to limit the persistence of immature or temporary works into the future. Authors therefore require the ability to manage their unpublished works in order to successfully navigate these privacy boundaries.

Some materials are dependent on a specific context to derive the appropriate meaning, and are unsuitable once removed from that limited sphere. Many types of personal material would fit into this context, such as letters or family photos, meant for a restricted private audience who would understand the lives of the people involved. Other materials may be relevant only to those in a particular temporal, social, or geographic context; for example, articles written for a small social group may involve in-jokes or other references that would be confusing or misleading to "outsiders." Other works may represent one point in a continuum of evolution, such as a novel undergoing revision or a film being edited. The intermediate stages may be a source of embarrassment to the creator; the intermediate stages do not represent the artist's final public version, and to exhibit these pieces publicly (without the artist's permission) is a violation of the spirit of the creative process.

We therefore believe that there is a crucial need for people to create and control non-public online material, which should be protected as much as possible. Copyright law already contains provisions for dealing specifically with these works, with the sections on unpublished works; we will visit these elements in the following subsection.

3.2 Unpublished Works: Definition and Examples

Under American and Canadian law, any tangible work is protected by copyright when it is created, which includes both published and unpublished works alike.[1]. Note that copyright also includes the right *not* to publish. The cases we are considering here — works such as drafts, not intended for full disclosure — are considered unpublished, even if the materials have been circulated to another party (other than the author). The distinction here is that there is a limit to the intended audience, rather than a full release to the public.

Canadian copyright law uses the term "published" to refer to "making copies of the work available to the general public in quantities that satisfy the reasonable demands of the public, taking into account the nature of the work in question." [12, p. 97] Deliberately restricting the circulation of material (for example, to a group of friends) demonstrates that the work was not published, as it was not intended for the public. The US Copyright Act contains a similar notion of publication: "'Publication' is the distribution of copies or phonorecords of a work to the public by sale or other transfer of ownership, or by rental, lease, or lending." [18] Addtionally, the U.S. Copyright Office states that, "The legislative reports define 'to the public' as distribution to persons under no explicit or implicit restrictions with respect to disclosure of the contents." Therefore, when materials have been given to a group with the understanding that the author wants no further distribution, the work is not considered to be published.

3.3 Unpublished Works and Copyright

Copyright law distinguishes between unpublished works and published works. Until recently, Canadian law protected unexploited works in perpertuity (or until they were published or performed in public); this led to such situations as historians being unable to ever reproduce unpublished mate-

[1]The authors will deal primarily with US and Canadian law in this paper

rial (e.g., letters, journals) when creating scholarly works, except limited amounts under fair dealing provisions [12]. This changed in 1997, under Bill C-32, which now allows unpublished material (with a few exceptions) to pass into the public domain 50 years after the calendar year in which the author dies [12]. In this case, the duration of copyright of unexploited and published works is the same, but the Canadian Copyright Act, by specifically addressing and defining 'published' and 'unpublished' material, recognizes that these are two distinct categories of works under the law. Similarly, Access Copyright — a Canadian agency that licenses access to copyrighted works to such institutions as universities and non-profits — specifically omits unpublished works from the list of items that its licenses allow to be reproduced [1].

American copyright law also recognizes the special status of unpublished works, although these distinctions have been reduced via updates in copyright law (particularly starting in 1978). Similar to Canadian law, unpublished and published works alike revert to the public domain after a set period: under US law, life plus 70 years. However, within the text of the Copyright Law of the United States of America, these two categories of works are still discussed separately, and a variety of grandfather clauses have been to allow for legal changes affecting pre-existing unpublished works [19]. Therefore, despite the similarities in duration, there are sufficient differences in the nature of these materials to require distinct treatment under the law: the two categories have not been simply merged into one. As we will discuss in the following section, the most striking difference between unpublished and published works, under copyright law, is how they are considered under fair use provisions.

3.4 Unpublished Works and Fair Use

Fair use (or in a related sense, fair dealing) is the ability for a party to reproduce portions of copyrighted works without permission of the copyright holder, under specific limited circumstances. Canadian law uses the term fair dealing; this is not explicitly defined in the Copyright Act, but instead, a specific set of exceptions are laid out (e.g., for research or private study, criticism or review). Within this set of exceptions, reproduction is not considered to be an infringement of copyright. Although these provisions are restrictive, there is some leeway for the courts to determine whether or not a person has engaged in fair dealing. In a 2004 Supreme Court of Canada case, the court outlined a set of criteria that could be used to determine the bounds of fair dealing:

> Fairness depends upon the context and facts of each particular case. There are a number of identifiable factors that might influence an assessment of fairness, none of which is conclusive or binding. They are: the purpose of the dealing, the nature of the dealing, the amount of the dealing, alternatives to the dealing, the nature of the work in question, and the effect of the dealing on that work. [7]

One of these items specifically deals with published versus unpublished material, under the category of "the nature of the work":

> The nature of the work in question should also be considered by courts assessing whether a dealing is fair. Although certainly not determinative, if a work has not been published, the dealing may be more fair in that its reproduction with acknowledgement could lead to a wider public dissemination of the work — one of the goals of copyright law. If, however, the work in question was confidential, this may tip the scales towards finding that the dealing was unfair. [7]

Note that the confidential aspect of the work is part of determination of fairness; here we see a clear link between copyright and privacy, within a legal context. It is therefore fitting that we look to DRM as a possible method for protecting privacy in some cases.

American law differs from Canadian law slightly on the topic of fair dealing; the terminology is 'fair use', and the chief difference is that no specific exceptions have been granted. Instead, the concept is more open-ended, allowing for determinations on a case-by-case basis of whether a copyrighted work has been used fairly [9]. Guidelines are available to assist people in deciding whether or not their intended use falls within the proscribed limit. Section 107 of the US Copyright Act lists four factors to be considered in determining fair use:

1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
2. the nature of the copyrighted work;
3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
4. the effect of the use upon the potential market for or value of the copyrighted work.

An addendum states that, "The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors." Therefore, unpublished work may, in some cases, be reproduced under fair use guidelines: it is not automatically exempt.

The topic of fair use of unpublished materials has been debated in legal circles, as it is not always clear whether such material has been used appropriately. In the majority of cases, when a work is unpublished, this puts it at a disadvantage for being allowed. The Copyright Management Center at Indiana University has created a checklist for fair use, which lists conditions that favour and oppose fair use (see `http://copyright.iupui.edu/checklist.htm`), informed by the Copyright Act. Within this framework, "unpublished work" falls under the category of "opposing fair use": this does not automatically disqualify use, but it does indicate that a stronger case must be made to allow copying.

Legal scholar Kenneth Crews, in a comprehensive article on fair use of unpublished materials, discusses court decisions on several key cases in copyright law [8]. His summary indicates that in the majority of cases, unpublished materials

have been denied as part of fair use provisions. This is generally due to economic factors; for example, if the author wishes to publish the material at a later date, having excerpts already published may interfere with their ability to profit from first publication. However, Crews argues that the assumption that there are overriding privacy concerns with unpublished works may be poor grounds for legal decisions, as no such interests may exist in all cases; he suggests "a restructuring of the burden of proof in fair-use cases to put the burden on the party best able to present proof of the peculiarities surrounding unpublished works" [8, p. 68] For example, Crews cites the case of J.D. Salinger, who wished not to have some of his archived letters reproduced in a biography; because the author was known to aggressively protect his privacy, this would indicate intent to suppress publication, and therefore the courts would be unlikely to grant fair use (and did not, when the case was tried in 1987). DRM may therefore serve as a means for expressing the author's desires in terms of the privacy of their unpublished materials, which would provide substantive material for legal cases.

3.5 Summary of Legal Status of Unpublished Works

Let us return to the example of a limited-distribution draft. What do these legal protections mean for such a work? Under Canadian and US copyright law, such works are already protected by the same copyright provisions as published works, allowing the author to legally restrict further distribution. Copies may not be made without permission, including archived copies of electronic material. In addition, the author retains the right of first publication, which means that the material cannot, for example, be posted on a website without permission. If the author decides to never publish the material, the only copies that may exist are those that were originally distributed (until the work passes into the public domain after the author's death).

The only means for further reproduction of the copyrighted work (without the author's permission) would be through fair use provisions, and in many cases, these would not be granted for unpublished works when privacy concerns exist. (We are not suggesting that DRM should be used to circumvent appropriate fair use; in fact, we do not support such an approach. However, this is not an overriding consideration in this specific domain, given the degree to which privacy aspects dominate the balance of rights.) Therefore, the use of DRM to enforce the author's right to limited publication is well-supported under existing copyright law. The legal framework already in place incorporates the author's right to control the spread of unpublished works: adding DRM to this framework provides enforcement without contravening the spirit of the law.

4. TECHNICAL SOLUTIONS

Often, security solutions focus on technologies that provide perfect security, or at least attempt to severely limit the ability of an adversary to circumvent the security mechanisms. In some cases, this approach is warranted. For example, we would expect strong security to be in place in electronic voting systems, or in the protection of documents relating to issues of national security. In these cases, usability is often sacrificed in order to achieve greater security.

However, we are dealing with the ability of individual users to provide reasonable security for their on-line documents. Unlike in the scenarios provided above, strong security is not a requirement. Thus, while a motivated attacker might be able to circumvent a security system (for example, by cut-and-pasting the protected document into an unprotected document, and then posting that to a web site), we assume here that it is unlikely that the user will have an adversary who is intent on posting their personal information in such a forum. Thus we require "sufficient" security, which achieves a balance between providing a reasonable amount of security and ease-of-use.

We focus our discussion here to support for limited distribution of material in the context of the World Wide Web. We do this for three reasons:

1. the web has become one of the primary means for disseminating personal information,
2. our previous studies have focused on the web and discovered a desire for limiting distribution on the web, and
3. despite its popularity, this context has been largely ignored when considering how information exchange might be limited.

From a user's perspective, the advantages of publishing information on the web include a large amount of exposure, and the ease of access both for the publisher and for those wishing to access the published material.

It should be noted that, while we are using DRM systems as a guideline, we are not specifically developing systems to protect copyright. Copyright systems have at least two functions that are contrary to web-based content: content integrity (e.g. content is not changed during storage) and document persistence (e.g. "ensuring that a document survives unaltered and can be located" [6]). L. Jean Camp refers to these two functions, in combination with the "binding of reputation to document", as "copy accuracy", and proposes that we use this concept for the basis of DRM design [5]. However, providing the ability to ensure high levels of integrity (e.g. through mass distribution), as well as archival capabilities (to prevent the change of a document's content), are not features that are available, or necessarily desirable, in web-based systems. We instead focus on the requirements to limit distribution of material over the web, both in terms of who can access material and for how long material remains available.

The requirements imposed on limiting distribution in the context of the web are more complicated than in other environments due to the varied nature of the web. Some of the requirements for such a system are:

1. the ability to update content and ensure that browsers are accessing the most recent version (e.g. resume),
2. the ability to delete an item (e.g. old photos),

3. the ability to keep an item from being indexed by a search engine,

4. the ability to specify who can view a particular item,

5. the ability to expire an item (e.g. specify that a newsgroup posting will only be available for three years from the date of posting),

6. the ability to prove authorship.

These requirements must be met in a variety of circumstances, including static html pages, images, blog entries, newsgroup messages, and postings on web forums. One of the main challenges in web based systems is the variety of forms that information can take, and the variety of locations where a user may add content. Additionally, any solutions must be interoperable. Current DRM solutions focus on the use of particular hardware or software to enforce access restrictions. In contrast, to limit content distribution on the web requires the co-operation of many software vendors providing browsers (e.g. Internet Explorer, Safari, Firefox) and servers (e.g. IIS, Apache), as well as any site providing services (e.g. Google, Yahoo, AOL, blogger.com, etc.) and any products that assist users in developing web content.

We describe a few current solutions to similar problems and discuss how they might be adapted to the web. No one solution is able to fulfill all of the requirements outlined above, however it may be possible to combine a variety of solutions that together provide the document limitation features specified above.

4.1 Deployed Systems

4.1.1 X-No-Archive

An early form of limiting distribution was based solely on time, and not on limiting who could access the information, and was aimed at newsgroup messages. X-No-Archive is a Usenet newsgroup message header that can be added by the poster to indicate that their message should not be archived for later viewing. This header was first created and used by DejaNews in 1995, in response to privacy concerns people had over DejaNews's decision to archive the entire Usenet news feed. Google and others also honour this header when archiving news feeds. However, the recognition of X-No-Archive when dealing with Usenet newsgroups has been "entirely voluntary." [21]

This approach to limiting distribution could be modified and applied to web pages and search engines. For example, a meta tag could be created that indicates that a particular web page should not be indexed by search engines. A tag could also be created that specifies an expiry date, so that a user could provide a date after which search engines would no longer index the web page or provide links to it.

However, this approach does not address any of the other system requirements, such as the ability to ensure that only the most recent version of a document is accessed, or specifying that only particular people can access a specific document.

4.1.2 Password Protected Sites

In the context of limiting distribution of personal material to a small number of people via the web, we find that some mechanisms are already in place and, in fact, in widespread use. The most common method for limiting access is to set up password protected portions of a web site, and to then provide passwords only to those people to whom we wish to provide access. However, this provides only a coarse-grained solution to the problem in that restrictions can not be provided on a per-file basis. Rather, if two files are to be released to two different groups of people, they must be placed in separate directories, each of which will be password-protected. A second method of restricting access to publish information in a public directory, but to not provide any public linkis to the directory. Access is controlled by limiting who is told about the existence of the directory. However, as in the case of password protection, access is only controlled at the level of a directory, and is less easily controlled at the file level.

However, there are other situations in which a limited distribution may be desired. For example, someone may wish to post a comment to a newsgroup or blog, but have that comment only be available to a limited set of people, and not to everyone who reads the newsgroup or blog. Another example is wishing to limit the distribution of photos or a draft of a short story to a few select friends. In this case, it may be desirable to have all of the items in the same directory, but to still provide different access to different people.

4.1.3 Adobe LiveCycle Policy Server

Digital rights management solutions have largely been developed in the context of multimedia, such as the protection of music (e.g. Apple's FairPlay DRM) or movies (e.g. Intel's High-bandwidth Digital Content Protection, HDCP). In addition, a large number of solutions have focused on hardware-based approaches, such as Digital Video Express (DIVX), Content Scrambling System (CSS), and product activation (e.g. Microsoft operating systems). Very little has been produced in the arena of protecting text. One exception to this is Abode Systems, who provide digital rights management as part of the Digital Edition of Adobe Reader (formerly eBook) [3] and their LiveCyle Policy Server [2], which is marketed as a document control solution.

The Adobe LiveCycle Policy Server allows an enterprise to create PDF files that contain the security policy of the document as part of the PDF. The policy can specify parameters such as who can view the document, if they can print it, and what the expiry date is on the document. This security is achieved through the use of a central server that serves as the authentication mechanism, via LDAP. Offline access can be granted to a user for a limited amount of time, however is not necessarily granted as a default action. By using a central server, Adobe also provides the capability for a person to revoke the ability to view a document, to change the security policy dynamically, to track document access, and to ensure that any updates to the document are pushed out to the user [3]. However, this approach requires that the organization maintain a central server which can authenticate all possible users. Additionaly, it supports only PDF documents, and not the variety of web-based documents and forums that are currently in use.

A mechanism similar to this would address many of the requirements for limiting distribution on the web, such as the ability to limit who can view a document, to delete (revoke) a document, and to expire a document. However, to move this to a web context would require a method of specifying the server where a document could be found. This could be difficult given that users can move between servers and take their web content with them. Additionally, this approach only works if browsers are co-operative.

4.2 Research Systems

4.2.1 Off-the-Record

A system called off-the-record communication has been developed by Borisov et al. [4]. This system focuses on the protection of conversations, and was developed specifically with instant messaging systems in mind. A number of requirements were identified that were based on the qualities of face-to-face converations, including the ability to deny having stated something (repudiation), the ability to prevent an eavesdropper from overhearing the conversation, the ability to authenticate the other party *during* the conversation, and the ability to ensure that past messages can not be recovered (perfect forward secrecy). This system uses a novel application of encryption techniques to ensure that the above requirements are met. While this approach was intended for use with instant messaging, an extension to support email communications was also discussed. Unfortunately, this system is focused on interactive communications between two individuals. In contrast, the web consists of a large amount of static information. Additionally, when there are interactive communications on the web (e.g. web forums), they are often between a large number of people and not just two individuals.

4.2.2 UCON and ORCON

Park and Sandhu have developed a system called UCON, for usage control [16]. This is a system that combines elements from traditional access control, trust management and digital rights management to provide a comprehensive information management system. Like traditional access control systems, it consists of subjects and objects, which are related by rights. Rights can be subject to authorization rules, conditions and obligations. For example, the user (subject) might be required to pay (obligation) before having the right to access a particular song (object). This model, like the Adobe model, addresses the case where the distribution of material is not payment-based. Also similar to the Adobe model, this paper proposes the idea of a client-side reference monitor that would provide the access control capabilities.

The concept of UCON was extended in a later paper that discussed ORCON, for originator control [15]. This paper described a ticket-based system where the originator of an object could maintain control over the re-dissemination of an object. This paper also provided more deployment details on UCON, noting that a virtual machine would be required, and that any UCON-controlled object must only be accessed through this virtual machine. It also described the digital container that would be used for protecting the content on the virtual machine. Again similar to the Adobe system, the control set containing the usage rights for an object is included in the digital container and moves with the object. There is also a control center that might provide additional access information.

This approach suffers from many of the same drawbacks as the Adobe LiveCycle system when applied in a web-based environment. For example, it requires that all users use a particular piece of software for access, and so reduces interoperability. Unlike Adobe, this system focuses more on static information and so does address the issue of how to update content, however it should be easy to have any access of an object result in a check to the control center for updates.

5. CONCLUDING REMARKS

In this paper we have argued for a system that protects the privacy of individuals by providing methods to limit distribution of electronic material on the web. In particular, we note that individuals should be able to limit information across two dimensions: the people to whom they provide the information and the amount of time that information is available. We provide the legal context for such a system, noting how privacy is already a consideration in copyright law for both published and unpublished works.

We then provide the requirements for deploying such a system, and examine some deployed systems and research systems, noting how each meets (or does not meet) the requirements for deployment on the web. The two systems that come the closest to providing the necessary functionality are Adobe's LiveCycle Policy Server [3, 2] and UCON [16, 15]. However, both systems require specific software at the client side (Adobe Reader and a virtual machine, respectively), which does not meet the interoperability requirements of web-based information. Additionally, the system needs to be designed to respect security in those cases where a user has requested it (e.g. they want to limit the content to a particular group of people, or they want to have it expire in two years), but to be completely unobtrusive in those cases where the user wants to post information with no restrictions. The current systems assume that security is desired in all cases. Additionally, users should not need to identify themselves to access content unless the content specifically requires it *because* there is a limitation on who can access it. If there is no security on the material, or the only restriction is a simple expiry date, then any user of that data should not be expected to reveal their identity. This, too, is contrary to how current DRM and access control systems are designed.

Acknowledgments

The authors would like to gratefully acknowledge the support of the IBM Canada Centre for Advanced Studies in Toronto and the National Sciences and Engineering Research Council of Canada (NSERC).

Bibliography

[1] Access Copyright. Canadian copyright act. `http://www.accesscopyright.ca`, 2005. Last visited: 18 April 2005.

[2] Adobe Systems. Adobe livecyle policy server overview. `http://www.adobe.com/products/server/policy/overview.html`, 2005. Last visited: 18 April 2005.

[3] Adobe Systems. Adobe reader 7.0 frequently asked questions for digital edition users. http://www.adobe.com/products/acrobat/pdfs/95003923_eBook_FAQ_UE.pdf, 2005. Last visited: 18 April 2005.

[4] Nikita Borisov, Ian Goldberg, and Eric Brewer. Off-the-record communication, or, why not to use PGP. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society*, pages 77–84, Washington, DC, USA, 2004. October 28, 2004.

[5] L. Jean Camp. DRM: Doesn't really mean digital copyright management. In *Proceedings of the 2002 ACM Conference on Computer and Communication Security*, pages 78–87, Washington, DC, USA, 2002. Novemebr 18-22, 2004.

[6] L. Jean Camp. First principles of copyright for DRM design. *IEEE Internet Computing*, 7(3):59–65, 2003.

[7] CCH Canadian Ltd. v. Law Society of Upper Canada. 1 S.C.R. 339, 2004 SCC 13, 2004.

[8] Kenneth D. Crews. Fair use of unpublished works: Burdens of proof and the integrity of copyright. *Arizona State Law Journal*, 31:1–93, 1999.

[9] Dictionary.LaborLawTalk.com. Fair use. http://encyclopedia.laborlawtalk.com/Fair_use, 2005. Last visited: 18 April 2005.

[10] Gramm-Leach-Bliley Act. 15 USCS §6801-6809, et seq. 106 P.L. 102, 113 Stat. 1338, 1999 Enacted S. 900; 106 Enacted S. 900, 1999.

[11] J. Grudin. Desituating action: Digital representation of context. *Human-Computer Interaction*, 16(2-4):269–286, 2001.

[12] Lesley Ellen Harris. *Canadian copyright law: the indispensible guide for writers, musicians, visual artists, filmmakers, publishers, editors, teachers, librarians, students, lawyers and business people.* McGraw-Hill Ryerson, 2001.

[13] Health Insurance Portability and Accountability Act. Public Law 104-191, 1996.

[14] L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 129–136, Fort Lauderdale, FL, 2003.

[15] Jaehong Park and Ravi Sandhu. Originator control in usage control. In *Proceedings of the Third International Workshop on Policies for Distributed Systems and Networks*, pages 60–67, Monterey, CA, USA, 2002. June 5-7, 2002.

[16] Jaehong Park and Ravi Sandhu. Towards usage control models: Beyond traditional access control. In *Proceedings of the 2002 ACM Symposium on Access Control Models and Technologies*, pages 57–64, Monterey, CA, USA, 2002. June 3-4, 2002.

[17] Personal Information Protection and Electronic Documents Act. S.C. 2000, c. 5 (PIPEDA), 2000.

[18] US Copyright Office. US Copyright Act, 17 U.S.C., 1976. http://www.copyright.gov/circs/circ1.html#pub, Last visited: 18 April 2005.

[19] US Copyright Office. Copyright law of the United States of America and related laws contained in title 17 of the United States code. http://www.copyright.gov/title17/circ92.pdf, 2003. Last visited: 18 April 2005.

[20] Tara Whalen and Carrie Gates. Public perception of private information on search engines. Research Technical Report CS-2005-01, Dalhousie University, January 2005.

[21] Wikipedia. X-no-archive. http://en.wikipedia.org/wiki/X-No-Archive, 2005. Last visited: 13 April 2005.