PROCEEDINGS OF THE FIRST INTERNATIONAL MOBILE IPR WORKSHOP: RIGHTS MANAGEMENT OF INFORMATION PRODUCTS ON THE MOBILE INTERNET

Olli Pitkänen (ed.)

August 27, 2003

HIIT
Publications
2003-2

MobileIPR 2003

Proceedings

First International Mobile IPR Workshop

Rights Management of Information Products on the Mobile Internet

August 27-28, 2003, Helsinki, Finland

MobileIPR 2003

ISBN: 951-22-6675-X
ISSN: 1458-9451

Editor: Olli Pitkänen
Helsinki Institute for Information Technology HIIT
olli.pitkanen@hiit.fi
http://www.hiit.fi/
P.O.Box 9800, 02015 HUT
Tammasaarenkatu 3, Helsinki
FINLAND

Technical editor: SuviSoft Oy Ltd
Hermiankatu 3A
FIN-33720 Tampere
FINLAND

Sponsor

Table of Contents

SESSION ThuPmOR2: Session 6

PREFACE

The 90's were blessed by two simultaneous and symbiotic killer apps: the Internet and the mobile telephony. The relatively simple basic services of Internet and GSM access were introduced by the industry, and enjoyed phenomenal commercial success that made the rapid build-up of the required infrastructure feasible. Ever since, the industry has sought for the Next Big Thing that will bring back the golden boom days.

In the present sobering economical climate, many actors are starting to recognise that no such thing is likely to emerge on the stage. There is no Next Big Thing, universally demanded by an unsatisfied market with deep pockets filled with extra money. Instead, the future market looks more like the medieval map of Central Europe: a landscape consisting of an endless variety of niches defined by user segments, professional occupations, age groups, cultural variations, tastes, and quirks of economic infrastructure and most of all a multitude of turnpikes for collecting an exotic array of supposedly legitimate payments from travellers.

We are thus moving away from an environment characterised by economy of scale to a much more challenging environment characterised by economies of speed and scope: instead of simple services provided to a large mass of homogeneous users, we should learn to work on a rapidly changing market requesting segmented, tailored, personalised, user-configurable set of niche services.

Digital content, its creation, distribution and use occupy a central role in this new landscape. In addition to the traditional model of content produced by content providers for what were thought of as a homogeneous mass of consumers, we must also take into account content created, shared, and used by the users themselves. Similarly, we should also consider content derived from public services, such as much of educational and cultural content.

Indeed, we believe that the viewpoints and concerns of several stakeholders – content owners, end users, public authorities, and civil society – should be taken into account in a balanced manner as a basis of developing the infrastructure and

platforms of the next generations of mobile and fixed networks ultimately leading to the Mobile Internet.

Digital Rights Management (DRM) is an area where many of these issues converge. How can the legitimate rights of content owners be balanced with the rights of the end users and societal issues such as free speech, fair use, and bridging the digital divide? How should the underlying rules and regulations of the digital content marketplace be formulated to make sure that it becomes a dynamic driver of future applications, and provides opportunities for new service creation for all stakeholders?

Much of the research at the Helsinki Institute for Information Technology is aimed at studying these and related issues to give us a firm basis for our work aimed at building the technological platform for the Mobile Internet. Obviously, these themes are extremely complex and challenging, and require a fruitful integration of many viewpoints and disciplines from computer science to law, economics and societal studies. They also require truly international co-operation to build a basis of understanding the various conditions and issues shaping the field in various countries and societies.

It is therefore a great pleasure for me to welcome the participants of the *First International Mobile IPR Workshop: Rights Management of Information Products on the Mobile Internet* to Helsinki, Finland. As seen from the title, we hope that this event will be just the first of a series of workshops to be devoted to this central and challenging theme that will delineate so much of the future. I trust that the workshop will prove to be a useful opportunity for all participants to share ideas, views, and concrete research results, and that it will be an important step forward to create an intellectual basis of the future Mobile Internet for all.

Helsinki, July 2, 2003

Professor Martti Mäntylä
Director, Helsinki Institute for Information Technology

Acknowledgement

Dear participants,

On behalf of the local organizing committee, I would like to thank all the people and organizations that have made this workshop possible.

The major funding was granted by the Finnish National Technology Agency *Tekes*. Also, the following companies have generously supported our work: *Nokia*, *Sonera*, *Elisa*, and *Finnish Broadcasting Company YLE*.

The keynote speakers Professor *Hal Varian* (the School of Information Management & Systems, SIMS, University of California, Berkeley) and Professor *Ross Anderson* (Computer Laboratory, University of Cambridge) are vastly important to the success of the workshop.

The program committee members and the reviewers have made an effort to ensure the quality of the workshop:

- Professor Jaime Delgado, Universitat Pompeu Fabra, Spain
- Professor Kalevi Kyläheiko, Lappeenranta University of Technology, Finland
- Professor Martti Mäntylä, Helsinki Institute for Information Technology, Finland
- Professor Matti Rossi, Helsinki School of Economics and Business Administration, Finland
- Professor Jon Bing, Universitat of Oslo, Norway
- Professor Bernt Hugenholtz, University of Amsterdam, Netherlands
- Professor Sirkka-Liisa Järvenpää, University of Texas at Austin, United States
- Dr. Jukka Kemppinen, Helsinki Institute for Information Technology, Finland
- Dr. Till Jaeger, Institut für Rechtsfragen der Freien und Open Source Software, Germany
- Dr. Juha Laine, Electronic Cormmerce Institute, Finland
- Mr. Habtamu Abie, Norsk Regnesentral, Norway
- Mr. Stamatis Karnouskos, Fraunhofer FOKUS
- Ms. Lilian Edwards, Edinburgh University, United Kingdom
- Ms. Aura Soininen, Helsinki Institute for Information Technology, Finland

- Mr. Perttu Virtanen, Helsinki Institute for Information Technology, Finland

Authors have created the real essence of the workshop.

Local organizers, especially MobileIPR project team (in addition to me, Mr. Mikko Välimäki, Mr. Ville Oksanen, and Mr. Tommo Reti), Ms. Pirkko Miettunen, Mr. Herkko Hietanen, and Mr. Jan Fagerström (Conference Consultant at Helsinki University of technology) have taken care of arrangements. Suvisoft provided us with the web service.

And of course, the participants finally make the workshop complete.

Olli Pitkänen

Local Organizing Chair

Helsinki Institute for Information Technology HIIT

Keynote: The Social Cost of Sharing

Hal Varian
School of Information Management and Systems, University of California, Berkeley

I examine which sorts of intellectual property will and will not be produced under sharing. Several types of seller behavior are examined including monopoly, limit-pricing monopoly, and regulated monopoly.

Bio:

Hal R. Varian is the Dean of the School of Information Management and Systems at the University of California, Berkeley. He is also a Professor in the Haas School of Business, a Professor in the Department of Economics, and holds the Class of 1944 Professorship.

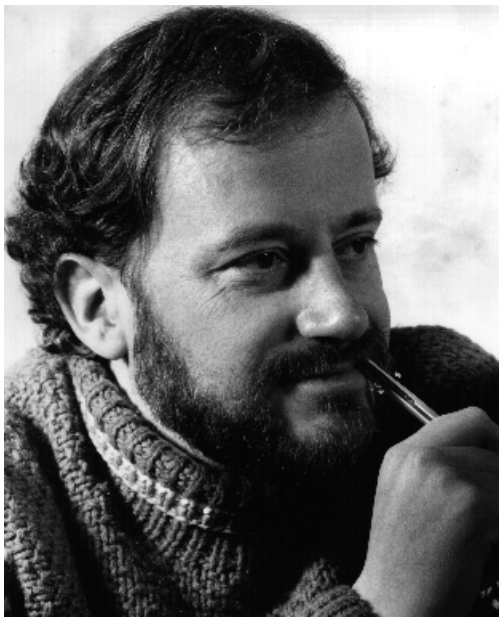
He received his S.B. degree from MIT in 1969 and his MA (mathematics) and Ph.D. (economics) from UC Berkeley in 1973. He has taught at MIT, Stanford, Oxford, Michigan and other universities around the world.

Professor Varian is a fellow of the Guggenheim Foundation, the Econometric Society, and the American Academy of Arts and Sciences. He has served as Co-Editor of the *American Economic Review* and is on the editorial boards of several journals.

Professor Varian has published numerous papers in economic theory, industrial organization, financial economics, econometrics and information economics. He is the author of two major economics textbooks which have been translated into 22 languages. His current research has been concerned with the economics of information technology and the information economy. He is the co-author of a bestselling book on business strategy, *Information Rules: A Strategic Guide to the Network Economy* and writes a monthly column for the *The New York Times*.

Keynote: Cryptography and Competition Policy Issues with `Trusted Computing'

Ross Anderson
Cambridge University



The most significant strategic development in information technology over the past year has been `trusted computing'. This is popularly associated with Microsoft's `Palladium' project, recently renamed `NGSCB'. In this presentation, I give an outline of the technical aspects of `trusted computing' and sketch some of the public policy consequences. Also, issues such as the new draft directive on IP enforcement and the emerging position of the EU DG Competition on the abuse of security and IP as a means of thwarting competition policy are relevant and I will include them into the talk.

Bio:

Ross Anderson leads the security group at the Computer Laboratory, Cambridge University, where he is Reader in Security Engineering. He is a Fellow of the Institution of Electrical Engineers and of the Institute of Mathematics and its Applications. He has well known research publications on a number of security policy topics including medical privacy and banking systems, as well as on the underlying technologies such as cryptography and tamper resistance. He is one of the fathers of peer-to-peer systems: his 1996 paper on the `Eternity Service' held out the vision of a censorship-resistant file store spread across the whole Internet, a concept now implemented by many file sharing networks. He is also one of the pioneers of the study of security economics. His book `Security Engineering - A Guide to Building Dependable Distributed Systems' has become a standard text.

Committee Members

- Local Organizing Committee:

Olli Pitkänen (chair)
Mikko Välimäki
Ville Oksanen
Tommo Reti
Jan Fagerström (Conference Consultant at Helsinki University of Technology)

- Technical Program Committee members:

Professor Jaime Delgado, Universitat Pompeu Fabra, Spain
Professor Kalevi Kyläheiko, Lappeenranta University of Technology, Finland
Professor Martti Mäntylä, Helsinki Institute for Information Technology, Finland
Professor Matti Rossi, Helsinki School of Economics and Business Administration, Finland
Professor Jon Bing, Universitat of Oslo, Norway
Professor Bernt Hugenholtz, University of Amsterdam, Netherlands
Professor Sirkka-Liisa Järvenpää, University of Texas at Austin, United States
Dr. Jukka Kemppinen, Helsinki Institute for Information Technology, Finland
Dr. Till Jaeger, Institut für Rechtsfragen der Freien und Open Source Software, Germany
Dr. Juha Laine, Electronic Cormmerce Institute, Finland
Mr. Habtamu Abie, Norsk Regnesentral, Norway
Mr. Stamatis Karnouskos, Fraunhofer FOKUS
Ms. Lilian Edwards, Edinburgh University, United Kingdom
Ms. Aura Soininen, Helsinki Institute for Information Technology, Finland
Mr. Perttu Virtanen, Helsinki Institute for Information Technology, Finland

Of Trusted Computing, Gnutella and their Emerging Ecologies: Risk Impasses and the Subtle Art of Cultivating Regulation

Prodromos Tsiavos
Department of Information Systems
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom
p.tsiavos@lse.ac.uk

Abstract

Shifting from Risk regulation to Risks produced by regulation, this study attempts to explore the emergence and operation of regulatory Ecologies as an organic approach for dealing with the issue of Risk and regulation. Objective of this paper is to investigate the reasons behind regulatory failure and set a research agenda for dealing with the issue of technology regulation, in a two-step approach. The first step is of a methodological nature: it calls for the adoption of an alternative stance for studying the regulatory phenomenon by viewing the content of regulation as a direct result of its creation process and focusing on the latter at least as much as on the former. The next step is to re-conceptualise regulation not as a means for controlling and delimiting contingencies but as an instrument for their proliferation. The Open Source development of the Gnutella protocol is illustrated as such an alternative regulatory paradigm and contrasted to a model of regulation based on control of contingencies such as that of Trusted Computing Platform or Technical Measures of protection.

1. Instead of an introduction: from regulating risks to risky regulations

Regulating Risks can be a risky business. This is by no means a new observation. Sunstein's [56] example of pollution regulation is an illustrative one: the introduction of regulations that render mandatory the installation of emission control equipment on new motor vehicles could alleviate their cost and lead car owners to retain their existing old vehicles thus increasing instead of delimiting the air-pollution levels. The case of oil spills regulation exhibits a similar pattern: cleaning oil spills often involves the use of methods and substances that may themselves be dangerous for certain ecosystems whereas the possibility of further pollution dispersion as a result of the cleaning efforts cannot be excluded [30]. It seems that attempting to intervene in any complex system entails a multiplication rather than mitigation of Risks. Grabosky [26], in an attempt to illustrate the typology of risks emanating from a variety of regulatory interventions, presents an array of regulatory examples ranging from asbestos removal [63], disposal of hazardous waste [12] and cross-border pollution [4] to liability laws that stifle business innovation [52] or tax regulation that creates parallel markets for tax avoidance enterprises [28].

Building regulation is notoriously difficult. Implementing it is even harder [46]. But what happens when the unintentional consequences become the norm, when regulation produces uniformly patterned behaviour, which is not the one desired by the creators of the regulation? We could put the blame on the politicians that are more interested in short term results and risks perceptions rather than long term solutions and the addressing of the actual risks [39]; we can talk of "bad science" or even "engineering flaws in the design and implementation of regulatory activities" [26]. However, the problem of "counter-productive regulation" [26] remains and is here to stay for as long as we approach it epidemically refusing to challenge its underlying premises [6]. A closer look at the evolution of regulation -both as theory and as practice- in conjunction with the trajectories of various technologies is essential to gain some insights into the problem of Risk.

Regulation theory has for a long time attempted to address issues of environmental, health or even financial risks [50]. As theories and practices of regulation came to a level of maturity in the mid 1990s [6] questions related to the risks created by regulatory intervention started emerging [26]. The issue was not just how regulation could contribute to the mitigation or management of risk of complex systems like the environment or financial markets,

but also how risks originating from the operation of regulations themselves could be handled. This development coincided with the "emancipation" of regulatory theory from the realms of market regulation and competition law. Regulation "vocabulary" was now to be used in areas as diverse as family law, corporate governance [57] or –most importantly for our study- internet regulation [47], [10], [40]

Research concerned with the risks related to the deployment and enforcement of regulation has been particularly intense in areas where information technology has had a profound impact, such as the Internet or any kind of digital networks. The works of Lessig [40], [41],[42], Biegel [10] or Benkler [8], [9] are excellent examples of a stream of research that sought to address the adverse effects that regulatory interventions can have on technological innovation or the balancing of rights within techno-legal ecosystems [29]. It is interesting to note that although Risk does not appear as a technical term in the literature concerning regulation of (or through) digital technologies, the theme of Risk and blame have a constant presence. To quote Baldwin [6], "as 'risk' becomes increasingly politicized and construed as 'danger' rather than its original technical meaning of statistical or mathematical probability", it becomes a central preoccupation of contemporary regulatory studies [50].

2. The importance of being ecological

In the case of the unanticipated effects that regulation concerning information technology often has, the vocabulary used to describe the phenomenon is illustrative of the reasons why a Risk driven approach is suggested [26] as a possible way of dealing with the arising issues. Boyle [14], [15] refers to "Environmentalism on the Net" and the need to develop regulatory policies that could sustain the preservation of the "ecology" of the commons in a world dominated by technical measures of intellectual property protection and over-stretched copyright regulation. Similar is the argument made by Lawrence Lessing in "The Future of Ideas" [42]. Yochai Benkler [8], [9] talks of "(t)he battle over the institutional ecosystem in the digital environment" to refer to the evolutionary process by which different production modes (such as "peer production" that characterises open source development and peer-to-peer dissemination of content) instigate the creation of different regulatory creatures. Hosein, Tsiavos and Whitley [29] have referred to the interaction between technology and regulation as a technological ecology of regulation or as *T-ecology* of regulation. The discourse on risks stemming from regulation is indicative of the image that regulation has for itself and in a great extend resembles the systems that Risk regulation is concerned with: in a rather self-referential fashion, regulation is viewed by itself as a complex system that when interacting with technology produces side-effects harmful for itself and all surrounding systems. When referring to his concept of "counterproductive" regulation, Grabosky makes an important remark concerning the systemic nature of society, the importance of an ecological approach for regulatory intervention and the role of Risk in the way late-modern [31] society operates:

"In addition to inadequate understanding of basic casual processes, there is often among policy entrepreneurs an inadequate appreciation of the systemic nature of modern society. Interventions can trigger other causal processes. The functional disruption of related systems is familiar to students of ecology. Similar principles apply in regulatory life. Regulatory policies, like public policies generally, have wider implications (…) Given the density of contemporary social space, efforts to influence one variable are likely to influence others, directly or indirectly. Engineers of a given regulatory domain are often insufficiently aware of the wider social ecology – the complex, interdependent systems of social life in which the target behaviour resides." [26]

The fact that regulatory interventions are often the source of more problems than the ones they purport to solve indicates the need for a radical reconsideration of both the way in which regulation is built and conceived. The model in which regulation is constructed and enacted in information-technology-intense settings is under increasing strains: instead of achieving regulatory integration and control of contingencies, it causes fragmentation of the regulatory modalities and proliferation of unanticipated contingencies.

This paper aims at providing an alternative perspective on how regulation should be treated in relation to Risks stemming from its very own operation. It argues that Risk as the occurrence of unintended and adverse consequences in relation to regulation cannot be handled by introducing tougher enforcement mechanisms or by re-enforcing the rights of the existing stakeholders. The lack of matching between the intentionality embedded in regulatory structures and the results of its enforcement is only the symptom of a wider phenomenon that relates to the way in which regulation is produced. Therefore, research that focuses just on the content and impact of regulation misses the point. The focus should also be on the

way regulation is produced and the degree in which different actors have the capacity to inscribe their interests into different regulatory instruments.

Such a perspective has significant implications in the way regulation is conceptualised. We suggest a shift from a mono-dimensional and state driven to a multi-source regulation model. In the mainstream regulatory model the primary goals is achievement of control of contingencies and the enforcement of a particular set of possibilities that are considered as acceptable by a central authority. On the contrary, the model suggested by this paper advocates a proliferation of contingencies by allowing more actors to have input in the regulatory production process. When control is concentrated in a single point as it happens in the case of traditional legal regulatory making is practically difficult to achieve its goals and susceptible to abuses of power. On the contrary, when control is distributed we witness the emergence of a regulatory landscape that although a first reading would suggest being a chaotic, is in fact a much more integrated to the interests of the actors solution.

The implications of such a viewpoint for Risk and regulation problems are profound as it calls for an opening of the regulation to all possible inputs and for a proliferation instead of delimitation of contingencies.

The argumentation of this paper is supported by a study of two regulatory *ecologies*, that of Trusted Computing Platform and the Gnutella Protocol. At the end of the paper a more in depth analysis is conducted, the limitations of this work are highlighted and future research tasks are set.

3. Methodology and Data collection approaches

The hypothesis and main arguments employed in this paper require a particular approach both for the treatment of the empirical data and the literature itself. In terms of dealing with the case of the Gnutella development process, my research is methodologically informed by Actor Network Theory (ANT) [55]. Although the ANT vocabulary is not slavishly followed, it has exercised a great influence in the way the empirical objects were approached.

The choice of ANT comes mainly because of the settings in which it has been used in the past by other researchers. ANT has almost invariably used in order to trace the transformation/ construction of "artefacts" [16] of all kinds, from scientific facts [34], [35] to tangible objects, markets [18], beliefs [37] and -recently- even laws [36]. Having close links with Foucault's work [37], power theories [53], the sociology of translation [18] and Science and Technology Studies [11], ANT seems a useful tool for assessing the creation process of an alternative regulatory paradigm. The term paradigmatic change is used in the way being employed in the work of Thomas Kuhn [33]. Despite the fact that Kuhn refers to revolutions in normal science rather than regulatory changes, the increased influence of the scientific and technological factor in the regulation making process and enforcement makes his work exceptionally relevant.

One of the main objectives was to avoid oversimplifications and representations of technology as a monolithic entity. Therefore, I sought to capture complexity rather than reducing it through a detailed collection of the processes needed for the development of the Gnutella protocol [20], LimeWire [43] and the Trusted Computing Platform [2], [59], [45] respectively. The fact that there was not set boundary in the phenomena under study indicates that the description "case study" [53] would be an oversimplified account of the unit of analysis I used to approach the problem domain. Instead, and because the field under study remained in a state of constant negotiation and evolution through a dialectic process between its constituent elements, we will use the term "discourse".

In order to identify the phenomena from which we would draw our data we "followed the actors" with a "rolling snowball" [11] letting them reveal the unit of analysis appropriate for testing our hypothesis. By tracing the development trajectory of the both human and non-human actors [49], i.e. the Gnutella and its developers or Copyright Law and peer-to-peer networks, I was able to draw the main involved parties and validate our hypothesis.

4. Data Collection

The data collection processes were accordingly compatible with the epistemological paradigm [23], [7] I decided to follow. In a first stage I collected all the data related to the development of the Gnutella protocol. These were of the following kinds: web sites that were used for hosting forums and file repositories related to the development of the protocol that could be either archived or still operational; messages posted on discussion groups, forums and newsgroups; the design documents of the Gnutella protocol. In a second stage I gathered material related to the Limewire application. These

included: operational and archived web sites having been used for the development of the application; applications such as Concurrent Version Systems (CVS) or Bug reporting tools (such as Issuezila), design and implementation documentation and relevant press reports.

The material gathered, covered a time span from early 2000 to 30th of April 2003. Further data are still collected for subsequent study of the topic. Sources for the material were the World Wide Web, the Internet Archive, and Usenet Groups hosted by Google. The websites analysed were the archived sites of DSS Clip2, the Gnutella Developers (GnutellaDev), the Gnutella New Generation (GnutellaNG), the Gnutella Developers Forum (GDF), the LimeWire LLC (Limewire.com), the Limewire Open source (LimeWire.org) and the LimeGroup. A variety of other web sites were also used mainly of informative character, such as the Slashdot.org or the WiredNews, as they played a very important role in the proliferation and evolution of the protocol and the applications. However, I will not make extensive use of theme as they are only marginally relevant to our argument. I also do not provide any account of the IRC mediated communications or private messages exchanged between the various developers, as the focus of this study is the crystallised forms of communication between participants as expressed in different regulatory forms.

In order to trace changes in Copyright Laws, primary sources were used in the sense of comparing different versions of the relevant Acts; secondary sources such as articles in legal journals and anecdotal information by individuals participating in the law making process, were employed.

Regarding Trusted Computing (TC), the data gathered to this point are relatively few. The main sources are the official documentation by the Trusted Computing Platform Alliance, Microsoft and Hewlett Packard as well as a set of semi-structured interviews with participants to TCP projects and European Commission employees.

Once the data were collected, they were analysed using textual analysis tools (Atlas). For the part of the web sites that was in a textual form, I attempted to draw links with other texts or to use it as guidance for the revelation of other relevant documents. The non-textual/ functional elements of the web sites were deconstructed in the steps they included in order to explicate the processes employed for the inclusion of developers and the allocation of tasks. The postings between the GDF are associated in thematically related areas and related to the various protocol versions [38].

5. A flavour of Trusted Computing

In this section we make a very basic presentation of the Trusted Computing (TC) initiative with a particular focus on its institutional infrastructure, the development process of the various technical solutions related to it and the objectives set by the various participant organisations. It is by no means an exhaustive explication of TC and it only serves as stimulus for the exploration of the regulation development question. Hence, this material will be revisited in the analysis section in order to relate to the research questions set at the outset of this paper.

According to Ross Anderson [2], the idea of TC firstly appears in a paper by Bill Arbaugh, Dave Farber and Jonathan Smith [5]; is related to the work of Markus Kuhn titlted TrustNo1 Processor [32] and the origins of the TC idea date in the early 1970s in a paper written by James Anderson for the USAF [3]

The foundations for TC as we know it today were set by an alliance formed between Compaq, HP, IBM, Intel and Microsoft in October 1999 having as a stated objective to improve “trust and security on computing platforms” [59]. Since its inception, the Trusted Computing Platform Alliance [59] has grown substantially to include now over 150 participating organisations.

Microsoft’s implementation of TC is known as “Palladium” [45] but the name has changed since January 2003 to Next-Generation Secure Computing Base for Windows [45].

As stated by the formal TCPA FAQ site, goals of the TCPA are:
Through the collaboration of hardware, software, communications and technology, vendors drive and implement TCPA specifications for an enhanced H[ard]W[are] and Operating System based trusted computing platform that implements trust into client, server, networking and communication platforms [59].

The key deliverables for the TCPA will be:

TCPA White papers, which describe the specification and how it improves computing.

Specification version 1.1 has been developed by members of the TCPA and published in July 2001.

Define platform specific implementation guidelines.

Provide advocacy for the proper use of TCPA on computing platforms.

The basic operation of TC is based on the principle of creating a set of "several software components implemented on a secured hardware platform", the key component of which is a security operating system, nicknamed "the nexus", which runs at a higher level of security than the rest of the system [51]. This new security level may only be implemented through architectural changes in the level of the processor hardware and operating system, therefore the collaboration between companies like Intel and Microsoft is required for such a level of integration to be achieved. For the same reasons, changes are required at the level of the drivers, the memory controller, chipset and peripheral devices [51]. Another important component of a TC system -at least in its Microsoft implementation- is the Security System Component (SSC), which is "a smart-card-like component that manages the platform certificates and returns keys in exchange for the nexus signature. This is an evolution of the Trusted Platform Module (TPM) and is of pivotal importance as it stores the measurements of the components of the user's systems on which the security evaluation is based [59].

By positioning the security mechanisms at the microprocessor level and integrating the overall computer and operating system architecture so that it supports the TC functions an increased level of security is indeed achieved. What remains debatable whose interests are best served by this "security" and what kind of "trust" is developed [3].

TC claims to provide a secure operating system environment where applications that are TC-certified may communicate with each other. This is primarily achieved through checking the hardware and software components a system comprises of and then comparing these measurements against the policies contained in the particular applications used in the particular transaction [59]. This happens through the Trusted Platform Module (TPM) or Security System Component (SSC).

In that sense TPM is indeed a passive component and is not a Digital Rights Management (DRM) system in itself. However, seeing TPM in isolation does not make much sense and the institutional structure of the TCPA is a good reason for adopting a more holistic approach. The TPM is a measurement mechanism that is well integrated with the TC platform architecture so as to discourage the user from tampering with the applications and allowing the applications to communicate securely with the vendor [3]. The measurements taken by TPM-like devices may only be used within a broader PKI-like (Public Key Infrastructure) setting. The "trust" vocabulary inevitably leads to such solutions. In a PKI scenario the rules as contained in the policies that will govern the behaviour of the software will be set by the Certification Authorities (CA). These are parties that "issue and sign certificates that can be used by an entity (person, website, platform etc) to convey information that can be trusted" [59]. In other words the measurements taken by the TPA -and associate a particular hardware/ software configuration with a particular user- may then be used in order to "certify" the trustworthiness of this user against particular policies contained in the application software through a CA.

Indeed, an insulated -and in that sense secure- communication channel may thus be achieved. However, it is questionable how much can the user negotiate the terms of a policy in order to be certified as appropriate user of a service that will be TC-based. The click-wrap license precedent should not make the users particular optimistic. Indeed, TC is not DRM *per se*. However, such an infrastructure is ideal for any kind of content management systems. Companies involved in the development of TC are also involved in DRM-like technologies (HP and Microsoft to name just two) and a CA infrastructure is an essential condition for Intellectual Property Rights management in digital environments.

Even if the Security and Trust promises of TC are fulfilled –and that is debatable because of a series of side-effects [3], - the type of security that it provides seems to reflect the objectives of the parties that participate in its development. It seems that the *locus* of the control is shifted from the end user to the content owners and software vendors, since these are going to be the drafters and implementers respectively of the policies that will constitute the charter of a TC-enabled world.

6. A Gnutella Narrative

In early March 2000, Justin Frankel and Tom Pepper developed a small Windows program that they called Gnutella. The name is a merge of the GNU license that stands for **G**NU is **N**ot **U**nix and under which the software was to be released and the Nutella chocolate and hazelnut spread produced by the Italian confectioner Ferrero, which it was said that the original developers consumed during the fourteen days that the software was developed. (Oram 2001). The original Gnutella stood both for the software and the protocol upon which it was based and is what we will refer to when we use the term "Gnutella" in this paper.

Although Gnutella is based on the Internet Protocol (IP) it operates in a rather different fashion, as it does not give meaningful and persistent identification to its nodes. As a matter of fact the underlying Internet structure is completely hidden from the Gnutella users. We get to think of the World Wide Web (WWW) as a relatively stable infrastructure, where the information resides on stable "places" such as the web sites. On the contrary, Gnutella creates a virtual infrastructure where the computer literary is the network. There is no backbone infrastructure that the Gnutella user accesses: each computer *is* part of the Network; if there were no Gnutella-based software users then there would not be any network either.

The true peer-to-peer paradigm on which Gnutella is based has as a result the absence of any central points of control. As it is often stated "the client is the server is the network". [48] Gnutella does not rely on any central authority to organize the network or to broker transactions [48]. This means that in order to perform a search, for instance, the only thing you need is to know any arbitrary host to connect with. Once the connection is established the host send another message to all the hosts that are in your immediate vicinity. If a search is to be made, you send a search message to the host and then the host passes the message to all the other hosts until it receives a return message that the requested file has been found. Then you are connected directly to the host that has the requested file and the transfer of the file can happen directly between the two nodes. In the latest versions of Gnutella, each file is assigned a unique ID so that the search is for the ID and not the file itself. It is possible that many hosts have the file with the same ID. The search happens in the same way but the transfer happens simultaneously from many hosts and that increases the speed of the network. Finally, the non fixed nature of the Gnutella network is slightly compromised with the concept of Super-nodes or ultra-peers that allows for certain nodes that have greater computing power or bandwidth to operate as a broker infrastructure for the facilitation of indexing and file transferring. Although that way there is a kind of backbone network created, it is still not a fixed one. This development in a sense follows the pattern that the Internet development has exhibited but we will return to this issue in the analysis section.

7. Gnutella early development phases and the Web sites used as development hubs

Gnutella was initially developed as an application by Justin Frankel and Tom Pepper. They were working for Nullsoft a company that produced the well known music media player Winamp. Nullsoft has been purchased by America On Line AOL since June 1999 and if we are to believe Tom Pepper, "Winamp was developed primarily to play digital music files (…) Gnutella was developed primarily to share recipes." [48]. The original plan was Gnutella to be released after reaching version 1.0 under a GNU license General Public License (PL). Under the GNU General Public License [22] everyone has access to the source code of the software but in case any changes are made there are no proprietary rights over the final product [54].

Because of the legal problems that the Napster service was facing at the time, AOL declared Gnutella an "unauthorised freelance project" and decided to remove it from the Nullsoft site. At the time Gnutella was still in version 0.56. The removal of Gnutella from the Nullsoft web site is attributed to what is often quoted as "the Slahsdot Effect".

The attention that the Nullsoft site attracted resulted in its subsequent removal by the AOL people that owned it. However, the same Slashdot effect was responsible for the rescue of the Gnutella protocol. The original Gnutella application was based on a communication protocol that has been reverse engineered and further developed by an initially small but increasingly growing community of open source developers. The Gnutella communication protocol was reverse engineered by Bryan Mayland, who then posted on a website called www.gnutella.nerdherd.net. The web site does not exist any more but at the time Ian Hall-Beyer and Nathan Moinvaziri created the environment that could sustain the initial phases of such a project. Besides the Nerdherd website, the link to Gnutella's Internet Relay Chat (IRC) Channel *#gnutella*. Gene Kan emphasises the importance of the IRC channel for the development of the protocol as it allowed for the instantaneous interaction between the developers when responses in very short time circles was required.

The Clip2 Distributed Search Systems (DSS) was the first entity to start monitoring the operation and performance of the Gnutella network; it provided a forum for discussion between the Gnutella developers, metrics for the Gnutella network, IP addresses for Gnutella hosts and news on different products, but was terminated after May 2001 [20]. Other forums that were used for the early development stages of the Gnutella protocol were the http://gnutellang.WeGo.com or the http://Gnutelladev.WeGo.com but both URLs are used for other purposes today. According to the Internet Archive log, the former has been

terminated after November 2001 [25] and the latter after having been transformed to the gPulp web site in October 16th 2001 [25], which was "not a working group on Gnutella" according to the gPulp people themselves, has not been changed after August 2001 and was finally terminated after November 2001 as well. Even since January 2001 the problems with the existing developers' forums have been identified and the Clip2 DSS people decided to form another forum, known as the Gnutella Developers Forum (GDF) hosted by Yahoo!. [24] The reason behind the formation of the GDF was the lack of an administrator in the WEGO groups as well as that WEGO required a substantial amount of money for hosting the forums. [24] Another reason behind the migration of developers to the GDF was that other forums hosted by particular Gnutella application vendors were not considered as neutral as they should be.[24] Although the GDF was initiated by clip2 people, it was a conscious –and as proved by the events that followed, the correct- choice to select a platform other than that of the Clip2 web site for the hosting of the GDF. As the GDF founding document states:

We believe placing the forum at a third-party provider gives it the best chance of neutrality and longevity. The "gdf_sysop@yahoo.com" eGroups account owns the GDF eGroup. Because of this, Clip2 has the ability to transfer administration rights to another party, such as a future developer organization. Another reason for using eGroups is the number of useful free tools provided by the service, including a chat system, file repository, link list, database system, poll service, etc. [24]

Clip2 has ceased to exist after may 2001, but the GDF's is still active with the latest additions to the Gnutella protocol having been posted in September 2002. It is important to highlight the existence of other forums related either to the Gnutella protocol itself or to the various Gnutella applications, the www.gnutellaforums.com being the one with the maximum traffic. Another important web site for the development of the Gnutella protocol is the RFC-Gnutella (Request For Comments) Gnutella project, which is hosted by www.sourceforge.com, although it has been virtually inactive since August 2002. The RFC-Gnutella seems to operate at the moment as the most comprehensive repository for Gnutella documents, rather than as a communication medium as it is explicitly stated in the project description section.

8. Re-visiting the Regulation development process

It is rather common in regulatory debates to focus on the content of a particular regulatory instrument and understate the importance of other components, in particular its institutional basis and dialectics that lead to its creation or the specific form (technological, legal, market- or norm- based) in which a regulation is manifested. Lessig has greatly contributed to the emergence of an alternative approach for examining the regulatory phenomenon by highlighting the concept of the four regulation modalities [40] and highlighting Open Source [42] as a process that can provide some minimum transparency, accountability and participation when regulation is contained in technological artefacts.

In both the cases we presented in this paper, there was a conscious attempt to link the process and tools used for the creation of the regulatory instruments to their end-results. Starting from the working assumption that both the TCP and the Gnutella protocol constitute forms of regulation [40] -or at least have regulatory characteristics- [60], [61] the objective was to explore the relationship between the actors involved in the creation of a regulatory instrument and the content of the rules it contains.

In the case of TC, the industrial group that is behind the TCP initiative has clearly managed to inscribe its interests into the various Trusted Computing technologies. In a similar fashion other groups that had an input in the development process, like the content industry or enterprises that would like to use TC as their organisational infrastructure have managed to represent their interests in the TC products and processes. The positioning of TC in direct relation to PKIs also provides indications about the actors that had the opportunity to provide input in the creation of these kinds of technical solutions. On the contrary, the interests of users have been greatly underrepresented and this becomes apparent from the reactions that the TC has caused an instance of which has been Microsoft's position to even change the Palladium name into Next-Generation Secure Computing Base for Windows [45].

Needless to say that this is a first reading of the situation and the data that were collected when this paper was drafted were not sufficient for a fully supported analysis of the situation. Nevertheless, the preliminary data collection provides a strong indication that if an initiative like TC is to be examined from a regulatory perspective the question of participation and openness in the creation of TCPs needs to be addressed. Gartner [51] reports that opening up the platform would increase its

market acceptance but "Microsoft is culturally protective of software, reducing the likelihood that Palladium will be open". Microsoft has actually agreed to open part of the NGSCW for viewing but not changes are allowed to be made under the existing licensing scheme that it proposes [45]. Anecdotal information indicates that there is an effort from other companies participating to the TCPA to open up the development processes as much as possible. The idea of having open-source security solutions is indeed rather old and there are already attempts for open TC solutions [1]. However, there are inherent problems with such an approach: firstly, there is a serious question of how compatible would an attempt to make a TC technology be with the culture of openness that exists in any open source setting. The openness should then be moved into the produced system and that is –at least according to some- inherently incompatible with a TC solution that by definition requires closeness to achieve its objectives. Secondly the TC concept is not just a technical one: it includes a large organisational and procedural part with trusted third parties and CA that will implement particular policies through technical measures. Such systems often are based on a control system where the control is shifted from the user to third entities. If the processes are to change as a result of an open source development that would advocate a different security model, then the system could be more open but is doubtful whether it would by TC.

Analogous are the results we get from reading the Gnutella protocol case. It has been developed in the background of the Napster case and as such both organisationally and technically it advocated a distributed model that would be immune to various forms of legal attacks. The development was open-sourced precisely because the original developers could not work on the protocol being employees of a company with vested interests in IPRs. The choice of a totally decentralised system for indexing and transfer of files has also been influenced by the legal battles related to file-sharing and the relation between the Gnutella development and copyright issues is a recurrent theme in the related developers forums.

Moreover, the various developers from the original duo of Pepper and Franklin to the latest GDF group have used infrastructures that were supportive of the mode of development they have chosen. Especially in the GDF -but in previous forums as well- developers are the principal participants whereas simple users or content owners are not really present. Gnutella has been a very important vehicle for the peer-to-peer community for the development of new ideas and particular products but the user experience that Gnutella clients are providing is not of the level that other protocols are offering. The number of Gnutella users is nothing but a fragment of the overall peer-to-peer users that are mainly FastTrack-KaZaA users. Equally under-represented are the content creators who are loudly silent in all kinds of Gnutella forums.

The absence of both content creators and unsophisticated users from the Gnutella forums is also because of the amount and complexity of the existing discussion groups as well as the self-referential nature that the discussion tend to adopt: as the time advances the participants of the forum tend to continue discussions that have started some time ago an is difficult for an outsider to participate unless she creates her own discussion topic [62].

This first round of data from both cases indicates that the degree of openness in the production stage of the regulation has direct implications for the content that the regulation is going to have. At the same time, the complexity of the phenomena and the interrelated nature of a diverse range of issues indicates that the regulation production should not be viewed in mechanistic terms but through an ecological perspective where regulation is not approached as something that is deterministically produced but organically cultivated.

In terms of its methodological implications, such an approach would entail the use of methodologies that trace the inscription of interests from one actor to another. In this paper we suggested Actor Network Theory as such a tool, but further research is required for identifying a solid methodology for dealing with these issues.

9. Regulation As dialectics

Choosing a particular perspective for the studying regulation allows a different approach to the phenomenon of regulation as well. The two cases under consideration represent two different worldviews about how regulation should be contacted.

The TC view is part of greater trend that has emerged in the late 1990s and views legal regulation as an imperfect creatures that needs to be completed with the technological intervention. The problems of undesired side effects from the introduction of particular regulatory measures has been viewed –at lest by the copyright literature- as an instance of the greater enforcement problem or

as an expression of the inherent problem of law to follow technological development. Technical measures of protection in all kinds of forms tend to be more predictable and in that sense they create trajectories of action that are more determinable and thus provide a sense of increased control.

In this paper we argue that such a conceptualisation of regulation is in a great extend illusionary. Particularly in the case of Copyright Law the recent history of the arms race between copyright regulation and peer-to-peer technologies has shown that an approach that seeks to maximize control is in a great extent unfruitful. Every new legislative initiative, court decision or technical measure of protection was met with a response in terms of licensing, technical solution or norm creation in the community of the users that cancelled the effect of the original action. This is a phenomenon that was described in the beginning of this essay and is repeated with the case of TC.

TC attempts to insulate a realm of activity from the rest of world by creating a more controlled and thus isolated environment. Such an approach follows the traditional model of regulation that sees as its prime objective the limitation and control of contingencies. However, reality seems to be inherently irreducible and forcing a particular course of action as the only allowed one has all sorts of side-effects that create greater problems than the ones they claim to solve. In the case of the TC some of the problems are related to the claims of security and flexibility that seem to be incompatible with each other: in order for a secure environment to be achieved it needs to be predictable as well and predictability may only be achieved if the control lies with centralised or semi-centralised authorities. The fact that numerous users will have intentionalities different from those inscribed in the policies of the CAs will most probably be the cause of massive disruptions and attempts to by-pass the TC systems. That will initiate a new cycle of dissidence with pretty much unexpected consequences in the same way that the copyright-p2p conflict has evolved.

The Gnutella protocol is an indication of a different approach to the regulatory problem. There, the objective is not to create a particular course of action or a secure environment but to open up the possibilities for content dissemination by proliferating instead of controlling contingencies. This becomes apparent in the development process as well as in the end product, which is a distributed and decentralised system. The idea behind peer-to-peer networks is the creation of "distributed systems without any centralised control or hierarchical organization, where the software running on each node is equivalent in functionality" [48].

The concept thus introduced is one that bases regulation in *dialectics* not in *compliance*. The integration that the TC model advocates is illusionary, as it separates the receiver from the producer of regulation in the development phase and thus creates more opportunities for dissidence from the end regulatory product. In addition, TC-based regulation is based on the concept of controlled environments and as such it is more likely not to match with the intentions of users that could have different from the pre-set objectives. Therefore it leads to a fragmentation and increases the Risk of counter-productive regulation. This is an effect that is further intensified by the fact that TC is an infrastructure technology that will be used in massive scale. On the other hand the Gnutella protocol development model is not as anarchic as it firstly seems to be. It attempts to solve the same problem with different means. If the problem is that the consequences do not map on the original regulation, then regulation should be made more open and reflexive. By allowing participation and proliferating possible regulatory avenues the intentions of the receivers of the regulation match more accurately the regulatory product and the integration is more successfully achieved.

The latter is an ecological approach not only in the sense that supports a holistic view of the problem but also in the sense that does not try to violate the complexity of a socio-technical environment. Solutions that seek to transfer control massively and in an unbalanced fashion are deemed to operate disruptively and the first data of all technical measures in place advocate such an argument (e.g. PressPlay vis-à-vis KaZaA). Every techno-legal systems is what Von Foerster [21] describes as a "non-trivial" machine; "it is synthetically determined but not analytically determinable: it is dependent on the past, but cannot be predicted" [58]. This independency is the main feature of the system's autonomy and as Hejl [27] explains "inputs which appear identical to the outside observer do not necessarily have the same internal effect". This essential indeterminacy of the law is what makes it possible to operate in the complexity of a social setting. Introducing all encompassing technical measures will not really solve the problems that legal regulation tries to address.

Of course the solution to the Risk from regulation is not a straightforward one. The Gnutella development process also suffers from lack of representation of interests and in that sense it

resembles the TC development process. Moreover, the Gnutella protocol is an artefact with regulatory properties but with a strong utilitarian nature as well. A development process that puts participation as an absolute principle will not necessarily lead to the best technical result or not even a result that will be the most user-accepted.

Throughout this paper the term “regulation” was indiscriminately used to refer to technological artefacts and laws alike. This was done for practical reasons, but it contains a great danger as the utilitarian nature of technologies that have regulatory features creates a conflict of interests in the way these are produced: should the most technically efficient or the most transparent and participative solution be chosen? How is the development methodology of the Gnutella protocol related to the fact that it has regulatory features? Why do more closed protocols like FastTrack achieve better user acceptability than Gnutella?

This paper finishes with a series of questions, different from those that were set in its beginning but triggered by the answers that were provided to them. The focus in regulatory studies should be on the process of regulation development as much as on the product and such a perspective calls for an understanding of regulation not just as compliance but as dialectics as well. In the same way as the empirical object that triggered this research, the latter is not a set of definite answers but another link in the chain of the regulatory discourse.

[1] Adamantix, Adamantix Frequently Asked questions, available at http://www.adamantix.org/faq.html, 2003

[2] J. P. Anderson, "Computer Security Technology Planning Study", 1972, ESD-TR-73-51, Vol II

[3] R. Anderson Cryptography and Competition Policy: Issues with `Trusted Computing', http://www.cl.cam.ac.uk/~rja14/, 2003

[4] R. Andrews, "Long range planning in environmental and health regulation" Ecology Law Quarterly 20, 1993, 515-582.

[5] B. Arbaugh, D. Farber and J. Smith, ``A Secure and Reliable Bootstrap Architecture", in the proceedings of the IEEE Symposium on Security and Privacy, 1997 pp 65-71.

[6] R. Baldwin, C. Scott, C. Hood C, A Reader on Regulation, Oxford University Press: Oxford,1998

[7] R. Barthes, Mythologies. Trans. Lavers, A. Vintage Press: London. 1957

[8] Y. Benkler, The Battle Over the Institutional Ecosystem in the Digital Environment, 44 Communications of the ACM No.2 84, 2001

[9] Y. Benkler, Coase's Penguin, or Linux and the Nature of the Firm, 112 Yale L.J. (Winter 2002-03)

[10] S. Biegel, Beyond our control? : confronting the limits of our legal system in the age of cyberspace. Cambridge, Mass.: MIT Press.2001

[11] W. E. Bijker, Of bicycles, bakelites, and bulbs : toward a theory of sociotechnical change Inside technology. Cambridge, Mass: MIT Press. 1995

[12] A. Block and F. Scarpitti, Poisoning for profit, William Morrow, New York, 1980

[13] Boland, R. J. "Phenomenology: A Preferred Approach to Research on Information Systems." In Research Methods in Information Systems, ed. E. Mumford et. al., 1985, 193-201. North-Holland: Elsevier Science Publishers.

[14] J Boyle, "A Politics of Intellectual Property: Environmentalism for the Net" 47 Duke L.J. 87 (1997)

[15] J Boyle, "The Second Enclosure Movement and the Construction of the Public Domain" 66 Law and Contemporary Problems 33 (2003)

[16] Callon, M. and B. Latour, Unscrewing the Big Leviathan: how actors macrostructure reality and how sociologists help them to do so, in Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies., K.D. Knorr-Cetina and A.V. Cicourel, Editors. Routledge and Kegan Paul: Boston, Mass. (1981), p. 277-303.

[17] Callon, M., Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay, in Power, Action and belief: a new sociology of knowledge, J. Law, Editor., Routledge & Kegan Paul: London. (1986) p. 196-233.

[18] Callon, M., Actor-Network Theory - the market test, in Actor Network Theory and After, J. Law and J. Hassard, Editors. Blackwell Publishers The Sociological Review: Oxford.(1999)

[19] Clapes L. et al Silican Epics and Binary Bards: Determining the Proper Scope of Copyright

Protection for Computer Programs, 34 UCLA L.Rev. 1493 (1987)

[20] Clip2 DSS. FAQ on founding of Gnutella Developer Forum, Clip2 DSS.(2002)

[21] H. von Foerster, Principles of self organization: in a sociomanagerial context, In H. Ulrich and G J B Probst (eds.) Self-Organisation and Mangement of Social Systems: insights, promises, doubts and questions, Berlin: Springer, 2-24, 1984.

[22] FSF (2003) What is Copyleft by Richard Stallman, available at www.fsf.org/copyleft/copyleft.html

[23] Garfinkel, H. Ethnomethodology. Polity Press: Cambridge.(1967)

[24] GDF (2002) The Gnutella Proposals, available at http://groups.yahoo.com/group/the_gdf/files/Proposals/proposals.htm.

[25] GnutellaNG, (1999) Welcome to gnutellaNG, available at http://web.archive.org/web/20000816001207/http://gnutellang.wego.com/.

[26] P. N. Grabosky "Counterproductive Regulation", International Journal of the Sociology of Law, 23, 1995, 347-369.

[27] P Hejl, Towards a theory of social systems: self-organization and self-maintenance, self-reference and syn-reference, In H. Ulrich and G J B Probst (eds.) Self-Organisation and Mangement of Social Systems: insights, promises, doubts and questions, Berlin: Springer, 60-78, 1984.

[28] B. Hutter and P. Sorensen (Eds), "Business adaptation to legal regulation", Law and Policy 15, 1993, 169-270.

[29] I. Hosein, P. Tsiavos and E. A. Whitley, "Regulating Architecture and Architectures of Regulation: Contributions from Information Systems." International Review of Computing Law and Technology, Volume 17, Number 1, 2003, pp. 85-97.

[30] W. Katz, "Don't do it! An alternative approach to the management of oil spill cleanup", Environmental Progress 13, 1994, M2-3.

[31] J. Kallinikos, The Age of Flexibility, Managing Organizations and Technology.
Academia-Adacta, 2001.

[32] M Kuhn The TrustNo 1 Cryptoprocessor Concept, 1997-04-30

[33] Kuhn T., The Structure of Scientific Revolutions, 3rd Edition, University of Chicago Press: Chicago (1996)

[34] Latour, B. and S. Woolgar, Laboratory Life: the Social Construction of Scientific Facts. Beverly Hills and London: Sage.(1979)

[35] Latour, B., Give Me a Laboratory and I will Raise the World, in Science Observed, K.D. Knorr-Cetina and M.J. Mulkay, Editors. Sage: Beverly Hills.(1983)

[36] Latour, B., Scientific Objects and Legal Objectivity, translated by Alain Pottage, in Making Persons and Things, Cambridge University Press: Cambridge. (2002)

[37] Law, J. After ANT, Complexity, Naming and Topology in Law, J. and Hassard, J. (eds.) Actor Network Theory and After. Blackwell: Oxford.(1999)

[38] Lee, A. S. "Integrating Positivist and Interpretive Approaches to Organizational Research." Organization Science, Volume 2, Number 4, pp. 342-365.(1991)

[39] R. Leone, Who profits? Winners, Losers and Government Regulation, Basic Books, New York,1986.

[40] Lessig, L. "The New Chicago School." Journal of Legal Studies, Volume 27, Number June,pp. 661-691.(1998)

[41] Lessig, L. Code : and other laws of cyberspace. New York, N.Y.: Basic Books.(2000)

[42] Lessig, L. The future of ideas : the fate of the commons in a connected world . 1st ed. New York: Random House.(2001)

[43] Limewire.org, Getting Started,LimeWire LLC. 2003.

[44] LimeWire.org, Project Help for Developers, LimeWire LLC. 2003.

[45] Microsoft, Microsoft "Palladium": A Business Overview, available at http://www.microsoft.com/presspass/features/2002/jul02/0724palladiumwp.asp, 2003

[46] D. McBarnet, "Legitimate rackets: tax evasion, tax avoidance, and the boundaries of legality", Journal of Human Justice 3, 1992, 56-74.

[47] A Murray C Scott "Controlling the New Media: Hybrid Responses to New Forms of Power", Modern Law Review, Vol.65.4, 2002

[48] Oram, A. (ed.) (2001) Peer-to-peer: Harnessing the Benefits of a Disruptive Technology, Cambridge O' Reilly

[49] Pouloudi, A., and Whitley E. A. (2000) Representing Human and Non-human Stakeholders:

On Speaking with Authority in Organizational and Social Perspectives on Information

Technology,. pp. 339-354, Kluwer, Aalborg, Denmark, June 10-12, 2000

[50] M. Power, The Audit Society: Rituals of verification, Oxford University Press, Oxford, 1997

[51] M. Reynolds, Palladium Security's Brave New World, Gartner, 2002

[52] J. Sigler and J. Murphy, Interactive Corporate Compliance, Quorum Books, New York, 1988.

[53] R. E.Stake, "Case Studies." In Strategies of Qualitative Inquiry, ed. Norman K. Denzin and Star, S.L. 'Power, Technology and the Phenomenology of Conventions: On Being Allergic to Onions.' In A Sociology of Monsters: Power Technology and the Modern World, ed. J.Law. Oxford :Basil Blackwell, 26-56. (1991)

[54] R. Stallman, The GNU Operating System and the Free Software Movement in Open Sources: Voices of the Open Source Revolution, 53, 56, (edsChris DiBona et al) (1999)

[55] S. Strum, and B. Latour, The Meanings of Social: From Baboons to Humans in Schubert, Yvonna S. Lincoln, (1998) 86-108. London: Sage Publications.

[56] C. R. Sunstein, After the Rights Revolution: Reconceiving the Regulatory State, Harvard University Press, 1990.

[57] G Teubner, In Corporations, Capital Markets and Business in the Law. Liber Amicorum Richard Buxbaum. Edited by Baums, T.; Hopt, J.; Hom, N. Kluwer, 2000.

[58] G Teubner, Law as an Autopoietic System, The European University Institute Press Series, 1993

[59] Trusted Computing Platform Alliance, TCPA Frequently Asked Questions, 2003, available at http://www.trustedcomputing.org/tcpaasp4/index.asp

[60] Tsiavos P., Whitley E. A., Hosein I. (2003) The Footprint of Regulation: How information Systems are Affecting the Sources of Control in a Global Economy. IFIP (WG 8.2- 9.4) IS Perspectives and Challenges in the Context of Globalization, Athens, Greece, June 15-17.

[61] Tsiavos P., Hosein I., (2003) Beyond Good and Evil: Why open source development for peer-to-peer networks is not necessarily equal to an Open Society, is as imbalanced as Copyright Law and definitely is not going to make you a better person. ECIS 2003, 11th European Conference in Information Systems, New Paradigms in Organizations, Markets and Society, Naples, June 19-21

[62] Tsiavos, Participation and Exclusion in the Gnutella Development Process: Tasting the Fruits of Knowledge The third Social Study of IT Workshop at the London School of Economics, London 24-25 April 2003

[63] E. Warren and G, Marchant "More good than harm". A first principle for environment agencies and reviewing courts", Ecology Law Quarterly 20, 1993, 379-440.

A Framework for Evaluating Digital Rights Management Proposals

Rachna Dhamija
UC Berkeley, SIMS
rachna@sims.berkeley.edu

Fredrik Wallenberg
UC Berkeley, SIMS
fredrik@sims.berkeley.edu

Abstract

In this paper, we analyze the strengths and weaknesses of the various solutions to compensate intellectual property rights holders. Specifically we look at digital rights management (DRM) based systems, extensions to DRM to support fair uses, monitor-and-charge schemes, compulsory licensing schemes and alternative business models.

Our main contribution is to provide a framework from which current and future proposals may be evaluated. In order to realistically evaluate any compensation scheme, we suggest that the following questions are important to ask:

- *Is the proposal technically feasible?*
- *What are the incentives to circumvent legal and technical protections for all parties in the transaction?*
- *What is the burden of monitoring for compliance in the system, and on which parties does this burden fall?*
- *What is the efficiency of the collection and distribution of funds from consumers to rights holders?*
- *What are the impacts on user privacy and fair use?*
- *What is the feasibility of legal enforcement, both domestically and internationally?*

1. Introduction

Over the last few years the debate over protection, or lack thereof, of copyrighted works has flourished. Proposals on how to reimburse the creators of these works range from strict proprietary encryption locks to new business models that rely on revenue streams from ancillary products. Each new proposal points out the shortcomings of previous schemes and highlights the benefits of its own solution. However, no consistent framework exists for analyzing the different solutions.

In this paper, we analyze the strengths and weaknesses of the various solutions. Specifically, we look at DRM based systems, extensions to DRM to support fair uses, monitor-and-charge schemes, compulsory licensing schemes and alternative business models. From this comparison, we extract important dimensions such as technical feasibility, incentives to cheat, burden of monitoring, privacy, and the feasibility of legal enforcement. Our main contribution is to provide a framework from which current and future proposal may be evaluated.

Digital Information as a "Public Good" Economists sometimes refer to certain goods as *public*. This does not imply that they are in the public domain as defined by intellectual property law. Rather, a public good is a product or service that has two properties. First, it is *non-rival*, which simply means that consumption by one person doesn't limit consumption of the next. Second, it is *non-excludable*, implying that once the product exists, the benefit cannot be limited to those that have paid for it.

Ideas and information captured in physical media traditionally fall into some middle ground. While the information itself certainly has the characteristics of a public good, the physical media that it is tied to is rival and excludable. This gives rise to business models involving the sale of physical artifacts whose only value is the embedded information such as books, CDs and DVDs. These business models have taken a serious blow with the introduction of information in digital form combined with communications media such as the Internet. The question at hand is whether or not it is possible to devise a scheme under which money can be transferred from those consuming information goods to the providers of the same.

We use the characteristics of a public good to distinguish between the following classes of proposals to compensate intellectual property rights holders:

- The first approach is to make the product rival. Solutions in this category use DRM copy protection to make sharing of information goods hard (or impossible).
- The second approach is to make the product excludable. This includes watermarking schemes that allow owners to monitor who is using the product in order to charge for its use and to pursue those who don't pay.
- A final approach is to accept that a product is a public good that is non-rival and non-excludable. This cate-

gory of solutions relies on financing through a general collection system (such as levy or tax schemes) or on revenues from alternative business models and voluntary contributions.

The three classes of solutions present different challenges from the standpoint of incentive compatibility. They also differ with respect to cost and enforceability, and who bears the burden of each.

For example, DRM systems that artificially make a product either excludable or rival invite circumvention activities by end users (who realize that, if they could remove the technical barriers, the product is neither excludable nor rival). In compulsory licensing schemes, which tax users independently of their actual consumption, there is less incentive for users to cheat. However this scheme invites another problem. If the disbursement of funds to rights holders is based on the observed consumption of their products, rights holders now have a strong incentive to bias the observed traffic in their favor. With either solution, there is one party that will have to be monitored for cheating unless technical barriers are put in place that cannot be circumvented, an unrealistic assumption.

In section 2, we summarize the general and specific proposals that have been made for compensating rights holders. In section 3, we analyze how the different solutions fare in areas such as technical feasibility, incentives to cheat, burden of monitoring, efficiency, privacy and fair use. Finally, we present our conclusions in section 4.

2. Proposed Solutions

2.1. Creating Rival Goods

Traditional DRM Digital rights management (DRM) systems aim at protecting ownership and copyright of electronic content by restricting what actions an authorized recipient may take with respect to that content. In traditional DRM systems, content is distributed in protected form and relies on a compliant device or secure execution environment to allow the user to make use of the work. In these schemes, the content may be protected through encryption (as is the case with the DVD Copy Control Association Content Scrambling System[1] and cable and satellite transmissions) or through labeling (as is the case with Macrovision[2]) and the Digital Television Broadcast flag [1]. In permissions-based systems, the encrypted content is delivered with a machine-readable license, which specifies the license terms and specific permissions for which a user is authorized to make use of the work (as is the case with eBooks and many audio and video players). For examples, see the Windows Media DRM[3], the RealNetwork Helix DRM[4] and the Apple FairPlay DRM[5].

DRM Extension Proposals One of the strongest criticisms of DRM systems has been their inflexibility in allowing end users to make *fair uses* of works. U.S. Copyright laws give copyright owners the right to prohibit others from making some uses of the work, such as copying, distributing or making a derivative work. However, there are many exceptions to this rule that allow users to legally make further uses of the work, even when such uses are not authorized by the copyright owner (for example, the ability to make private backups, or the ability to make excerpts for commentary or criticism). The term fair use strictly refers to the four factor test given in 17 U.S.C. §107. Here, we use the term more loosely to refer to other exceptions that apply to copyright, e.g., Special Rules for Libraries and Archives (17 U.S.C. §108), narrow exemptions (17 U.S.C. §110), first sale rights, term expiration, public domain, privileges for reverse engineering and backup of computer programs.

In their "Fair Use Infrastructure" proposal, Burk and Cohen examine how fair use can be retained under a DRM system that provides strict access control [2]. The proposal has two components. First they suggest that there is a subset of all possible cases of fair uses that can be well-described and encoded as automatic defaults into DRM systems. For the cases of fair use that cannot be easily encoded in this way, the authors propose that users could make a request for access to the work with a trusted third party. The third party is responsible for determining if the requested use falls under fair use, and if so, it is able to grant access to the encrypted work via a key escrow system (in which keys to decrypt works are deposited by copyright holders).

Mulligan and Burstein propose modifications to Rights Expression Languages (REL), i.e., XrML, in order to better support fair uses [3]. They argue that REL syntax and vocabulary should enable rights holders to express license terms in a way that more closely matches copyright law. Specifically, rights holders should be able to express fair use exceptions, as clearly and easily as they are able to express restrictions on the use of a work. Furthermore, they propose that DRM systems should be designed so that all parties in a rights transaction (both the rights holders and end users, for example) can express their rights through REL. Erick-

[1] http://www.dvdcca.org/css/
[2] http://www.macrovision.com/
[3] http://www.microsoft.com/windows/windowsmedia/drm.aspx
[4] http://www.realnetworks.com/products/drm/
[5] http://www.info.apple.com/usen/musicstore/musicstore.html?topic=music_authorization

son proposes a DRM architecture in the same vein [4]. In his model, users can make rights requests (such as a fair use request) to a third-party licensing authority. Like the other proposals in this section, he argues that a third party licensing authority will be more impartial than the rights holders in deciding whether to grant a use request. Also, like Burk & Cohen, his solution to achieve a closer approximation to fair use depends upon giving users the opportunity to engage in a rights negotiation with rights holders.

2.2. Creating Excludable Goods

In order to provide monitoring and tracking, some DRM schemes rely on watermarking, which embeds a visible or invisible mark in content such as audio, images and computer software. Fingerprinting is similar in concept and usually refers to embedding a unique serial number in content (as opposed to general copyright information). Fingerprinting can also involve the extraction of unique features from a particular piece of work in order to identify it.

There are a number of examples of the use of ex-post monitoring by copyright holders to detect unauthorized uses of their work. For example, some companies have proposed an automatic system to detect unauthorized distribution of images that consists of a watermarking scheme and a web crawler that downloads pictures to check if they contain the watermark (for example see Digimarc MarcSpider[6] image tracking). In their attempts to stifle piracy, the music and motion picture industries are using the services of third parties such as BayTSP[7] and Ranger Online[8] to track down infringing copies without the need for a-priori watermarking. BayTSP claims to have detected 10,000 infringements with a 95% compliance and removal rate resulting from takedown notifications.

In his "ISPs as Digital Retailers" scheme, Sobel proposes to use watermarking and fingerprinting techniques, not only to find copies of work but also to charge consumers directly. In Sobel's proposal, ISPs can license content from copyright holders at wholesale prices and then re-sell the content to their customers, with whom they have an established billing relationship [5]. He suggests that digital fingerprinting (and/or watermarking) could be used by the ISPs to monitor the flow of copyrighted materials over their networks. In order to ensure that the transaction costs associated with negotiation are minimized, the author proposes a statutory license that forces the copyright owner to provide a license, however, it does not regulate the prices that copyright holders may charge.

[6] http://www.digimarc.com
[7] http://www.baytsp.com
[8] http://www.rangerinc.com

2.3. Public Goods

Even among those who agree that digital information should be treated as a public good, there is a wide discrepancy in the solutions proposed with respect to government regulation. At one end of the spectrum, we have proposals that rely on legal intervention, typically in the form of compulsory licensing schemes. At the other end, we have abolitionists who suggest that doing away with intellectual property rights will best allow market solutions to flourish.

Compulsory Licensing One conclusion of accepting that digital information goods have the characteristics of public goods is that creation needs to be subsidized through compulsory licensing policies. In this case, the rights holders are required to license their works at a set rate and under certain conditions. One way to characterize the problem is how to "collect a pool of money from Internet users, and agree on a fair way to divide it among the artists and copyright owners" [6]. In this paper, we use the term "compulsory licensing" as von Lohmann does, to refer to the taxation model that is commonly used for public goods. However, other compulsory licensing models are possible, see for example the Music Online Competition Act of 2001[9] and U.S. Copyright Office's Copyright Arbitration Royalty Panel[10].

Specific compulsory licensing proposals, such as those by Netanel and Fisher, suggest that the money should be collected based on consumption of devices (such as CD and DVD burners), media (blank CDs and DVDs) and services (such as ISP access) [7, 8]. The efficiency of any collection scheme will depend on how close the consumption of products and services that are taxed is to that of the digital goods that are consumed. Funds raised in this manner would then have to be disbursed to the rights holders based on some approximation of the use of their respective products.

Netanel's Non-Commercial Use Levy (NUL) proposal builds upon the basic concept of compulsory licensing by specifying that the license should only be for non-commercial use and should not include all forms of digital goods [7]. Specifically he intends the model to cover "literary works" that are not primarily tools. That is, he expects creative content such as music, movies, text and graphics to be covered, but not computer programs.

To refine the collection mechanism, he proposes that the levy should be imposed upon "commercial providers of all consumer products and services the value of which ... P2P file sharing substantially enhances" [7, page 32]. He further proposes that the NUL should strive to raise as much money as the sales supplanted by P2P sharing.

[9] http://www.house.gov/boucher/moca-page.htm
[10] http://www.copyright.gov/carp/

While the remuneration should ideally be tied closely with the users' aggregate private value of the goods, Netanel acknowledges that such monitoring would imply high transaction costs and privacy costs. He suggests that metering of downloads, streams and uses should be done both at the ISP level and, in some cases, on user devices and be supported by digital fingerprinting and sampling techniques.

Fisher proposes that we replace the current copyright model with a government administered reward system [8]. In order for creators to collect revenue under this system, they are required to register with the copyright office and will receive a unique filename in return (that would allow the work to be tracked). Similar to Netanel, Fisher proposes that the government put a tax on devices and services that are used to access digital entertainment.

The distribution would be determined through the analysis of a number of metrics including surveys and usage data provided by file sharing systems, such as KaZaa. Fisher also recognizes that the amount of compensation needs to differ depending on the type of good, thus the fact that both a new Britney Spears song and the recent Spielberg movie have the same "market share" doesn't mean that they should receive the same remuneration.

Alternate Business Models Boldrin and Levine, among others, believe that copyright (or other intellectual property rights) is unnecessary in order to stimulate the creation of information goods [9]. There have been many calls for the development of new business models that don't require control of the content per se, but where the revenue comes from excludable actions such as showing a movie in a full size theatre or giving live concerts. Further, just as the content drives demand for "performances", it may also provide room for merchandising. However, a number of artists have expressed skepticism about such ancillary revenue streams [10].

Finally there are those that suggest that voluntary payments may work. There are plenty of examples within the shareware industry of products that are made available for free (without any limitations on functionality) and of "contributions" that are sufficient enough to support the developers. Yet, there are examples where the "tips" were insufficient to pay for the development of the product. For example, Stephen King's novel *The Plant* was originally offered under the "tip model" but the model failed to raise sufficient revenue and the book was withdrawn [11].

As an enabler for alternative business models, Creative Commons licenses can be used by authors to indicate that their copyrighted works can be copied and distributed, usually under certain conditions (for example, only with attribution, or only for non-commercial use). Related efforts are The Free Software Foundation's General Public License for software licenses and the Electronic Frontier Foundation's Open Audio License for digital sound recordings.

3. Discussion

In this section, we compare and contrast the proposals for compensating rights holders along the following dimesions: technical feasibility, incentives to cheat, burden of monitoring, efficiency of collection and distribution of funds, privacy, fair use, feasibility of legal enforcement and flexibility.

Technical Feasibility A number of security researchers have commented on the technical futility of copy protection and DRM [12, 13, 14]. One cryptographer has stated that DRM approaches will never be successful because "all digital copy protection schemes can be broken and, once they are the breaks will be distributed" [12]. Other security researchers are more optimistic that DRM models can have more success by focusing on risk management and the ability to adapt to security compromises [15].

All of the attacks that traditional DRM system are vulnerable to also apply to the DRM extension proposals. While the DRM extension proposals aim to provide better support for fair use, they acknowledge that it is impossible to create a DRM system that will allow all fair uses. A fundamental technical challenge is how to create exceptions that are flexible enough to allow legitimate fair uses, but not so flexible that they can be exploited as loopholes by infringers.

The centralized key escrow scheme proposed by Burk & Cohen would be an enormous technical undertaking. Many of the technical criticisms of key escrow systems in general also apply to this proposal [16]. For example, a centralized key repository creates a very high value target and introduces many new vulnerabilities and threats regarding the improper disclosure of keys. Due to the large number of users and copyrighted works, such a system would be extraordinarily complex to administer and extremely costly to implement.

Sobel's monitoring and charging scheme is also infeasible to implement securely. It is simply too easy for users to alter digital fingerprints and watermarks, especially given that they have a strong incentive to do so. For example, users may easily be able to remove the mark, or to place a watermark from one work into another [17, 18, 19]. Another simple attack is for users to encrypt their files to prevent detection of the watermark. It would be difficult (or impossible) for the ISP to differentiate between legitimately encrypted content and encrypted copyrighted content without banning

encryption entirely.

The main technical challenge in the compulsory licensing schemes is how to track digital copies of content. Fisher suggests two approaches to tackle this problem [8, Ch. 6, p.6]. The first is for the creators to imbed digital watermarks into copies of their work, which could then be tracked and replicated in each copy of the original work. Unfortunately, as discussed above, the watermarking approach is subject to a number of attacks that may make such a plan infeasible.

Fisher's second approach relies upon the existence of a centralized registration system, which would require artists to register their work in exchange for a unique serial number for that work. Again, Fisher does not provide any details for how the serial number would be tracked. One possibility is to embed the serial number as a watermark, but this is subject to the problems discussed above. Any centralized registration service of this type will be an enormous technical undertaking.

Even if watermarking systems were impossible to defeat, both the monitor-and-charge schemes and the compulsory licensing schemes are subject to "distributed" cheating attacks, where the consumption of a good is artificially inflated across a large number of (real or illegitimate) users. These types of attacks are challenging to detect, and any usage monitoring or sampling scheme must be designed with this in mind.

Incentives to Cheat DRM schemes that tie payment directly to consumption inherently give users a larger incentive to circumvent copy-protection or monitoring devices in order to avoid payment. Also, onerous security restrictions on DRM-wrapped content make compliance less attractive, given the availability of unrestricted content [20]. DRM schemes that allow a certain threshold of private use copying may be perceived as more "fair" and may therefore enjoy wider adoption and higher compliance rates. Apple's recently announced Fairplay DRM allows more private use copies to be made compared to other DRM schemes [21].

DRM extension proposals, such as those proposed by Mulligan & Burstein and Erickson, may increase user compliance, because users are now able to engage in fair use without circumvention. In the Burk & Cohen model, the incentive for the users to comply is that those who fail to obtain access via the escrow agent would be subject to prosecution for circumventing technical measures. Under Burk & Cohen the incentive for rights holders to deposit keys with the escrow agent is that they would otherwise be unable to invoke legal protection against circumvention. If they choose not to escrow their works, users would be given a "right to hack" as a substitute for access to the work via escrow keys. One problem with these proposals is that rights holders currently have no economic incentive to express fair use terms or to allow user negotiation in DRM enforced licenses, as proposed by Mulligan & Burstein and Erickson, absent a change in the law [22].

In a monitor-and-charge model, such as that proposed by Sobel, users have the incentive to under-report consumption to avoid payment. Irreputable rights holders also have an incentive to push unrequested information to the user to increase their revenue. In fact, spammers could potentially construe their spam as copyrighted material and be paid for it.

In compulsory licensing schemes, which tax users independently of their actual consumption, there is less incentive for users to cheat (users may still have an incentive to skew the reporting, either because they would like to favor certain artists or because they are concerned about the tracking of their specific consumption.) This is one of the prime motivators for any compulsory licensing scheme as outlined in section 2.3. Fisher recognizes that, under these types of schemes, it is now rights holders that have the incentive to cheat, or engage in "ballot stuffing." This ballot stuffing is very similar to the spamming problem that monitor-and-charge schemes are subject to. However, in the compulsory license case, cheating will be more challenging to detect since end users are not being directly charged and therefore have less incentive to complain about products they have not consumed. As discussed above, cheating that artificially inflates the consumption of a good over a large number of (real or illegitimate) users, will be hard to detect in both schemes.

Burden of Monitoring Under DRM-based systems, the burden of monitoring user compliance falls on the rights holders.

In the monitor-and-charge environment, as proposed by Sobel, the monitoring burden falls on the ISP. The Sobel proposal correctly identifies that the ISP is in the best position to monitor individual users [23]. What the author fails to acknowledge is that the ability to successfully monitor depends on the effectiveness of the tracking mechanism (which isn't very effective) and the user's incentive to circumvent the protection (which is high). The security responsibility for the former falls entirely on the shoulders of the rights holders who insert the watermarks. The incentive for users to cheat will depend on price and usage restrictions, both of which are also determined by the rights holder. With that in mind it isn't surprising that the ISPs are less than thrilled about the proposal. Furthermore, while ISPs do bill their individual users, the complexity implied by this system (where everyone can be a copyright holder and consumer) would result in "the worlds most complicated billing sys-

tem" (Sarah Deutsch, council for Verizon, at the UC Berkeley Law and Technology of DRM conference in February 2003).[11]

In compulsory licensing schemes, the burden of monitoring falls on the government to ensure that rights holders do not "game" the system. One concern is that this vests a large amount of power and discretion over creative culture with a government agency. In this case, public monitoring will be critical to ensure that the rules established are fair. For example, what criteria will be used to determine who is authorized to register as a legitimate artist with the copyright office? How do we ensure that organizations, such as RIAA and MPAA, who have large lobbying power, will not tilt such a system to their advantage and to the disadvantage of smaller independent artists? How will the agency respond to the opposition that is sure to arise over the funding of politically unpopular art? For example, The South Carolina House of Representatives passed a bill to renounce the Dixie Chicks for their "unpatriotic" criticism of President G.W. Bush prior to the second Gulf War [24]. As another example, we cite the legal battles that arose over "standards of decency" in government funding of artists by the U.S. National Endowment for the Arts [25].

Efficiency of Collection and Distribution of Funds In the case of DRM-based solutions, the revenue received by the rights holders is a direct function of the value assigned by the users (specifically, we know that the user's value is at least as high as the price). In the case of a monitor-and-charge system like Sobel's, the fact that the payment is made well after the decision to consume tends to have an impact on purchasing behavior.

Under compulsory licensing schemes, direct ties between funds collected and true consumption cannot, per definition, be established. The precision of the estimate of what is consumed can be improved by tying the collection of funds to the consumption of (non-public) goods that have a high correlation with the public information good. In this context, Netanel's proposal fares better than most compulsory licensing schemes, because funds are raised based on levies on (non-public) goods whose values are tightly linked to the digital work.

A further effect of decoupling collection and consumption is that "sales" can no longer be used to determine how funds are disbursed. Rather, some metric must be used to estimate the aggregate value of the consumption of a particular work. When users no longer "vote with their wallets", even perfect observation of every copy acquired (through downloading or otherwise) is unlikely to yield a perfect estimate, because consumption patterns are changed. That is, when the marginal cost is lowered (to near-zero), not only will the users consume more, but the mix will most likely change since the user will consume goods with a value to them lower than the previous cost but higher than the new marginal cost. The further away from the individual actions that the sampling is done (by, for example, monitoring the traffic on the backbone), the worse the precision becomes.

Privacy There is an inherent tension between the goals of DRM copyright-enforcement and the privacy goals of end users. Rights enforcement technologies may compromise user privacy through the restrictions they place on users, by tracking and monitoring users and their usage patterns, and also through the data that is collected by network operators [26, 27, 28].

The DRM extension proposals also present a serious challenge to user privacy by creating a centralized database of user requests for access. The most privacy-optimal solution would fulfill fair use requests (or under Burk & Cohen, release escrow keys to applicants) without retaining any personally identifying records. However in order to prevent abuse and prevent would-be infringers from exploiting such a system, it is likely that records will be kept. The danger is that copyright industries will demand the ability to match keys with identities so that pirated materials can be linked to the suspected infringers. In their proposal, Burk & Cohen recommend that identifying information be released only pursuant to a court order and only on a showing of actual piracy. This issue is currently being tested in U.S. courts. As of this writing, ISPs are required to turn over subscriber information to copyright holders upon "reasonable suspicion of a violation", a much lower standard than that suggested by the authors. See The U.S. District Court (DC) opinion in RIAA v. Verizon, holding that the issuance of a subpoena by a Clerk of the District Court to obtain the identity of an anonymous peer to peer infringer from his ISP does not violate either the First Amendment of the Constitution, or the justiciability requirements of Article III [29]. Furthermore, Burk & Cohen acknowledge that in any scheme where users must request access, even the most stringent system of privacy protections for fair uses is likely to chill some lawful uses. For more treatment of the chilling effect, see [30, 26].

Under a monitor-and-charge scheme, such as that proposed by Sobel, the impact on privacy should be expected to be much higher than Sobel acknowledges since *all* copyrighted traffic to and from an *identified* user will be monitored. The observations on the chilling effects of monitoring apply here as well.

In general, compulsory licensing models would require less precise monitoring of individual activities than DRM-

[11] `http://mindjack.com/relay/archive/2003_02_01_index.shtml`

based or monitor-and-charge models. Even if file usage is monitored at the same level, metering for the purpose of redistribution of funds does not require identification of the end user. Neither Netanel nor Fisher provide any details of their watermarking schemes, if they will embed any personally identifying information, for example. But presumably, such a scheme could depend on aggregate sampling and would not require ISPs or P2P operators to record how much any particular individual has downloaded or uploaded a particular file (however, ISPs and P2P operators could choose to record this data, as it would be commercially valuable). Netanel's approach is more problematic than Fisher's from the perspective of privacy since he, in an effort to improve precision, suggests that usage should be tracked not only on the network (down to the individual user level) but also on the user's devices.

Fair Use The legal definition of what constitutes a fair use is ambiguous in U.S. copyright law, and only a court can determine with authority whether a particular use is a fair use. It is unlikely, therefore, that we will be able to build DRM systems that can reason about what uses are fair in the foreseeable future.

One solution to resolve the tension between rights holders desires for copy controls and users desires to make fair uses is to encode special exception cases of fair use into the DRM system. The exceptions must be broad enough to be useful, but cannot be so broad as to allow infringement to occur. Regardless of how broad the encoded exceptions are, there is always the spectre of future fair uses that have yet to be thought of [31].

All of the DRM extension proposals in section 2.1 enable the introduction of a third party decision maker. The aim is to approximate case-by-case determinations, which cannot be emulated by fair use defaults alone. In particular, the proposal by Burk & Cohen strives to encode some flexibility to handle borderline cases as well as new uses.

Solutions that rely on ex-post monitoring have an inherently better chance of supporting fair uses. The difficulty inherent in Sobel's monitor-and-charge model is that the ISPs must determine which uses are fair use. The likely result would be a very strict interpretation by the ISPs resulting in severe limitations on fair use, because the ISPs would face legal liability for infringing uses and they do not benefit financially from fair use copies.

Compulsory licensing schemes will also inherently more easily allow fair uses to be made. However, one criticism of the compulsory licensing schemes is that the cost that is borne by users will account for all uses, whether they are licensed and authorized uses or whether they are unauthorized fair uses.

Feasibility of Legal Enforcement Even if we can track and identify infringers from a technical standpoint we still have to worry about whether or not we can enforce laws effectively.

Currently, with DRM based systems, there is a very large set of possible entities that the rights holders may pursue when infringement is discovered, including individual users, providers of circumvention tools, operators of file sharing networks and ISPs. Little changes under the monitor-and-charge proposal. However, the rights holder can now prosecute one entity, the ISPs, for failing to enforce copyright laws, and it will be up to the ISPs to pursue everyone else.

The major difference is seen under a compulsory licensing scheme. In this case, the rights holders only have to concern themselves a smaller subset of users that infringe the license, such as unauthorized commercial users under Netanel's NUL. The entity in charge of disbursement of royalties will have to monitor the rights holders, but it benefits from having the means to punish them by virtue of withholding funds [8, Ch.6, p.29]. Penalizing individual users who do not explicitly act on behalf of a rights holder, but are simply trying to distort the system, will be much harder.

Another obvious problem stems from international users (and content). DRM-based systems can function well internationally, however, prosecuting infringing uses in other countries presents a challenge. For example, in the DeCSS case, Jon Johansen was acquitted by Norwegian courts [32]. Similarly, the monitor-and-charge and compulsory licensing schemes only contemplate raising money from U.S. users, without any concern for how foreign users would be induced to pay or how the U.S. would handle payments to foreign entities.

Currently, rights holders are seeking broader legal anti-circumvention legislation in the U.S., European Union and other countries to protect DRM-based business models [33]. It remains to be seen whether rights holders would also seek legislative protection in the case of compulsory licensing (for example, in the form of prohibiting users and network operators from circumventing watermarking schemes).

Flexibility There is another important consideration when choosing a mix of market based solutions and government intervention. In general, government mandated solutions foreclose development of many new solutions. Adopting a compulsory licensing solution will lock us into a specific solution and may halt the evolution of new business models for distributing digital goods.

4. Conclusions

Based on the analysis in the previous section, it is quite clear that there are many tradeoffs to consider when evaluating proposals to compensate intellectual property rights holders. In order to realistically evaluate any compensation scheme, we suggest that the following framework of questions be applied:

- Is the proposal technically feasible? No proposed technical protection measures are strong enough to sustain a determined attack. Only in combination with models where the incentives to circumvent are limited, can technical solutions succeed.
- What is the feasibility of legal enforcement, both domestically and internationally? It is easy for researchers and market actors to forget that a solution that requires significant government intervention and enforcement is inherently bound to the confines of country boundaries and international treaties. Reducing the reliability on legal enforcement may improve the chance of international effectiveness.
- What are the incentives to circumvent legal and technical protection for all parties in the transaction? The incentives for users to cheat will depend on the price per copy of digital works and the restrictions that are placed on usage. Decoupling revenue collection from the act of copying may reduce incentives for the user to cheat. Privacy concerns may also affect these incentives.
- How efficient is the proposed solution? Efficiency is a concern in the collection and disbursement of funds from consumers to rights holders. It is also a concern when analyzing the burden of monitoring for compliance and where that responsibility is placed.
- What are the impacts on user privacy and fair use? Privacy concerns frequently run counter to desires for economic efficiency. Therefore, any proposed solutions must acknowledge that there is a trade-off to be made. Fair use is important on its social merits alone, however, a broader adoption of fair and private uses will also serve to reduce user incentives to circumvent.
- How flexible is the solution? Some proposals will, if adopted, foreclose other types of solutions. It could be that it is better to support an inferior solution now, but one that leaves us with an opportunity to adapt other, better solutions in the future.

References

[1] "Final Report of the Co-Chairs of the Broadcast Protection Discussion Subgroup to the Copy Protection Technical Working Group," June 3 2002. [Online]. Available: http://www.cptwg.org

[2] D. L. Burk and J. E. Cohen, "Fair Use Infrastructure For Copyright Management Systems," *Harvard Journal of Law & Technology*, vol. 15, no. 1, Fall 2001. [Online]. Available: http://jolt.law.harvard.edu/articles/pdf/15HarvJLTech041.pdf

[3] D. K. Mulligan and A. Burstein, "Implementing Copyright Limitations in Rights Expression Languages," in *2002 ACM Workshop on Digital Rights Management*, Washington DC, November 18 2002. [Online]. Available: http://crypto.stanford.edu/DRM2002/mulligan_burstein_acm_drm_2002.doc

[4] J. S. Erickson, "Fair use, DRM, and trusted computing," *Communications of the ACM*, vol. 46, no. 4, pp. 34–39, April 2003.

[5] L. S. Sobel, "DRM as an Enabler of Business Models: ISPs as Digital Retailers," *18 Berkeley Technology Law Journal*, forthcoming 2003. [Online]. Available: https://www.law.berkeley.edu/institutes/bclt/drm/papers/sobel-drm-btlj2%003.pdf

[6] F. von Lohmann, *The Daily Princetonian*, April 14 2003. [Online]. Available: http://www.dailyprincetonian.com/archives/2003/04/14/opinion/7930.shtml

[7] N. W. Netanel, "Impose a Noncommercial Use Levy to Allow Free P2P File Sharing," *U of Texas Law, Public Law Research Paper*, no. 44, November 15 2002. [Online]. Available: http://www.utexas.edu/law/faculty/nnetanel/Levies_chapter.pdf

[8] W. Fisher, "PROMISES TO KEEP: Technology, Law, and the Future of Entertainment," available at http://cyber.law.harvard.edu/people/tfisher/PTKprivate.htm, last revised: March 22 2003. [Online]. Available: http://cyber.law.harvard.edu/people/tfisher/PTKprivate.htm

[9] M. Boldrin and D. K. Levine, "The Case Against Intellectual Property," *American Economic Review*, no. 92, pp. 209–212, 2002. [Online]. Available: http://www.dklevine.com/papers/intellectual.pdf

[10] J. Toomey and K. Thomson, "Virtual Tip Jar or Charity Case?" available at http://www.futureofmusic.org/articles/arsfinal.cfm, Future of Music Coalition, last revised: May 20 2000. [Online]. Available: http://www.futureofmusic.org/articles/arsfinal.cfm

[11] *Associated Press*, December 8 2000. [Online]. Available: http://www.cnn.com/2000/books/news/12/08/stephen.king.ap/

[12] B. Schnier, "The futility of copy prevention," *Cryptogram*, May 15 2001.

[13] P. Biddle, P. England, M. Peinado, and B. Willman, "The Darknet and the Future of Content Distribution," in *2002 ACM Workshop on Digital Rights Management*, Washington DC, 18 november 2002. [Online]. Available: http://www.cse.ogi.edu/~krasic/cse585/darknet.pdf

[14] E. W. Felten, "A Skeptical View of DRM and Fair Use," *Communications of the ACM*, vol. 46, no. 4, pp. 57–59, April 2003.

[15] P. Kocher, J. Jaffe, B. Jun, C. Laren, and N. Lawson, "Self protecting digital content," CRI Content Security Research Initiative, Tech. Rep., 2003.

[16] H. Abelson, R. Anderson, S. Bellovin, J. Benaloh, M. Blaze, W. Diffie, J. Gilmore, P. Neumann, R. Rivest, J. Schiller, and B. Schneier, "The risks of key recovery, key escrow, and trusted third-party encryption," Counterpane Labs, Tech. Rep., 1998. [Online]. Available: http://www.counterpane.com/key-escrow.html

[17] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Proceedings of Information Hiding Workshop*, 1998.

[18] S. A. Craver, A. Perrig, and F. A. P. Petitcolas, "Robustness of copyright marking systems," in *Information hiding techniques for steganography and digital watermarking*, S. Katzenbeisser and F. A. P. Petitcolas, Eds. Artech House Books, January 2000, ch. 7.

[19] S. A. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. S. Wallach, D. Dean, and E. W. Felten, "Reading Between the Lines: Lessons from the SDMI Challenge," in *Proc. of 10th USENIX Security Symposium*, Washington, D.C., August 13–17 2001. [Online]. Available: http://www.usenix.org/events/sec01/craver.pdf

[20] F. Hill Slowinski, "What Consumers Want in Digital Rights Management," AAP and ALA, Tech. Rep., March 2003. [Online]. Available: http://www.publishers.org/press/pdf/DRMWhitePaper.pdf

[21] "Apple Computer pressrelease," June 23 2003. [Online]. Available: http://www.apple.com/pr/library/2003/jun/23itunes.html

[22] B. L. Fox and B. A. LaMacchia, "Encouraging recognition of fair uses in DRM systems," *Communications of the ACM*, vol. 46, no. 4, pp. 61–63, April 2003.

[23] R. Andersson, "Why information security is hard," Univesity of Cambridge Computer Laboratory, Tech. Rep., 2001. [Online]. Available: http://www.cl.cam.ac.uk/ftp/users/rja14/econ.pdf

[24] "House resolution 3818," South Carolina General Assembly 115th Session, 2003–2004, Sponsors: Rep. Ceips, Adopted March 19 2003. [Online]. Available: http://www.lpitr.state.sc.us/sess115_2003-2004/bills/3818.htm

[25] "NATIONAL ENDOWMENT FOR THE ARTS v. Karen FINLEY," No. 97–371. Supreme Court of the United States, Argued March 31, 1998. Decided June 25, 1998. (524 U.S. 569).

[26] J. E. Cohen, "DRM and Privacy," *Communications of the ACM*, vol. 46, no. 4, pp. 47–49, April 2003. [Online]. Available: http://www.law.georgetown.edu/faculty/jec/CommACMdrm.pdf

[27] J. Cohen, "A right to read anonymously: A closer look at 'copyright management' in cyberspace," *28 Connecticut Law Review*, no. 981, 1996.

[28] J. Feigenbaum, M. J. Freedman, T. Sander, and A. Shostack, "Privacy engineering for digital rights management systems," *Digital Rights Management Workshop 2001*, pp. 76–105, 2001.

[29] "RECORDING INDUSTRY ASSOCIATION OF AMERICA v. VERIZON INTERNET SERVICES," Case 03-ms-0040, (D.D.C. 2003), Memorandum Opinion issued April 24, 2003. [Online]. Available: http://www.dcd.uscourts.gov/03-ms-0040.pdf

[30] "EFF Open Letter to Universities," Electronic Frontier Foundation, November 6 2002. [Online]. Available: http://www.epic.org/privacy/student/p2pletter.html

[31] F. von Lohmann, "Reconciling DRM and Fair Use: Preserving Future Fair Uses?" in *"Fair Use by Design?" Workshop*. 12th Computers, Freedom & Privacy Conference, April 16 2002. [Online]. Available: http://www.cfp2002.org/fairuse/lohmann.pdf

[32] "Den offentlige påtalemyndighet mot Jon Lech Johansen," OSLO TINGRETT, January 7 2003. [Online]. Available: http://www.domstol.no/archive/Oslotingrett/Nye%20avgjorelser/DVD-jon.doc

[33] P. Samuelson, "Digital Rights Management {and, or, vs.} the Law," *Communications of the ACM*, vol. 46, no. 4, pp. 41–45, April 2003.

Standardization in Digital Rights Management
Trends and Recommendations

Abstract

Digital Rights Management (DRM) is one of the most heavily debated technologies currently. Despised by consumers who fear for their right to fair use, e.g., enabling back-up copies for personal use, it seems to be considered a necessity for a profitable content business in the digital age by many content providers. Its benefits, drawbacks and even implications to society itself are fiercely debated among and between consumer advocates, media power houses, governments, consumer electronics (CE) industry, IT vendors, service providers, and individual consumers alike. Technology standardization, on the other hand, is a process characterized by reaching industry consensus. DRM seems to be a perfect case study for standardization. In this paper, we analyze the effects of standardization of DRM. We first take a look at standardization in general, its purpose and functions, and its relation to patents. We then discuss a particular case of DRM standardization by using the example of the Open Mobile Alliance (OMA).We set out to illustrate that the technology to enable a successful deployment of Digital Rights Management for real-world implementations is best developed in an open standardization forum.

Note: The views and opinions presented in this article are those of the authors and not necessarily of the organizations that employ them.

1. Introduction

Digital Rights Management (DRM) is one of the most heavily debated technologies currently. Several attempts are being undertaken to introduce DRM technology into mainstream products. Despised by consumers who fear for their right to fair use, e.g., enabling back-up copies for personal use, it seems to be considered a necessity for a profitable content business in the digital age by many content providers. Its benefits, drawbacks and even implications to society itself are fiercely debated among and between consumer advocates, media power houses, governments, consumer electronics (CE) industry, IT vendors, service providers, and individual consumers alike. The very nature of DRM and the conflicting opinions of consumers and content providers surrounding it make it an extremely difficult topic to constructively discuss, let alone agreeing on.

Technology standardization, on the other hand, is a process characterized by reaching industry consensus. A significant number of interested parties with varying backgrounds collaborate in standardization in order to define technologies that serve the interests of the entire group. In short, technology standardization is a consensus driven activity for the common good.

DRM seems to be a perfect case study for standardization. In this paper, we analyze the effects of standardization of DRM. We first take a look at standardization in general, its purpose and functions, and its relation to patents. We then discuss a particular case of DRM standardization by using the example of the Open Mobile Alliance (OMA).

In the OMA, a multitude of different stakeholders such as operators, handset manufacturers, technology and content providers work together to meet the requirements from participants of the various content businesses.

We set out to illustrate that the technology to enable a successful deployment of Digital Rights Management for real-world implementations is best developed in an open standardization forum.

2. Standardization

2.1 Framework and Definition

Standardization itself is a well known concept. Ever since industrialization began, standards of one form or another came into existence. With the ability of producing goods in great quantities came the possibility of tapping into huge markets with decreasing costs. Standards at the very early stages were often of de-facto nature, i.e., created by market force through a single or few stakeholders. With a growing number of parties being able to meet the demands of the markets, joint standardization became a more open means of further growing the market and then sharing the benefits among multiple suppliers. Standardization, both open and de-facto, has retained its importance also in today's high tech markets (Shapiro/Varian 1998). While open standards are used to jointly grow markets, they also enable a multi-vendor environment which in turn can be employed to limit the extent to which a dominant player in one market can exert its superior position to break into and control new markets.

Despite its long history, no single definition of standardization has been adopted. One could state that the notion of standardization itself has resisted 'standardization'. It is for that reason that we present different definitions of standardization.

Germon defines standards from a socio-economic perspective as a construct that results from reasoned, collective choice and enables agreement on solutions of recurrent problems. It can be understood as striking a balance between requirements of the involved parties, the technological possibilities and associated costs of producers, and constraints imposed by governments for the benefit of society in general. (Germon 1986)

From a technical point of view, an industry standard represents a set of specifications, to which all elements of products, processes, formats or procedures under its jurisdiction must conform. The process of standardization is the pursuit of this conformity, with the objective of increasing the efficiency of economic activity. (Tassey 2000)

According to a recent definition by the EU, open standards must be consensus-based – involving all stakeholders, including consumer organization representatives – publicly available, transparently agreed, and commercially exploitable on a fair and non-discriminatory basis. The development of standards must therefore take the public interest into account, while standards themselves can play an important role in supporting public policy, and in providing tools for industry to meet regulatory requirements, or take account of public interest issues. (CEN/ISSS 2003)

As can easily be seen, the definitions above are related by similar underlying concepts yet they have divergent characteristics when it comes to the exact scope of standardization.

In this paper, we define *technology standardization* as a process taking into account requirements from multiple stakeholders in the value chain of the market for which the technology is determined resulting in a set of technical specifications potentially accompanied by IP licensing requirements enabling real-world implementations.

Open technology standardization extends the above definition by requiring stakeholders from the entire value chain to be able to jointly and equally collaborate in scoping, defining, developing and governing technical specifications enabling real-world multi-vendor implementations as well as conducting interoperability testing of implementations based on the specifications which are made publicly available.

The success of a standard, either de-facto or open, is ultimately measured by its interoperable adoption of stakeholders and its penetration in relevant markets. Achieving interoperability within a standard eliminates levels of complexity in implementing limited or partial standards. With interoperability among system components, such a market retains advantage of diversification at the component level, but also achieves the efficiency advantages of interoperability.

2.2 Purpose and Function

The importance of technology standards has risen for several reasons. An especially significant role in the area of high-tech standards is played by an ever faster development and replacement of technology paired with the constantly growing complexity of products entering mainstream markets.

Over a technology's life cycle standardization can affect economic efficiency – both positively and negatively. Several competing standards – either locally or industry segment-specific – can coexist for some time, but will be resulting in complaints about inefficiency. In mobile networks, e.g., for GSM, more coordinated efforts were undertaken in order to gain first mover advantages especially in the EU, which resulted in technology leadership in the EU compared to the US.

The function of standards and their purposes can partially be derived from the above definitions. Standards can be perceived as serving several purposes. The following characteristics describe the functions of standards (Tassey 2000):

- *Quality and reliability*: specify acceptable performance and behavior such as functional levels, security, robustness, scalability

- *Information*: provide common languages such as engineering information, dictionaries, describing and testing, even product attributes

- *Compatibility and interoperability*: specify properties that a product must have in order to work with complimentary products within a system. This can be achieved through standardized interfaces between components and protocols

- *Variety reduction:* standards limit the choice to attain economies of scale. This applies to data formats, Meta data, algorithms, and architectures. Naturally, with high economies of scale involved, the involved companies tend to grow to large companies in this process

As already mentioned above, standards are used to jointly grow markets for whose shares the participating parties compete later on. Open standards enabling multi-vendor implementations can also be used as a tool to limiting the extent to which a dominant player can exert its superior position to break into and control new markets.

The more distributed the participants in the market, the more critical to technological innovation are open systems. Open standardization creates multi-lateral governance, thus promoting a multi-vendor environment by preventing a single company from changing the standard to render its competitors' products incompatible. The advantage of open governance can only be stifled by patents.

2.3 Innovations and Patents

Today, patents are an integral part of technology creation and development. Introduced centuries ago aimed at spurring innovation and information sharing, they are an important tool to protect intellectual property. They reward those who made the investment in R&D (Research and Development) ultimately leading to new ideas and technology. However, opinions on the benefit and usefulness of patents are split. There are two diverging schools of thought.

- The first group believes that patents stifle competition. The process of applying and finally being granted a patent can be lengthy and costly, especially for the budget of smaller companies. Big corporations usually hold the biggest patent portfolios. IBM, for example, has been leading the list of companies with most patents granted per year (IBM 2002).

- The second group advocates the innovation fostering aspect of patenting novel ideas and inventions. IBM, for example, is granted a high number of patents not due to being a big corporation, but because, every year it invests a significant amount of their resources into R&D.

Products based on new ideas that are protected by patents usually reach the consumer faster than those for which the manufacturer has no assurance that he will be faced by imitator competition soon after. As such, patents form an integral part of assuring any company that it will be able to recover its investment into R&D by selling products based on the results of that R&D activity. Without assurance of return of investment (ROI), companies might not make this investment in the first place.

The impact of patents on standards depends on the nature of the resulting standard, i.e., whether it is proprietary or open.

In the case of a single company trying to establish a proprietary product as the de facto standard, patents can be used to hinder competition by not licensing the patent to the manufacturer of a competing product, whether it is proprietary or according to an open standard. However, it is seldom the case that a single company holds all essential patents to a technology. Thus, it is unlikely that a single company suffocates all competition on the grounds of patents because it could be subject to the same practices by another company, resulting in a lose-lose situation. Large corporations sometimes form strategic relationships and agree on cross licensing of patents in their respective portfolios to create win-win solutions where the participating companies are able to enter the market and compete, e.g., with technical features of their products.

Open standards bodies often require participating companies to declare their intellectual property that relates to the technology being standardized. This provides the advantage of all companies being mutually aware of the patents held by other companies participating in the standardization process.

3. Digital Rights Management

3.1 Standardization - Perspectives

DRM is a very dynamic technology that is still in its infancy in terms of market penetration. While first patents in the field of DRM date back to the late 80s, the first standardization efforts in the field of DRM were started about 10 years later with initiatives such as the Secure Digital Music Initiative (SDMI 2000). Today, many standardization efforts related to DRM can be found. Lyon (2002) enumerates in his quick reference list of organizations and standards for DRM more than 60 efforts. In the past, this has lead to market segmentation in those areas and to confusion along the value chain.

The main reason for this segmentation can be found in the fact that requirements for DRM standardization vary across distribution channels and end devices. E.g. patient information has different security requirements than entertainment content. Additionally, also content providers from verticals like games, music, film or publishing have different views on the requirements to DRM to enable their respective businesses (Buhse/Wenzel 2003).

Still, DRM is a fascinating case study of standardization. It involves at its broadest consumer adoption, complex technological processes, varying requirements from a multitude of players in the value chain, while at the same time carefully balancing consumer experience and security requirements. Digital rights management and standardization thereof affect several parties with different benefits.

From the *content provider* perspective, which refers to the rights holder as well as to the distributor, standardization allows for the existence of several technology providers. With a broad supplier selection the technology costs for critical components are lower when compared to a market dominated by a monopolistic provider. Also switching costs are lowered and one-time hosting and packaging costs are lower compared to increased content-related costs for several non-standardized providers, while performance is optimized. The protected content market is still very immature while different business models are still being

explored. In this situation, the flexibility provided through open standards where components can be replaced as the innovation progresses seems to be the beneficial approach for content providers. An overall consumer demand aggregation will also lead to network effects and increasing returns for protected content. Still content providers fear negative lock-in effects of any single dominant, proprietary DRM technology supplier.

Looking at the *DRM supplier* perspective, standards in DRM can create bigger markets by earlier consumer adoption based on rapid technology penetration. Provided open standards are in place, it allows for continuous technology upgrades on both sides of standardized interfaces and thereby creating an innovation-friendly environment. In case of a de-facto standard in DRM, it might result in one dominant technology provider, while other providers will be pushed into market niches and potentially vanish over time.

From the *hardware manufacturer* perspective providing client devices (PCs, mobile phones, set-top boxes, etc.), standardization lowers the manufacturing costs and risk by adverting lock-in to a single technology provider. The requirement for interoperability testing in a multi-vendor environment is a small price to pay compared to the market not taking off altogether or leaving it open to proprietary technology vendors. Ultimately manufacturers benefit from substantial economies of scale in production fostered by adoption of a single (that is standardized) DRM technology.

The consumer ultimately benefits from an increased selection of valuable content previously not having been available for purchase as electronic media. Additionally, interoperability between different device categories adds to the positive end user experience and the ease of use by being able to legally consume and share protected content with a number of different devices.

Different approaches can be applied to standardization of DRM.

- Only the interfaces between different components in the back-end, on clients and between these two are specified. This leaves actual design and implementation of the internal functioning of these components up to individual manufacturers.
- Not only the interfaces, but also the behaviour of the different components themselves is specified. In DRM, this is, for example, the protocols between clients and back-end used to acquire content and rights, the format of the secure package that protects content, and the rights governing the usage of content.
- Not only the interfaces and the behaviour of different components, but also their exact internal implementation is specified.

De-facto standards based on proprietary technology are usually of the third kind since actual implementations must be available for manufacturers of clients and operators of back-ends to put a working system in place. Often, standardization bodies adopt one of the former two approaches. This yields situations where individual suppliers develop their own components that interoperate via the standardized interfaces. In section 3.2, we will have a closer look at a standardization forum following the second approach.

Moreover, in earlier markets, as can be observed with Internet-based DRM starting in 1998 and with mobile DRM starting in 2002, companies offer turnkey or end-to-end solutions where proprietary interfaces link components. In these cases, limited price competition through lock-in situations can be observed.

An effective design of an interface standard does not affect the design of the component itself. It provides open systems, allowing multiple proprietary component designs to coexist. With regard to DRM, these closed, proprietary components gain importance when it comes to security as encryption keys and other secrets have to be hidden within those components. Still, innovation can happen, allowing components from different parties working together and even the substitution of more advanced components as they become available over time. This greatly reduces the risk of obsolesce of the entire system also when it comes to security threats.

3.2 Open Mobile Alliance DRM

The Open Mobile Alliance (OMA) was formed in June 2002 through consolidation of the Open Mobile Architecture initiative and the WAP Forum. Since then, the Location Interoperability Forum (LIF), SyncML, MMS Interoperability Group (MMS-IOP), Wireless Village, and the Mobile Gaming Interoperability Forum (MGIF) have integrated into the OMA. The OMA counts more than 300 companies as its members (OMA 2003). Members of the OMA include operators such as 3, AT&T Wireless, NTT Docomo, Orange, T-Mobile, Vodafone, hand set manufacturers such as Motorola, Nokia, Samsung, Siemens, Sony-Ericsson, and technology providers such as Ericsson, IBM, Microsoft, Philips, Real Networks, Sony, Sun, DRM providers such as Digital World Services, Lockstream, SDC, and content providers such as Disney and others.

The OMA is uniquely positioned to develop an open standard for Digital Rights Management. It enjoys the participation of a multitude of players in the value chain, many of which are key players in a flourishing content market (Hartung 2003). Already in 2001, the sale of content in the mobile world in Europe was more than double of that in the wireless world (Jupiter 2002).

The Open Mobile Alliance has already released a set of three specifications constituting the world's first DRM standard targeted at mobile devices. This first release, commonly referred to as *OMA DRM release 1*, defines multiple components of a DRM system (Hartung 2003). These components comprise

- the secure format through which content in the OMA DRM system is protected
- rights according to which content may be rendered by client devices
- protocols for transferring content and rights from network servers to client devices

The approach taken by the OMA makes it an instance of the latter of the two approaches described in section 3. It not only specifies the interfaces but also goes so far to define the behaviour of components themselves. As such, DRM as standardized by the OMA provides the advantages of open standardization (section 2) while at the same time enabling manufacturers of clients and operators of back-end services to immediately deploy a system based on this standard.

The OMA also provides many of the functions of open standards such as enabling market growth, compatibility and interoperability (see section 2.2). Stakeholders from the entire value chain coming together in the OMA jointly grow the global market based on an open standard framework permitting the efficient and reliable development and deployment of applications and services in a multi-vendor environment (OMA 2003a). The DRM developed

by the OMA benefits from these functions of open standardization that are provided by the OMA.

The DRM architecture defined by the OMA enables super distribution of DRM protected content combining viral distribution of content known in a peer-to-peer fashion, yet retaining full control for content owners to allow and disallow consumption of the distributed content. This architecture explicitly allows for both centralized deployment, where there is a strong association between presentation server and download server, as well as decentralized deployment where there is a relatively low level of integration between presentation and download servers. The functionality enables the implementation of confirmed and reliable, and thus billable, transactions between a server entity (Presentation Server, Download Server) and a client device. The functionality allows any type of content to be delivered over any type of bearer to applications residing on clients independent of the operating system, thus fully conforming to the principles of the OMA (OMA 2003b).

Through its rigid IPR policy, the OMA fully acknowledges the importance of patents. The IPR policy of the OMA is based on reasonable and non-discriminatory terms (RAND). It thus protects each member company's continued investment into R&D by ensuring proper licensing of patents for those member companies whose technology becomes part of a standard. At the same time, it ensures fair licensing of patents to its members in order to provide a leveled playing field in which one member cannot refuse licensing its IPR in order to stifle competition. Furthermore, it provides assurance to participating companies through the requirement for member companies to declare essential IPR that they are aware of regarding the technology being standardized.

The success of standardizing DRM in the OMA gains further credibility through the consolidation that has already taken place in mobile standardization efforts. Before June 2002, there were, among others, the WAP Forum and 3GPP. Since the consolidation of the WAP Forum and the Open Mobile Architecture initiative into the OMA, the interests of many players with respect to DRM have come together in the OMA. Also, the 3rd Generation Partnership Project (3GPP) have input their requirements for DRM to the OMA further consolidating the efforts for an openly governed DRM standard.

Therefore, OMA can be considered as a good example for the consolidation of DRM standardization within a specific industry. Additionally, OMA tries to establish liaisons with other related standardization efforts in order to create synergies and in order to bring all value chain partners on board, including content companies from different verticals and from respective consumer groups

4. Discussion and Conclusions

Although various technologies for DRM have existed for quite some time now, it is at a relatively early stage in its life cycle. Not a single one of the proprietary solutions available to date has managed to establish itself as the de-facto standard for DRM in the market place. It could be argued that the market window has not opened up earlier and is currently about to provide the opportunity for a technology to separate itself from the rest of the field to become the de-facto DRM standard. While this is likely to have contributed to the current state of DRM, we argue that no single technology has emerged as the dominant DRM system due to the lack of an openly conducted standardization effort investing the time and resources in the development of a DRM standard.

The standardization of DRM is of particular interest since the flurry of high-tech start-ups creating a myriad of patents along the way. While many of these companies might be gone by now, the patents still exist somewhere, most likely as part of the patent portfolios of the companies that bought these start-ups. The patents generated by the start-ups, might very well be used by their new owners to prevent competitors from entering their market with DRM enabled products. The irony is that DRM – the technology aiming to protect intellectual property – might very well be hindered from taking-of by the intellectual property protecting the technology itself. Furthermore, DRM is a technology that effects a large number of stakeholders in the content business value chain without whose participation any DRM effort is doomed to failure. Especially, the perception that DRM has in the eye of the consumer make it a very difficult technology to introduce to the market.

We have demonstrated that the Open Mobile Alliance provides many of the advantages inherent to the joint development of technology through open standards. Moreover, the DRM effort conducted by the OMA is in the unique position to capitalize on the benefits that its vast range of member companies throughout the entire value chain contribute. In addition, the mobile market, already flourishing and surpassing that of the wireline Internet (Jupiter 2002), proves to be the ideal catalyst for the successful take-off of a commercially deployed real-world implementation of DRM. With the arrival of high bandwidth wireless connectivity, the promise of new services comes one step closer to reality. Content providers, device manufacturers, operators, IT vendors and consumers alike, will not be able to benefit from this new opportunity without the proper content to give life to these services. Whether it is a ringing tone, the latest in mobile gaming, today's number one hit in the charts, or a video clip of the decisive moment in a sports match, the content, and thus the great new services themselves, are unlikely to materialize without the proper insurances for all players in the value chain on their return of investment. The Open Mobile Alliance Digital Rights Management effort is well positioned to provide the protection for this very content.

As mentioned above, there are many standardization efforts for DRM across different industries. Ironically, this market segmentation has brought more DRM-related standardization efforts than DRM technology providers. In order to build on the promise open standardization of DRM provides, the authors strongly recommend all industry participants to

1. work towards consolidation within their industries
2. create liaisons with other such consolidated efforts
3. motivate all value chain participants to provide input to the respective standards

Ultimately, this will contribute to establishing a global DRM infrastructure in an open multi-vendor environment in which all stakeholders have their interests represented.

Bibliography

Buhse, W. / Wenzel, A. (2003): Creating a framework for Business Models for Digital Content - Mobile Music as Case Study, in: Becker, E. / Buhse, W. / Günnewig, D. / Rump, N. (eds.): Digital Rights Management. Technological, Economic, Legal and Public Aspects, Berlin u.a. (in Druck), S. 271-287.

David, P. A. and Steinmueller, E. W. (1994) “Economics of Compatibility Standards and Competition in Telecommunication Networks”, *Information Economics and Policy*, Vol 6 (December), 217-241

Germon, C. (1986) “La normalisation, cle d’un nouvel essor”, Paris.OECD

Hartung, F. (2003) Mobile DRM, in: Becker, E. / Buhse, W. / Günnewig, D. / Rump, N. (eds.): Digital Rights Management. Technological, Economic, Legal and Public Aspects. Berlin u.a. (in print), S. 138-149.

Lyon , G. E. (2002) “A Quick-Reference List of Organizations and Standards for Digital Rights Management”, National Institute of Standards and Technology, Gaithersburg

IBM (2002) IBM News. IBM tops U.S. patent list. Available from http://www.ibm.com/news/us/2003/01/131.html [Accessed 16 Mar 2003].

Jupiter (2002) Paid Content in Europe – Using Mobile and Alternative Payments for Incremental Revenues”. Jupiter. Entertainment & Media. Volume 1

OMA (2003a) Open Mobile Alliance. Overview. Available from http://www.openmobilealliance.com/docs/OMA%20public%20overview.pdf [Accessed 16 Mar 2003]

OMA (2003b) Open Mobile Alliance. Overview. Available from http://www.openmobilealliance.com/overview.asp [Accessed 16 Mar 2003].

SDMI (2000) SDMI Portable Device Specification Part 1, Version 1.0. Document Nr. pdwg99070802, New York.

Shapiro, C. and Varian, H. (1999) Information Rules: A Strategic Guide to the Network Economy, Boston.

Tassey, G. (2000) Standardization in Technology-Based Markets, Research Policy, Vol 20, p. 587-602.

MMS Content Copyright Protection using Watermarking

Luís Moura Silva[1,2], Nuno Carvalho[2], Paulo Germano[2], Paulo Reis[3], Amâncio Santos[1], Paulo Carvalho[1]

(1) CISUC – DEI, University of Coimbra, Polo II, 3030 Coimbra, Portugal
(2) WIT–Software, IPN, Av. Pedro Nunes, 3030-199 Coimbra, Portugal
(3) Ericsson, S.A., Lisboa, Portugal

Email for contact: luis@wit-software.com

Abstract:

This paper presents a DRM protection mechanism for Mobile Multimedia (MMS) content. The proposed scheme makes use of watermarking technology for copyright protection and requires the installation of a Proxy-Server inside the infrastructure of the Mobile Operator. It offers clear advantages over other approaches and can be easily adopted for the OMA-DRM standard. This paper explains in some detail the software infrastructure of our solution and the watermarking techniques that have been particularly devised for the size and limitations of MMS Images. The proposed DRM scheme is currently being trialed with a Mobile Operator.

Keywords: DRM, Digital watermarking, Multimedia Messaging, MMS.

1. Introduction

With the introduction of MMS technology [1] there are now new business opportunities for content providers to distribute their content in the mobile market. The predictions say that premium MMS content is expected to produce global revenues of €31 Billion by 2006 [2], so there are high-expectations in this business. However, there is a huge flaw in this business model for MMS: the lack of content-protection mechanisms. For instance, a mobile user is able to download a branded MMS image and pay for it. Then he is also able to forward this image to several other users that do not pay any extra cent for the premium content. This peer-to-peer distribution is a major concern for content-providers.

It is clear that will be necessary to provide some DRM solution to avoid the free distribution of branded content. Content-providers will not feel very motivated to format their content to MMS format if they don't have their share in the revenue. For this reason, several top-leading content providers are asking for content-protection schemes in mobile networks that would be able to protect, validate and avoid the peer-to-peer forwarding of premium content among mobile subscribers.

The Digital Rights Management (DRM) is a terminology generally used for a wide range of technologies aimed to protect the copyrights of media content. In the mobile environment, the DRM is the key feature for all parties involved in the content value chain from content providers, mobile operators and end-users. DRM solutions can be provided both by the handset vendors or the mobile operators.

A DRM solution targeted to the mobile device seems to be the optimal approach. This approach exploits the fact that copyrighted content is already at the mobile device and provides an inbuilt protection mechanism that prevents users from illegally forwarding the purchased MMS contents by Email, MMS, Bluetooth or iRDA. This solution is usually known as "*forward-lock*". It will only succeed if it will be fully adopted by all the major handset vendors.

The Open Mobile Alliance (OMA), founded in June 2002 by several companies (mobile operators, device & network suppliers, information technology companies and content providers) was created to be the center of mobile service standardization in order to help the creation of interoperable services across countries, operators and mobile terminals. Recognizing the need for DRM in mobile environments, the OMA has defined a set of open standards for Digital Rights Management. These standards will enable the protected delivery of MMS objects by allowing content providers and Value Added Services Providers to earn their right revenue.

The current OMA-DRM [3] standard provides three methods for protecting contents in mobile networks:

- **Forward-Lock:** provides a standardized way to prevent users from forwarding media objects. The media objects are delivered without including usage rights. When an object is received the user will not be able to forward it to other mobile phones;
- **Combined delivery:** enables content providers to package and deliver media objects together with copyright information. The handset DRM agent ensures that objects can only be used according to some defined rules;
- **Separate delivery:** allows media objects to be delivered separately from usage rights information. Before delivery, media objects are converted into DRM content format. This process includes symmetric encryption of the content, making it inaccessible without the content decryption key. This method provides more security and allows the superdistribution of media objects among mobile devices.

A DRM solution should not imply complex registration procedures and the mobile user should not be aware of the DRM solution. The transparency of the mechanism can be achieved in two ways: or there is an installation of a DRM agent in the mobile device or there is a centralized DRM solution provided by the mobile operator.

At the moment, there are already some content-protection solutions targeted to the mobile devices. They require the installation of a Symbian application into the mobile phone. However, these solutions seem to be hardly accepted by the market since they have some drawbacks:

- The multimedia objects are based on proprietary formats and therefore, they must rely on specific data viewers leading to software update constraints;
- These solutions are restricted to specific mobile device capabilities (e.g. support the installation of Symbian applications);
- The installation of a DRM agent into the mobile phones represents a huge headache for the mobile operator distribution department.

A DRM solution targeted to the mobile operator requires the installation of a DRM platform within the mobile network infrastructure. The DRM software platform will provide mechanisms for protecting and checking copyrighted information on the MMS messages. This solution can guarantee the compatibility with the existing phones in the market and with future phones, since it does not require the installation of a DRM agent into the mobile devices. However, it is not a perfect solution since it does not prevent users to forward content by Email, Bluetooth or iRDA.

The OMA-DRM standard is being adopted by some handset manufacturers and will be supported in their products in the near future. Currently, it does not represent a DRM solution to the existing MMS terminals that are being distributed in the market. Therefore, a DRM solution targeted to the mobile operator is currently the most appropriate solution. Based on the technological evolution and market demands, we have developed a DRM solution for MMS content that uses a centralized approach and makes use of digital watermarking technology.

The rest of the paper is organized as follows: section 2 describes the architecture of our solution; section 3 presents the content protection platform and the support for superdistribution. Section 4 explains the inner details about the watermarking algorithms that have been particularly devised for MMS content. Section 5 presents some performance results and section 6 concludes the paper.

2. Architecture of our DRM Solution

Presently, the MMS terminals do not support the OMA-DRM standard and the process of upgrading the MMS terminals with such functionality is quite expensive and time-consuming. Therefore, the mobile operators need a DRM solution that can guarantee compatibility with the existing and the future MMS terminals.

The DRM solution that was developed by WIT-Software is targeted to mobile operators and does not require any specific software installed on the mobile devices. Therefore, they are handset independent and are easy to deploy in the market. It does not require either any direct change to the MMC: it is only necessary to install DRM Proxy server located between some reference points within the Multimedia Messaging Service Environment (MMSE). This DRM Proxy makes use of digital

watermarking techniques to protect the premium MMS contents. The reference points in the MMSE are based on the 3GPP MMS interfaces [4]:

- **MM3 and MM7 interfaces:** The MM3 interface supports the flow of information between the MMC and External Servers. It uses the existing SMTP protocol. The MM7 interface supports the flow of information between the MMC and the MMS VAS Applications. The DRM Proxy server provides mechanisms for automatically protecting the image content sent by the MMS VAS Applications to the Multimedia Messaging Centre (MMC).
- **MM1 and MM4 interfaces:** The MM1 interface supports the flow of information between the MMS terminals and the MMC. The MM4 interface supports the flow of information between MMCs located at different MMSEs. The DRM Proxy server provides mechanisms for intercepting and checking for copyright information on the MMS media content, sent from a mobile phone to the MMC.

2.1 DRM Proxy in the MM3 and MM7 Interfaces

As shown in Figure 1 the DRM Proxy server is located between the MMS VAS Applications and the MMC server. The DRM Proxy server is responsible for: (i) receiving the premium MMS sent by the VAS Applications; (ii) protecting the attached media content using digital watermarking techniques; (iii) and relaying the protected MMS content to the MMC.

The MM3 interface is based on the SMTP protocol. The protection of MMS content on the MM3 interface is based on the originator address from the email message. A database server stores information containing a set of rules to apply according to the message sender. The rule associates an originator email address to specific information such as the VASP name and copyright policies. Every time the DRM Proxy server receives a premium MMS it retrieves the originator email address. If it matches a rule entry, the attached media content is automatically protected.

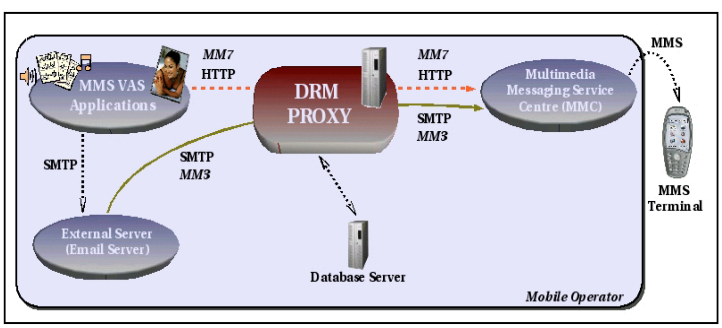


Figure 1: DRM Proxy Architecture (MM3 and MM7 interfaces)

The MM7 interface is based on SOAP over HTTP. The protection of MMS content on the MM7 interface is based on the VASP and VAS identifiers provided in the SOAP envelopes. The DRM Proxy server stores the MMS content on a database server in order to provide content superdistribution.

2.2 DRM Proxy in the MM1 and MM4 Interfaces

As shown in Figure 2 the DRM Proxy server is located between the MMS Terminal and the MMC. It is also located between MMCs of different mobile operators. The DRM Proxy server perform the following actions: (i) it receives the MMS messages sent by the mobile devices (MMS terminal) and the foreign MMCs; (ii) it verifies the attached media objects for copyright information; (iii) and relays the MMS messages to the MMC.

The MMS messages sent to the MM1 and MM4 interfaces are intercepted by the DRM Proxy server that verifies the attached MMS objects in order to see if they are protected or not. If any MMS media object is copyrighted and cannot be relayed, the originator address will be notified by an SMS or MMS messages. On the other hand, the recipient address may receive a WAP Push message redirecting the user to a web site where the media object can be downloaded and purchased. This feature is targeted for content superdistribution.

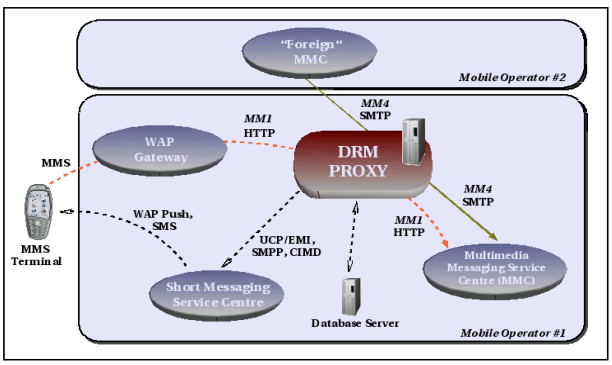


Figure 2: DRM Proxy Architecture (MM1 and MM4 interfaces)

The support for content superdistribution requires a connection to the Short Messaging Service Centre in order to send the SMS and WAP Push [5] messages.

This Proxy-based DRM solution has some advantages and drawbacks. They are presented in Table 1.

Advantages	Drawbacks
• It does not require the acceptance of the MMC Vendor neither the handset manufacturer; • It does not require any upgrade the existing MMS VAS Applications in order to provide a DRM solution. It uses an on-the-fly content protection mechanism for all parties involved on the MMSE; • It does not require the installation of a DRM agent on the MMS terminal. It is mobile phone independent and therefore can guarantee the compatibility with all existing and future MMS terminals; • It is easy to upgrade for supporting other media objects such as audio and video formats, in future versions; • It provides mechanisms for content superdistribution.	• It requires the installation of a new server in the MMSE. Although this is an off-the-shelf hardware server there is some operational impact on the MMSE; • Does not prevent users from illegally copying the purchased contents via Email, Bluetooth or iRDA. There is no solution for this issue unless there is any forward-lock scheme installed on the mobile phone.

Table 1: Proxy-based DRM Solution (advantages and drawbacks).

3. Content Protection Platform

Our DRM platform product is called PAMM. It provides two distinct content protection approaches:

- **Content Protection Management Platform (CPMP):** provides a lightweight Web-based back-office interface for protecting and verifying media content for copyrighted information;
- **DRM Application Program Interface (API):** provides a programming library for protecting and verifying media content for copyrighted information. The DRM-API is targeted to deploy a DRM solution for third-party vendors.

As shown in Figure 3, the CPMP and the DRM-API are located at the application layer and request the content protection and verification mechanisms to the DRM Policy Manager Server. The DRM Policy Manager Server is the middle-layer that allows the intercommunication between the application layer and the PAMM software core modules.

The flow of information between the application layer and the middle-layer is supported by several formatting languages and transport protocols such as SOAP over HTTP and Java RMI.

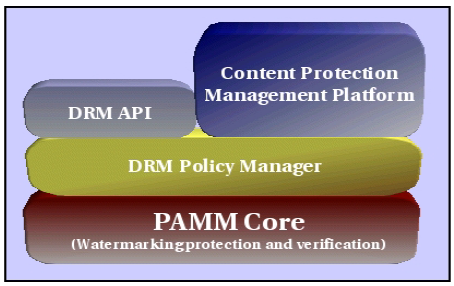


Figure 3: Layers of the Content Protection Platform

3.1 Content Protection Management Platform

The Content Protection Management Platform (CPMP) is a lightweight platform with a Web-based back-office interface aimed to protect and manage the MMS media content. Every copyrighted object includes additional information such as the author's name, creation date and object DRM policies. It provides the following DRM policies:

(i) prevents the media objects from being forwarded;

(ii) specifies the number of times the object can be forwarded;

(iii) allows the forwarding by decreasing the image's quality or applying a visible watermark.

The Content Protection Management Platform is mainly targeted for Content Providers that want to sell their premium MMS contents by using an ASP model with the mobile operators.

3.2 DRM Application Program Interface (API)

The DRM-API is a package that contains a programming library for content protection. It includes an easy-to-use interface able to protect and verify MMS media content for copyrighted information. The DRM-API is mainly targeted to the MMS Application developers and Content Providers. The MMS Application developers are able to directly use this API to protect the premium MMS before their submission to the network. The Content Providers are able to integrate the DRM-API into their existing content management solutions.

3.3 Support for Superdistribution

The PAMM platform provides a set of mechanisms for content superdistribution. The platform sends a WAP Push message to the recipient mobile phone every time it intercepts a copyrighted MMS message. The WAP Push message contains a URL where the user can preview and purchase the copyrighted content sent to him. This superdistribution option is a way to spread the content through the users who would not normally have found it. Figure 4 represents the main idea.

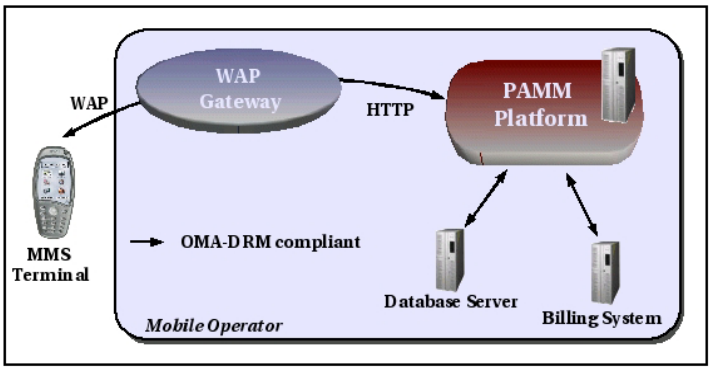


Figure 4: Content Superdistribution

In order to charge the user for downloading copyrighted content, there should be some integration with the billing system. The user can be charged by three distinct approaches: (i) generation of CDRs for off-line billing; (ii) submission of invisible SMS-MT billing messages; (iii) or the access to a third-party Billing API provided by the mobile operator.

The download of the MMS objects from the PAMM platform is initiated by the client. Therefore, the PAMM platform can detect the handset capabilities. If a handset is equipped with a OMA DRM agent the purchasing of the media content is protected by the OMA-DRM [6] standard.

4. Digital Watermarking for MMS

The protection of MMS content on the PAMM platform is based on digital watermarking techniques. These techniques are in fact the main core of the PAMM platform since they provide the required mechanisms for protecting and verifying MMS messages. The digital watermarking algorithms have been specifically devised by an expert research group from the University of Coimbra that aimed to develop an optimized package of watermarking algorithms to protect MMS media content.

To ensure the security requirements, permanent protection of MMS content is needed. Beyond simply granting digital licenses to authorized users, DRM systems should keep restrictions of the content usage rights even after the content being delivered to the end user. In order to ensure these features, many DRM systems are based on various cryptographic solutions. However, there are some drawbacks with this approach since they require some support from the MMS clients installed at the mobile phones. Watermarking does not have this drawback, since it does not change the content format.

4.1 Introduction to Digital Watermarking

Digital watermarking is a data manipulation technique that uses some data redundancy to store copyright information within the content itself. Depending on the type of data, there are several options for digital watermarking. Regardless the methodology, the most common properties [7] are:

- **Perceptual invisibility (often called transparency):** the modifications caused by watermark embedding should not degrade the perceived data quality;
- **Trustworthy detection:** watermarks should represent a trustworthy proof of ownership;
- **Robustness:** digital data can undergo a great deal of different modifications that deliberately (piracy attacks) or not (compression, filtering, resizing) affect the embedded watermark. A watermark should be detectable up to the point that the data quality remains within acceptable limits. The most robust watermarking strategies rely on the correlation between the inserted and the detected watermarks.

The protection of MMS content has some additional challenges, namely:

- **Small image size:** the image size of the MMS images is much smaller than Internet content. Therefore, it is harder to hide copyright information into the content itself;
- **Image transformation:** image resampling and image transcoding are some of the operations automatically used by some MMS terminals and MMC systems. At the moment, there are some MMS terminals providing auto-resize operations to the MMS content. Other operations, such as compression and filtering, had a special concern into the PAMM platform in order to provide a robust MMS watermarking solution.

Ideally, a watermark should be detectable up to the point that the host data quality remains within acceptable limits. Another important aspect for a watermarking algorithm is blind detection, i.e., the ability for detection without access to the original media (not-watermarked).

In the case of MMS content, the watermarking algorithms should code a reasonably number of bits into the watermark, in order to allow a computationally affordable identification of the content provider. This excludes the usage of highly robust watermarking strategies which rely on statistical measures, such as correlation [8] or the first order expectation [9], between the inserted and the detected watermark masks. These algorithms are not applicable to copyright protection of MMS content, for the following reasons: (i) they require the unmarked cover media; (ii) they require one statistical test to be performed per possible watermark, i.e., possible copyright owner and content; (iii) statistical testing is highly unreliable when applied to the limited size of the MMS multimedia objects.

The most probable transformation to the MMS images are those introduced by the system, both due to the limited communication bandwidth and the limited visualization capabilities of terminal equipments. To prevent large MMS messages, MMC servers usually limit the maximum size of MMS images. Each image larger than a predefined value is usually automatically resized by the MMC gateway. Other resize operations may be introduced by terminal equipment during storage and message forward operations. Terminal equipments may also introduce other types of data distortion such as compression (usually a small compression is applied) and colour depth reduction.

The most damaging operations under the described scenario are the geometrical resize distortions, which may introduce considerable smoothing and aliasing, depending on the interpolation method implemented during re-sampling. Some watermarking methods that are resilient to geometrical attacks were reported in recent papers. These methods can be divided into three categories: (i) template embedding based methods, (ii) invariant transform based methods and (ii) invariant features based methods. In template embedding-based methods a known synchronization template is embedded into the image along the watermark [10][11]. Other template embedding methods rely on autocorrelation peaks detection due to watermark replication [12]. These methods tend to be computationally very intensive while reducing the image fidelity as well as the watermark capacity. Linear geometrical image transformation resilience has also been designed using the Fourier-Mellin transformation [13]. However it suffers from several implementation difficulties mainly due its computational complexity and the required unstable log-polar mapping. Kim et al. [14] and Lin et al. [15] introduced other invariant methods with respect to linear geometrical distortions based on the Radon transformation. This method is for zero-bit watermarking and it is not straightforwardly extensible for multi-bit watermarking. Finally, feature based watermarking methods [16][17] rely on the extraction of invariant image features, which tend to be computationally very demanding.

In our MMS-oriented digital watermarking package (see Figure 5) we propose a low complexity watermarking scheme that provides a high level of robustness. To avoid multiple watermark detection, we include some id information within the image (typically the copyright owner and content identifiers). A noise visibility function (NVF) for texture masking enables to transparently embed the encoded watermark into the cover data. Due to the scaling property of the Fourier Transform, this task is performed in the frequency domain in perceptual relevant regions and using local features. Resilience to scale changes is obtained by inserting the watermark using a canonical scale. At detection, the image is scaled back to this canonical scale. To avoid cut-off frequencies of the interpolation filters, a set of possible frequencies, depending on the original image size, are applied to embed and to extract the watermark.

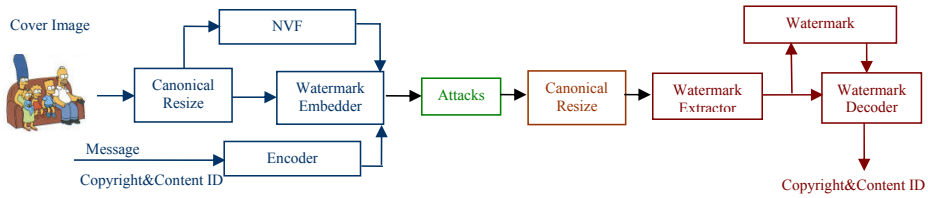


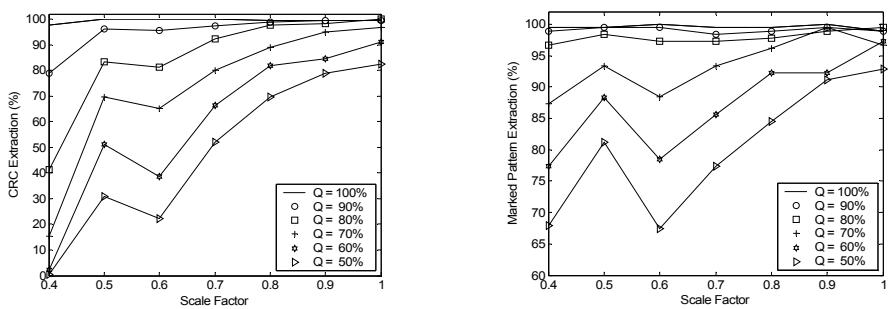
Figure 5: Proposed solution for MMS Watermarking.

Mobile MMS terminal equipments have limited visualization capabilities, both in spatial and in chromatic resolution. On the other hand, due to bandwidth limitations, MMS images are required to be smaller than a predefined size threshold. Usually the resize operation is carried out by the system's gateway, i.e., if the smallest physical dimension of an image is greater than a canonical value, then the image is proportionally resized. Otherwise the image is not changed. Image resize operations may also be introduced automatically by the terminal equipments during message forwarding.

For lack of space we do not present in this paper the details about the watermarking algorithm. We refer the interested reader to [18].

5. Some Performance Results

In Figure 6 we present some performance results that have been obtained with the described watermarking strategy using 183 photographic images obtained from the Internet [19]. Figure 7 depicts the Lena test image.

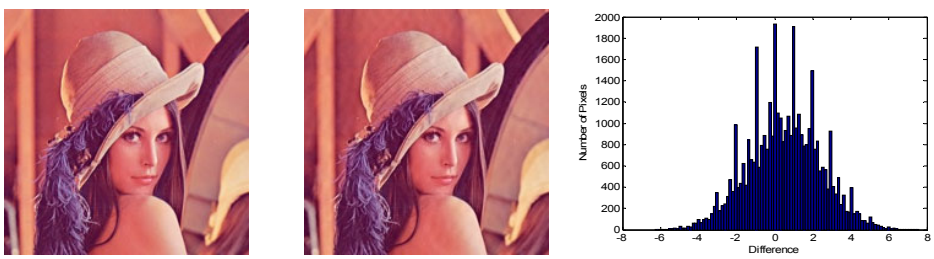


(a) (b)

Figure 6: Watermark extraction results for combined JPEG-compression (quality factor Q) and scaling.
(a) Message extraction results with correct CRC. (b) Message extraction results with correct watermarking pattern.

During these tests the gateway was configured to resize proportionally each image to fit into a 200×200 frame. To test the robustness of the watermarking scheme we have used a set of 30 combined scale factors (from 1 to 0.4) and JPEG-compression transformations (from a quality factor of 100% to 40%). For each test we have collected two different statistics: correct message extraction and correct watermark extraction. As can be concluded from the obtained results, the described

scheme is robust to geometrical resize operations even when combining with JPEG-compression. For low compression rates (the most probable case under MMS) the extraction probability of the embedded message is between 80% ($Q = 90\%$) and 99% ($Q = 100\%$) for the worst scaling case. Concerning the extraction of the watermark, which may serve as a zero-bit watermark, it is observed that the algorithm exhibits above 98% of correct extraction probability. This is a remarkable result, since, according to [14], the Digimarc watermarking algorithm is able to recover the watermark in about 72% of the situations (although under a different and unknown test). For compression attacks the algorithm exhibits an extraction probability of around 80% for CRC and 93% for the watermark identification pattern. Furthermore, as can be observed in Figure 7, the described scheme is able to preserve the perceivable image quality. These tests were performed with a 2GHz Pentium IV computer with Windows XP running Matlab. Using this development system, the largest images in the test bed took an average of 1.1 seconds for each watermark extraction.



(a) (b) (c)

Figure 7: (a) Unwatermarked image. (b) Watermarked image. (c) Intensity difference between images in (a) and (b).

The major challenges for watermarking in MMS are: (i) the small size of images; (ii) the required computational efficiency; (iii) and the robustness to geometrical resize operations. From the results presented in Figure 6 it can be concluded that our watermarking strategy is robust to this type of operations. Regarding colour depth attacks, the watermarking method is almost invariant, since the watermark is embedded into the image's intensity and not into its chromatic channels. In fact some preliminary tests with true colour images indicate that the algorithm is not sensible to a 256 indexed colour depth reduction.

The results in Figure 6 also suggest that the extraction performance for small scaling factors combined with large loss-compression rates is mainly induced by the smearing effects introduced by the interpolation filter. Namely, for a scaling factor of 0.5, where interpolation is equivalent to drop half of the image's columns and lines, it is observed that, regardless the compression ratio, extraction results are always better than for neighbouring factors, i.e., for scaling factors 0.4 and 0.6. This may suggest that the performance may be largely improved, if proper frequency magnitude attenuation effects induced by the low-pass interpolation filter are compensated during the watermark extraction phase.

6. Conclusions

This paper has described a DRM solution for MMS messages that uses a centralized approach and watermarking technology to detect and avoid illicit traffic of copyrighted content. The solution is easily deployed in any Mobile Operator infrastructure (only requires the installation of a Proxy Server between the WAP Gateway and the MMC) and is completely independent from the capabilities of the MMS terminals. The solution provides full interoperability between mobile phones and is also prepared for the adoption of OMA-DRM with future terminals that will support this standard. The watermarking algorithms that have been devised by the University of Coimbra are mainly targeted to MMS content. It has been proved to be extremely effective, it presents a high level of robustness and acceptable levels of performance. Currently, we are starting a trial in a Mobile Operator, and with the support of Ericsson we will conduct a more comprehensive benchmarking study. Future work includes the support for MMS video.

Acknowledgements: This work was partially financed by ADI (Agência de Inovação) and the POSI Program, supported by the Portuguese Government and the European Union.

References

[1] 3GPP (2002). *Multimedia Messaging Service*. 3GPP TS 22.140 v6.0.0
[2] http://uk.gsmbox.com/news/mobile_news/all/75393.gsmbox
[3] Open Mobile Alliance™(2002). *Digital Rights Management*. OMA-Download-DRM-v1_0
[4] 3GPP (2002). *Multimedia Messaging Service: Functional description (Stage 2)*. 3GPP TS 23.140 v6.0.0
[5] WAP FORUM (2001). WAP-250-PushArchOverview: *WAP Push Architectural Overview*.
[6] Open Mobile Alliance™(2002). *Download Architecture*. OMA-Download-ARCH-v1_07
[7] S. Voloshynovskiy, F. Deguillaume, T. Pun, *Content Adaptive Watermarking Based on Stochastic Multiresolution Image Modelling*, 10th European Signal Processing Conference, 2000.
[8] S. Pereira, *Robust Digital Image Watermarking*, Ph.D Thesis, Université de Gèneve, Switzerland, 2000.
[9] W. Zeng, B. Liu, *A statistical Watermark Detection Technique without Using Original Images for Resolving Rightful Ownerships of Digital Images*, IEEE Transactions On Image Processing, 1999.
[10] S. Pereira, T. Pun, *Template Matching for Affine Resistant Image Watermarks*, IEEE Trans. On Image Processing, 9(6):1123-1129, 2000.
[11] G. Csurka, F. Deguillaume, J. O'Ruanaidh, T. Pun, *A Bayesian Approach to Affine Transformation Resistant Image and Video Watermarking*, In Proc. 3rd Int. Workshop on Information Hiding, 315-330, 1999.
[12] M. Kutter, *Watermarking Resisting to Translation, Rotation and Scaling*, In Proc. SPIE Multimedia Systems Applications, 3528, 423-431, 1998.
[13] J. O'Ruanaidh, T. Pun, Rotation, *Scale and Translation Invariant Spread Spectrum Digital Watermarking*, Signal Processing, 66:303-317, 1998.
[14] H.-S. Kim, Y. Baek, H.K. Lee, *Rotation, Scale, and Translation Invariant Image Watermark using High Order Spectra*, SPIE (to be published), 2003.
[15] C.-Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, Y. Lui, *Rotation, Scale and Translation Resilient Watermarking for Images*, IEEE Transactions On Image Processing, 19(5): 767-782, 2001.
[16] M. Kutter, S. Bhattacharjee, T. Ebrahim, *Towards second Generation Watermarking Schemes*, in Proc. IEEE Int. Conference on Image Processing, 320-323,1999.
[17] S. Guoxiang, W. Weiwei, *Image Feature Based Second Generation Watermarking in Wavelet Domain*, in Lecture Notes in Computer Science, 2251:16-21, 2001.
[18] A. Santos, L. Moura Silva, P. Martins, P. Carvalho, *Frequency-based watermarking with spatial masking for DRM of MMS content*, Int. Conf. On Signal and Image Processing, Honolulu, USA, August, 2003
[19] http://www.microscopy-uk.org.uk/ovlib/london

Experimental DRM Architecture Using Watermarking and PKI

Mikko Löytynoja, Tapio Seppänen, Nedeljko Cvejic
MediaTeam Oulu
Information Processing Laboratory
University of Oulu, Finland
{mikko.loytynoja, tapio.seppanen, nedeljko.cvejic}@ee.oulu.fi
http://www.mediateam.oulu.fi

Abstract

This paper describes an experimental digital rights management architecture which uses digital watermarking and encryption to protect the content. In addition to cryptographically protected media content, the deliverable package can also include a teaser that the user is allowed to consume freely. The content is identified with an attack-resistant watermark that is used to automatically acquire a license needed to consume the rest of the content. The watermark is also used to prevent conforming players from showing protected content if the encryption has been removed. The architecture uses a public key infrastructure to handle licenses.

1. Introduction

Content is increasingly in digital form and is distributed using the Internet. The ease of copying has created a need to develop a means to protect it. Digital rights management (DRM) tries to find a solution to this problem inside a triangle set by technology, economics and law. The optimal solution is a compromise between technological possibilities, cost, ease of use, privacy and rights defined in law. The proposed DRM architecture uses some technological methods as well as threats of financial losses to protect the content.

The functional DRM architecture can be divided in three areas: content creation, content management and content usage. Content creation includes the creation of the media and defining the rights. Content management is about content distribution and trading of the rights. Finally, content usage is used to enforce that rights are adhered to and to track content usage. [1] A DRM system usually contains encryption and key management, access control, copy control, identification, tracing and billing mechanisms. Access control must be done using a flexible set of usage rules that define what the user can do with the content. Copy control is used to prevent making unauthorized copies of the content and this is usually very hard to achieve. Identification and tracking can be used as a last resort to track the source of pirated copies and enable legal action. [2]

Currently there are number of DRM solutions available, but they often use proprietary formats or need a plug-in for the player software. The current products lack interoperability and the content is usually locked to a terminal, which prevents the user from consuming the content using different terminals. This restricts what the users can do with the content and lowers the user experience. Open standards, such as MPEG-21 [3] and Open Mobile Alliance [4], aim to define a standard way to add DRM functionality to multimedia applications and promote interoperability. However, the MPEG-21 is a huge effort and covers many aspects of content creation and consumption. It remains to be seen how it will be adopted by the industry. There is also an open source project, called OpenIPMP [5] to develop a DRM framework that uses MPEG-4 IPMP extensions and rights definition languages. Trusted Computing Group [6] is an industry standard body that aims to develop a trusted computing platform specification for PC's, servers, PDAs, digital phones and other devices. Their intention is not to address DRM requirements but to focus on protecting user data and keys.

One of the hardest problems facing the DRM systems is how to define the user rights. The current rights expression languages restrict the user by allowing her to use the content only in the ways defined in the license. This often violates the rights the user would have according to the law. [7]

A weak point of any DRM system is that once the protection on the content is broken, it can be distributed all over the world, e.g., using peer-to-peer (P2P) networks. Using legal measures is one way to try to fight back against this kind of piracy, as it is often a small group of users who share most of the pirated content and others are only downloading it. This creates an asymmetry among users of P2P networks and it can be used to raise legal suits against these super-peers. But in the end, this alone cannot stop the copying entirely. [8] As the piracy cannot be removed, only solution is to convince users to use legal content e.g. by making it easier and cheap

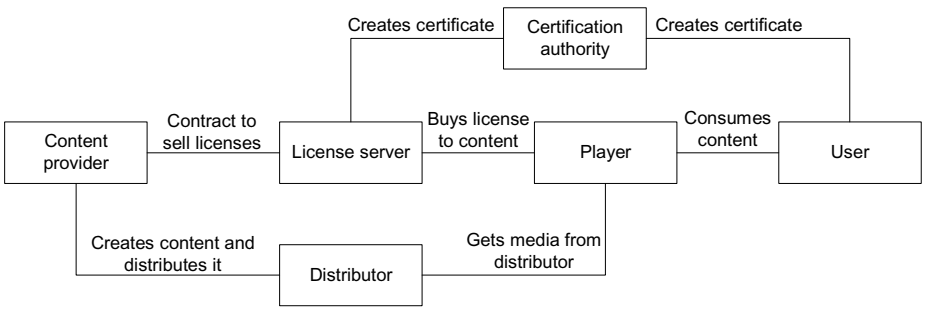


Figure 1. Overall architecture.

enough for users to buy legal content. It can be claimed that a certain amount of piracy can even be healthy, as it advertises the content and if the piracy rate is not too high, it can actually increase the revenues to content provider. Super-distribution can also be an efficient channel for content providers to distribute their content, provided that the content cannot be used without first acquiring a license for it.

Our proposed DRM architecture is designed to allow a super-distribution style of content distribution. The licenses are bound to the user and not the terminal they use. This allows users to access the same content with a desktop PC, a PDA and other mobile terminals. A network connection is only needed when the user first acquires the license from a server. The proposed system uses XML based licenses. They currently define only simple usage rights which enable the user to consume the content as long as she has a valid license for it. Later, the system should be extended to use some rights expression language, such as XrML [9] or ODRL [10].

The acquiring of the licenses is done using the model described in [11], in which the player identifies copyrighted content and acquires a license to consume it. The architecture uses services provided by a public-key infrastructure (PKI). The proposed architecture assumes that there is a PKI available, and we therefore will not be addressing the many problems related to creating one. A fully functional PKI offers many services such as certificate creation, revocation, updating, time stamping etc. [12]. The proposed system uses only a subset of them, such as certificate creation and revocation. The certificates we use are X.509 and they are used to buy and sign the licenses.

The content is protected using two methods, digital watermarking and public-key encryption. The architecture should provide multiple versions of algorithms with varying computational complexity in order to support mobile terminals with low processing power. The used algorithms must be a compromise between strength of protection and complexity. The malicious users can target their attacks against these both. The encryption algorithm used prevents direct attacks against the encryption, but this still leaves the possibility of attacks against the key management. We discourage users from sharing their private keys by using the same keys for buying the licenses, which places the users at a risk of financial loss if they share their keys. The watermarking algorithm used is robust against the most important forms of attacks.

In the next chapter, we describe the overall design of the architecture, the way licenses are handled and the watermarking method used. Chapter 3 discusses the current implementation of the architecture, as well as some attack scenarios and how to counter them. Chapter 4 concludes the discussion.

2. PKI-based DRM Architecture

The proposed architecture uses digital watermarking and public-key encryption to protect the content. The architecture uses standard media formats, and the DRM information is embedded into the content itself with watermarking. This allows the content to be converted to other formats, although there are some limitations set by the encryption. In the following, we describe the overall architecture and the watermarking method used. The current implementation uses audio content.

2.1. Overview

The overall architecture is shown in Figure 1. The architecture consists of content providers, distribution channels, license servers, certification authorities, player software and media. The content provider (CP) creates the content and distributes it to users using some media. The CP owns the intellectual property rights (IPR) to the content. Distribution channels are used to distribute the media content to users. This can be anything from web-servers to peer-to-peer networks used to distribute media from one user to another. A license server is used for acquiring licenses needed to consume the content. They

can be operated by the content providers or they can have a contract with the CP to grant licenses to users on their behalf. The certification authority (CA) is part of the public-key infrastructure (PKI). Its task is to link the identities of users and their encryption key pairs together using certificates. The architecture uses X.509 certificates, which are used to verify the authenticity of licenses and authorize the buying of them. The player software is used to consume the content. It enforces that the IPR are followed and is used to acquire licenses as needed. The media consists of the content and a user needs to have a valid media/license pair to consume the content.

The content created by the CP is protected using two mechanisms. First, the content is watermarked and second, it is encrypted. Currently the protection mechanism is implemented directly in the player, but in the future we plan to use downloadable tools in the player to extract the watermark and decrypt the content. The watermark is used to identify copy protected content and to carry information needed to acquire a license. We will further discuss the watermarks in the next section. Encryption is performed by scrambling the media on content level, not by encrypting the whole media file. This enables the encryption of selected parts of the media, as there are some formats that do not work if some parts of the file are corrupted. This selective encryption can be used to reduce computing power needed in the terminal, e.g. by encrypting only every other second of the content. This also allows the protected media to be consumed, although the encrypted parts appear as noise. Again there is compromise between computational complexity and protection level attained.

In our implementation, the media contains a teaser part that is not encrypted, and the user is free to consume it without a license. The player software then tries to find a watermark from the teaser and if it finds one, it asks the user if she would like to acquire a license to consume the rest of the content. If the user is willing, the player extracts the watermarked information. This information contains the URI to locate the license server and an ID to identify the content in question. Currently the content ID used is an alphanumerical string defined by the CP and it is assumed that IDs are unique in the context of the license server URI used.

The player sends to the license server the content ID and a user's certificate. The message is signed with the user's private key and the user is authenticated using the certificate. The server replies with the information about how much various licenses cost, how long they are valid, what they allow user to do, etc. If the user agrees to buy a license corresponding to one of the options, the player sends the agreement with the license information signed with the private key back to server. Finally, the server creates a license, signs it with its private key and sends it to the player. This process can be illustrated as

U → P: consume content
P → U: want to acquire license
U → P: agree
P → S: $\{U, N_P, \text{contentID}\}_U$
S→ P: $\{\text{previous message}, N_S \text{ license information}\}_S$
P → U: license information
U → P: agree
P → S: $\{\text{previous message, agreement}\}_U$
S → P: license

The left side identifies who sends a message to whom: U is the user, P is the player and S is the license server. The right side shows the contents of the message. N_X is random nonce and U is the user's certificate, $\{\}_K$ indicate that the information is signed with K's key. The communication between the player and the server is based on simple object access protocol (SOAP) [13].

The license is an XML file that contains the user's certificate, the content decryption key encrypted with user's public key, and the information about what the license allows user to do. The encryption key is encrypted using XML encryption [14] and the license is then signed using the XML signatures [15]. Before the player lets the user consume the content, it checks that the license is valid and that the user has the private key corresponding to the certificate in the license. Because the licenses are bound to the user's private key, they can be used in more than one terminal. This allows the user to consume the content both in her desktop PC and PDA. Therefore, in mobile terminals, a network connection is only needed when acquiring the license.

2.2. Watermarking

Digital watermarking is a process that embeds an imperceptible message to multimedia content that is difficult to remove. [16] The embedding process changes the content data so that the introduced distortion is kept below the just noticeable difference (JND) level of the human perceptual system. For example, the masking effects of the human perceptual system are applied.

In the architecture, the watermark has multiple functions. First, it is used to mark whether the content is protected so that a conforming player must not allow user to consume the content without a valid license. Second, the watermark is used to identify the content and the license server where licenses can be acquired. Finally, content providers can utilize the watermarking scheme for tracking of content distribution. If the watermark includes the identity of the end-user who purchased the content or the distributor, the watermark can be used to track down who is responsible for distribution of pirated copies of the content.

The proposed system uses improved spread spectrum modulation with an attack characterization method to

watermark audio content. The overall scheme of the watermark embedding algorithm is given in Figure 2. Samples of host audio are input to attack characterization and temporal masking analysis modules. The masking analysis module calculates the power level of the audio to determine the maximum power level of the watermark to be embedded. The purpose of the attack characterization is to test the host signal against mp3 compression and low-pass (LP) filtering, which are the most common removal attacks to the watermark, in order to optimize watermark embedding.

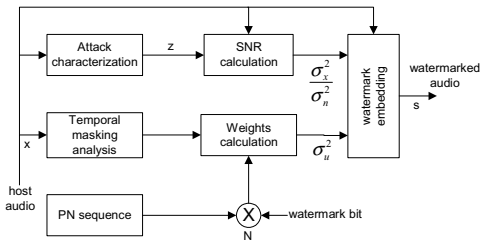


Figure 2. Watermarking embedding scheme.

The watermark embedding process can be described as:

$$\mathbf{s} = \mathbf{x} + (\alpha b - \lambda x)\mathbf{u}$$

where **x** stands for original host signal vector, **s** stands for watermarked audio vector and **u** for the pseudorandom (PN) sequence after the perceptual adaptation process. Watermark variable b is either +1 or -1, depending the bit to be embedded. Parameters α and λ control the distortion level and removal of host signal influence on the detection statistic. Variable x is defined as inner product of **x** and **u** divided with norm of **u**. The embedding can try to keep the bit error rate (BER) at a predefined value with variable watermark capacity or try to obtain a constant watermark capacity if a variable BER is allowed.

Watermark detection is performed by cross correlating the watermarked content with the inserted watermark. The principle is the same as that in a standard spread spectrum receiver that decides on a received bit according to the sign of a normalized sufficient statistic.

3. Experimental Results

In this chapter, we discuss how the architecture is currently implemented and consider some possible attack scenarios and how to counter them.

3.1. Implementation

Our first version of the architecture is implemented using Java 2. We use the IAIK Java cryptography extension to do the encryption and IAIK IXSIL to sign the licenses. [17] The basic functionality described above has been mapped to two network components, a client and a server. The client component consists of the player used to consume the content and also handles the licenses the user has acquired. The server component is used to create certificates, protect the content, to give licenses to users when needed, and to distribute the content. In the first version, we use audio content, but we have also done some tests to protect videos. The audio media is protected by first embedding the watermark to it and, after that, encrypting the audio data using advanced encryption standard (AES) algorithm [18]. This algorithm can decrypt the content in real time using a desktop PC, but mobile devices might not have the processing power to do this. CP can also encrypt the content at specific intervals, which allows for the use of less powerful terminals. Later versions of the architecture will also use other algorithms that require less computing power. Currently the watermarked data is an URI to the license server and a random ID number of the content.

If the media is compressed using lossy compression, the encryption must be done on the compressed domain of the media. If the compression used is lossless, then it is possible to do the encryption in the uncompressed domain. Using lossless compression also makes it possible to change the media to another format while still allowing the decryption of the content and the extraction the watermark. This is useful as the content is not restricted to some specific format.

The player finds the address to license server by extracting the server URI from the watermark. It connects to the server using SOAP. The SOAP implementation used is GLUE [19]. The license is an XML file signed with license server's private key using XML signatures and it contains the ID of the content, the user's certificate and the decryption keys needed to decrypt the media. The player also has the certificate of the CA to enable verifying signatures in licenses.

3.2. Attacking Experiments on Protection

The architecture protects the media using two methods that can both be attacked. The encryption prevents nonconforming players from showing the content and because of the algorithm used it is not feasible to try to break the encryption. Therefore, the obvious way to attack the encryption is through key management. The license contains the key needed to decrypt the content. The simplest way to get a valid key to content is to buy one and use that to extract the decryption key. To prevent

users from simply decrypting the license and sharing the key, the conforming players require that a valid license is present before decrypting the content. To prevent users from creating licenses themselves, the license must have a valid signature of the license server. For the user to be able to use a license allocated for her, she must have an equivalent private key. This prevents users from using other user's licenses, unless they also have the corresponding private key. Because private keys are used to authorize buying of licenses, this should discourage users from sharing their private keys. This cannot prevent getting pirated content, but aims to make it as hard as possible. Even if the encryption is broken, there is still the protection given by the watermark left: the conforming players refuse to play protected content without a valid license. The huge problem which our approach does not solve is with non-conforming players. They can be used to play the protected content, as long as the specification of the format used is available. Using downloadable DRM tools might help somewhat, but to really solve the problem a trusted computing platform is needed. This allows the player to be trusted, which currently is not the case. One possible solution could be the licensing terms of used technology, which could require that players enforce copyrights. This allows legal action against makers of non-conforming players.

The watermark is designed to be robust against attacks. We have tested our watermarking scheme, by watermarking large set of songs of different music styles and attacked them with mp3 compression and LP filtering. The SNR rating of the watermarked audio ranged from -25.5 dB to -27.3 dB, and the subjective evaluation involving test persons who listened to the original and watermarked audio sequences, gave an average score of 4.6 with a standard deviation of 0.42 on a 5-point scale (5: imperceptible, 4: perceptible but not annoying, etc). If the watermark is embedded at the rate of 50 bps the average BER in the presence of mp3 compression or LP filtering is about $3*10^{-4}$, and at rate of 200 bps the BER is around $3*10^{-3}$. These results show that it is hard to completely remove the watermark from the host signal without lowering the quality of the audio too much. Some bit errors are not fatal as the watermark can be embedded multiple times over the host media. Channel coding schemes can also be used to increase performance while making a compromise with the capacity to correct bit errors. [20]

4. Conclusions

It is impossible to create a perfect protection mechanism that cannot be broken. The goal of our research is to develop an architecture that supports various types of attack resistance. It should include strong mechanisms for protection while providing flexible policy options in order to test different e-commerce scenarios on the platform. We have tried to minimize the need for a network connection to allow users consume the content off-line with mobile devices.

As a framework for developing the architecture, we have chosen the PKI infrastructure due to its capability to provide effective protection and authentication functionality. Users are discouraged from sharing their private keys and thus license to content as keys can be used by the recipient to purchase the entire content on the charge of original user. As another key method of IPR protection we apply digital watermarking that can be used to embed in the content information on the identities of the content provider, the distributor and the end-user. This information can be utilized to discourage illegal distribution by supporting network monitoring applications. The watermark is also used to prevent conforming players from showing protected content if the encryption has been removed. An important part of our experimentation is the application of various watermark removal attacks. Our recent results indicate that a very high level of performance can be achieved against watermark removal attacks if latest technology is used.

The current implementation provides basic functionality of the proposed architecture. The future research will focus on the usage of downloadable DRM tools to enforce the content protection, as well as support for mobile terminals with restricted processing power. One of the big issues which are still open is how to prevent users from developing their own player software, which can be used to bypass the protection mechanisms.

5. Acknowledgments

The financial support from the National Technology Agency of Finland (Tekes), Graduate School in Electronics, Telecommunications and Automation (GETA), Infotech Oulu Graduate School, and the Stego project consortium is gratefully acknowledged.

6. References

[1] R. Iannella, "Digital Rights Management (DRM) Architectures", D-Lib Magazine, June 2001, Volume 7 Number 6.

[2] F. Hartung and F. Ramme, "Digital Rights Management and Watermarking of Multimedia Content for M-Commerce Applications", Invited paper, IEEE Communications Magazine, vol. 38, no. 11, pp. 78-84, November 2000.

[3] J. Bormans and K. Hill (ed.) "MPEG-21 Overview v.5", ISO/IEC JTC1/SC29/WG11/N5231, October 2002, Sanghai.

[4] Open Mobile Alliance, http://www.openmobilealliance.org/

[5] OpenIPMP, http://openipmp.com/

[6] Trusted Computing Group, http://www.trustedcomputinggroup.org/

[7] D. Mulligan and A. Burstein. “Implementing Copyright Limitations in Rights Expression Languages”, Proceedings of 2002 ACM Workshop on Digital Rights Management, November 18, 2002, Washington DC.

[8] P. Biddle, P. England, M. Peinado, and B. Willman “The darknet and the future of content distribution”, Proceedings of 2002 ACM Workshop on Digital Rights Management, November 18, 2002, Washington DC.

[9] The eXtensible Rights Markup Language, http://www.xrml.org/

[10] The Open Digital Rights Language, http://www.odrl.net/

[11] J. S. Erickson, “Fair use, DRM, and trusted computing”, Communications of the ACM, Volume 46, Issue 4, p. 34 – 39, April 2003.

[12] C. Adams and S. Lloyd, “Understanding the Public-Key Infrastructure: Concepts, Standards, and Deployment Considerations”, New Riders Publishing, 1st edition 1999.

[13] World Wide Web Consortium, Simple Object Access Protocol (SOAP) 1.1, W3C Note 08 May 2000.

[14] World Wide Web Consortium, XML Encryption Syntax and Processing, W3C Recommendation 10 December 2002.

[15] World Wide Web Consortium, XML-Signature Syntax and Processing, W3C Proposed Recommendation 20 August 2001.

[16] I. Cox, J. Bloom, M. Miller, “Digital Watermarking: Principles & Practice”, Morgan Kauffman Publishers, 1st edition 2001.

[17] Institute for applied information processing and communications, http://www.iaik.at/

[18] National Institute of Standards and Technology (NIST), “Advanced Encryption Standard (AES)” Federal Information Processing Standards Publication 197, Nov. 26, 2001.

[19] The Mind Electric GLUE, http://www.themindelectric.com/

[20] N. Cvejic, T Seppänen, Increasing robustness of an audio watermark using turbo codes. 2003 IEEE International Conference on Multimedia & Expo, Baltimore, MD, to appear.

A Generic DRM Framework for J2ME Applications

Nuno Santos, Pedro Pereira, Luís Moura e Silva

WIT-Software
Rua Pedro Nunes, IPN, 3030-Coimbra, Portugal
Email: luis@wit-software.com

Abstract

Recently, a new generation of mobile phones with support for Java has been taking widespread acceptance by the market, creating a business potential for downloadable Java Games and enterprise applications. However, it is relatively easy to forward Java programs between two Java phones. This opens the door for illegal peer-to-peer forwarding, with the consequent loss of revenues for content providers and operators. Therefore, DRM solutions are essential in order to protect copyrighted Java applications.

In this paper we present a generic DRM framework that supports different solutions for protecting the copyright of Java applications. This framework is mainly targeted to Mobile Operators, it is totally transparent to content providers and does not require any special support at the user's handsetss. It also allows the development of new custom built DRM solutions, providing a flexible platform for Java oriented DRM techniques.

Keywords: Java J2ME; DRM; copyright-protection; code instrumentation.

1. Introduction

The first generation of Java [1] enabled phones were very limited in terms of functionality. They were only able of downloading MIDlet[1] applications from the network and of executing them. They offered no simple way to copy a Java application to another terminal or PC. These limitations were a natural Digital Rights Management (DRM) [2] solution, since the only way to obtain a Java application was by downloading it from the network. However, the most recent mobile phones are much more feature-rich. These terminals often have a user visible file-system and are able to connect to another terminal or PC using USB, Bluetooth or infrared. This makes it easy for users to copy Java applications to other terminals. In this scenario, DRM is essential to prevent users from illegally forwarding copyrighted Java MIDlet applications.

[1] *MIDlets are small applications written in the Java programming language that run in all mobile phones that support the Mobile Information Device Profile.*

The current terminals are mostly based on the J2ME CLDC[2] 1.0 [3] and MIDP[3] 1.0 [4] specifications, which do not have any kind of DRM support. Hence, those terminals are an easy target for copyright violations, which can result in a significant loss of revenues for the application developers and the mobile operators.

This document presents a DRM framework that was developed by WIT-Software and has been integrated with a commercial Java Download Platform. The framework allows application providers to add DRM protection to MIDlets applications in their binary form, requiring no access to their source code. It works by doing some code-instrumentation of the MIDlet JAR file, adding copyright-protection code that will be executed on the terminal when the user starts that application contained in the JAR file.

The framework is sufficiently generic, allowing DRM solutions to be developed independently and deployed on a case basis. It can be used as a standalone DRM tool or as an Application Programming Interface (API). The former option is especially interesting for application providers, who can use the API to integrate the DRM framework with Over-The-Air (OTA) [5] provisioning systems. In this way, the instrumentation of the MIDlet is delayed until download time, allowing the system to choose the most appropriate DRM solution for the mobile phone that is requesting the MIDlet application.

The rest of this paper is organized as follows: Section 2 describes the related work on DRM solutions and code instrumentation. Section 3 discusses the suitability of J2ME MIDP 1.0 and 2.0 profiles for implementing DRM solutions. Section 4 presents the structure of the DRM framework. Section 5 describes three DRM solutions implemented for the framework. Finally, Section 6 concludes this paper.

2. Related Work

The Open Mobile Aliance (OMA) has defined a specification of DRM systems for mobile devices [6]. This specification addresses the protection of any type of media that can be delivered to mobile phones. This includes music, video, and applications, among others. It defines three methods of distributing content and right objects[4]:

[2] *Connected Limited Device Configuration.*
[3] *Mobile Information Device Profile.*
[4] *A rights object specifies the way the content can be used, like how many times it can be used, if copying is allowed or not, etc.*

- **Forward Lock:** The content is delivered unencrypted to the device, without any rights objects. It is up to the device to enforce a default set of rights and ensure that the content cannot be forwarded.
- **Combined Delivery:** The content is delivered unencrypted, together with a rights object. The mobile phone enforces the usage permissions specified in the rights object.
- **Separate Delivery:** The content is delivered encrypted. A rights object is delivered separately using WAP push. Since the content is encrypted, it can be forwarded freely. The receiving users will have to obtain a license before using it.

Several members of the industry, such as Ericsson, Siemens, Nokia [7] and DRMSecure [8], have already committed to the OMA DRM specification and are implementing parts of it in their products[5]. The main limitation of the DRM OMA specification is that it requires special support from the mobile device. Therefore, this specification does not solve the problem addressed by the framework presented in this paper, which is to protect Java applications delivered to the existing portfolio of Java-enabled phones.

There are some other companies with similar commercial offerings. The SDC Java DRM [9] is a technology for delivering content to mobile devices. The content is packaged inside a container together with the code necessary to access it. This container is protected using obfuscation and encryption techniques. In the device, the code is interpreted by a Java Virtual Machine on the device, enforcing the DRM rights. The available documentation was not very complete or clear, but it seems the system requires the presence of private keys in the mobile phone side. There was no description about key distribution and the portability of this scheme.

Other proposed scheme is MacroSafe [10], a product from Macrovision for content delivery. The solution is similar to the Separate Delivery mode of the OMA-DRM specification. Encryption is used to protect the content, which is delivered with a rights object. The client needs to retrieve the encryption key to be able use the content. This solution requires the presence of the MacroSafe's Client software on the client's device. There is no mention to whether the client will run on a J2ME device, but this seems unlikely since the specification of the client software seems to impose some device requirements that are not currently achieved by the J2ME devices.

[5] Nokia has recently launched a mobile phone supporting the OMA DRM standard: Nokia 6220.

3. J2ME MIDP 1.0 and DRM

Some hardware support for encryption and unique identification numbers is extremely important for DRM mechanisms. The presence of these resources is the basis for implementing strong DRM measures [2]. Unfortunately, the CLCD 1.0 and MIDP 1.0 specifications do not provide any adequate support. They define a very limited execution environment. In particular, some important features are missing, namely:

- There is only a very limited access to the file-system. This is done by means of a Persistent Record Store (PRS), which is a device-managed container. This makes it easy for any user to access the data that has been written by the MIDlet application on the mobile phone;
- There is no way of obtaining the terminal IMEI or any other type of device identification;
- Only a subset of the Java API is supported. There is no support for JNI (Java Native Interface), reflection or cryptography;
- Most devices running MIDP 1.0 are very limited in resources (CPU and memory). This is not a limitation of the specification, but it prevents the use of memory or CPU intensive DRM mechanisms. In particular, it makes strong cryptography almost unpractical.

These limitations severely constrain the type of DRM solutions that can be implemented. Traditional implementations like encryption, digital signatures, secure hardware and unique identification of the device are hard or even impossible to apply in such a restricted environment. Nevertheless, it is still important to have some kind of protection. Even if a DRM solution is not very hard to break, it will be useful if it prevents a significant number of DRM violations. Therefore, it is important to use the available support of the J2ME environment in the best possible way to protect the copyrights of J2ME downloaded applications.

There are some features of the MIDP 1.0 specification that can be useful for the implementation of a DRM solution:

- It is possible to read/write to the PRS, thus making it possible to keep a license together with the MIDlet JAR;
- It is possible to extract values from the JAD and the JAR manifest;
- It is also possible to obtain the current local time of the mobile phone.

Outside the specifications, some vendors provide proprietary extensions to MIDP 1.0. For instance, with the Siemens phones it is possible to obtain the IMEI identification. But exploiting vendor-specific extensions will most likely result in different and incompatible DRM

implementations, each one suited only for a certain type of terminal. This approach has some management difficulties, like ensuring that the right DRM solution is used on the user's device. There are several ways to solve this problem:

- MIDlet applications contain the implementation of all possible DRM systems. This would increase considerably the size of the application;
- Create different versions of the same MIDlet application for each type of terminal. When a client requests a MIDlet application, the most appropriate version will be downloaded. Even so, this may create some difficulties in managing a full package of MIDlet versions;
- Create only one version of the MIDlet application. When a client requests the MIDlet, the mobile phone is identified and the MIDlet is instrumented, by adding the corresponding DRM system. The drawback of this solution is the extra time it would take to instrument the code on-the-fly.

The framework described in this paper implements the last solution. It keeps the DRM implementations separated from the MIDlet applications. When there is a J2ME application download, the framework is used to instrument the MIDlet application with a specific DRM solution. This is done by including the DRM implementation classes in the MIDlet application and by performing some code transformations. The DRM code is executed the first time the Java application is started on the mobile handset. The generated MIDlet application can be sent directly to the terminal. Once there, the DRM code performs the necessary verifications and it only executes the MIDlet if there is a valid license.

4. J2ME DRM Framework

4.1 Overview

The main objective of the DRM framework described in this paper is to extend the Over-The-Air (OTA) provisioning servers with support for locking MIDlet applications at download time. In particular, it has the following goals:

- **Generic:** It should be possible to support different DRM solutions. This is necessary due to the limitations of the J2ME specification, which makes it difficult to implement a generic solution. Therefore, it is not realistic to expect that one solution will be adequate for all terminals and for all situations. It is more likely that a number of different solutions will co-exist, each one tailored to a different terminal. Hence, a framework that is able to support different solutions may have a significant advantage.

- **Transparency to the content provider:** The content provider should only provide an unprotected MIDlet application, without having to worry about DRM systems.
- **Transparency to the user:** The average user should not be aware of the presence of DRM mechanisms.

4.2 Description

The DRM framework consists of a component that integrates with the OTA server. Figure 1 presents an overall overview of the architecture.

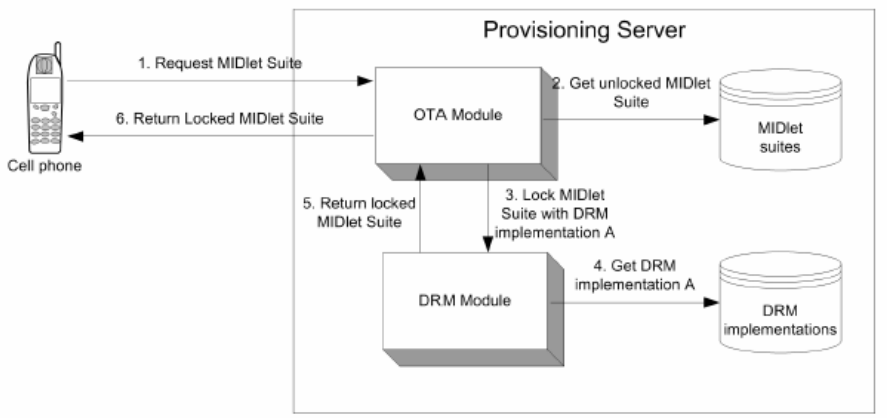


Figure 1: Architecture of the DRM framework

The DRM module maintains a database of DRM implementations. When a MIDlet application is requested by a client, the OTA server decides if it is necessary to protect it. If so, it chooses the DRM solution that is most adequate to the mobile handset, and uses the DRM framework to protect the MIDlet application with the corresponding DRM solution.

The DRM framework instruments the MIDlet JAR: it includes the DRM classes and changes the MIDlet's code so that when they are executed in the client's device, the DRM verification algorithm will be called before anything else. After being transformed, the MIDlet application is re-packaged, obfuscated and pre-verified. The result of this process is a MIDlet JAR ready to be sent to the mobile phone and protected by a DRM algorithm.

Structure of a DRM Solution

The DRM framework is general so that different DRM solutions can be developed separately and can be deployed into the framework in their binary form. For this to be possible, DRM solutions must follow certain rules of packaging, as described next:

- **Terminal side implementation:** A JAR file containing the classes that will be packaged together with the MIDlet application.
- **Server side instrumentation classes:** An optional server side implementation. These classes are only used at the server. Their purpose is to allow the DRM solution to perform custom instrumentation on the MIDlet and on the client side implementation. For instance, if the DRM solution needs to write a certain value to the MIDlet during code instrumentation, it is not safe to write it on the JAD, as it would be easily visible by the user. Instead, it is safer to insert the value inside the class files of the MIDlet application. This way, the value will be hidden from the average user.
- **Configuration file:** A standard Java properties file, specifying several properties, like the name of the main classes for the phone and server side implementations. It also describes the properties of the available solutions that can be accessed during instrumentation.

The terminal side implementation must provide a class extending an abstract class `wit.j2me.drm.DRMVerifier` (provided by the DRM framework). This class defines a `verify()` method, which should implement all the DRM verifications, starting the MIDlet if it succeeds and aborting the execution otherwise.

A similar requirement applies to the server side implementation, which must provide an implementation of the `wit.j2me.drm.CustomInstrumenter` interface. It defines an `instrument()` method, which should perform the custom transformations to be done at the server side.

Locking Process

This section explains how a MIDlet application is instrumented with a DRM solution. A MIDlet application has one or several MIDlets. The DRM algorithm must be executed before any other code of the MIDlet, so it is necessary to change the MIDlet to redirect execution to the DRM algorithm as soon as possible. When the user starts a J2ME application on its mobile phone, the application manager will show all the existing MIDlets, allowing the user to choose the one that will be executed. After that, the application manager will create an instance of the MIDlet's main

class and execute the `startApp()` method. In this process the following methods of the MIDlet are invoked, in this order: static constructor, instance constructor, `startApp()`. In order to execute the code corresponding to the DRM solution, it is necessary to intercept execution in the static constructor, which is the first method that is executed. Next, there is a description of what happens during instrumentation:

1. All classes contained in the terminal side implementation are inserted in the MIDlet application. This makes the DRM code available in the mobile phone. It is still necessary to change the original structure of MIDlet Jar, so that the DRM code is executed.
2. Each MIDlet is instrumented. The original static constructor is renamed and a new one is created. This new constructor calls `verify()` on the `DRMVerifier` implementation provided by the DRM solution (which was inserted in the MIDlet application in the previous step). This way, when any MIDlet is created, it will start the DRM algorithm.
3. All other public methods of the MIDlet, which can be called by the application manager, are renamed and replaced by stubs. These stubs will only call the corresponding methods of the original MIDlet if the DRM algorithm finished successfully. This is necessary because some DRM implementations might take a long time to execute. In this situation it is desirable to show a message to the user. This is not possible if the thread that is used by the application manager software to create the MIDlet (the dispatch thread) is blocked in the verification process. So, a new thread should be created and the dispatch thread released. In this situation, the dispatch thread will believe that the MIDlet is ready to execute and call the instance constructor and the `startApp()` method. If these methods were not replaced, the MIDlet would be started while the DRM algorithm is running.
4. (Optional) If a server side implementation is provided, it is used to perform custom instrumentation. The `instrument()` method of the `CustomInstrumenter` implementation is invoked, passing as parameters the MIDlet application and the client side implementation.
5. The MIDlet is re-packaged and obfuscated. Obfuscation makes it harder to understand the locking mechanism and to reverse engineer it. It also makes it smaller. This is important to minimize the extra footprint created by the DRM classes.
6. Finally, the MIDlet is pre-verified. This is necessary because the transformations that were performed removed the `StackMap` attribute that was on the original MIDlet classes.

At the end of this process, the MIDlet application is ready to be sent to the terminal. The instrumentation is performed using the `Javassist` library [11]. There are other options for

instrumenting Java code, the most significant one being the ByteCode Engineering Library (`BCEL`) [12]. An initial prototype of the DRM framework made use of BCEL. But it soon proved to be too complex for the task at hand. Therefore we decided to use `Javassist`, which focuses mainly on high-level manipulations. These higher-level abstractions proved to be more appropriate to the task, allowing us to implement the solution in less time and in a much simpler way.

Disadvantages

The DRM framework described before has a few disadvantages.

- The locking process takes a few seconds. Depending on the server and on the DRM solution, it may take something from one to five seconds. This may create a noticeable delay for the client who is waiting for the Java application.
- The MIDlet application size generally increases, as more classes are added to it. This mainly depends on the number and the size of the classes to will be added. It also depends on whether the original MIDlet was obfuscated or not. If it was not, it is even possible that the size decreases, as the size reduction obtained by obfuscating the MIDlet classes might compensate for the size of the new classes.
- Finally, there is an overhead in starting the MIDlet on the mobile phone, caused by the execution of the DRM algorithm. This depends on the complexity of the DRM solution.

5. Some Sample DRM Solutions

To test the framework described in the previous chapter, several DRM solutions have been implemented. They have different trade-offs. The first solution that will be presented does not require the mobile phone to connect to the server in order to obtain a license, but is relatively easy to bypass. The other solutions are harder to break, but require extra connections to obtain and verify the application license. The final solution works only on mobile phones supporting the Wireless Messaging API (WMA) [13]. These are three examples of DRM techniques for Java applications that can be deployed with the generic framework described in this paper.

5.1 Connectionless Approach

The first solution is transparent to the user and requires no support from the network other than the instrumentation phase during the download of the MIDlet application. Unfortunately, it is somewhat weak. This solution makes the following assumptions:

- After the MIDlet being downloaded to a mobile phone, the user will execute it for the first time within a brief period of time.
- During that timeframe, the user will not try to copy the MIDlet application to other devices.
- The clock of the user's mobile phone is set to the current time and the user will not try to change it.

The timeframe mentioned in the first and second assumptions is the validity period of the MIDlet application. This should be a period of a few hours. When the MIDlet is being instrumented, a timestamp is placed inside the DRM classes. In the first execution at the mobile phone, the MIDlet checks the time on the phone and verifies if it is still inside the validity period. If so, it makes an entry in its PRS indicating that the verification was successful, and executes the application. Otherwise, it registers an unsuccessful verification in its PRS and aborts the application. The next time it executes, it will check its PRS and execute only if it finds the record for successful verification.

Advantages

The main advantages are simplicity and transparency. Honest users, those that do not try to make illegal copies, will not notice the existence of a DRM solution. The size overhead of this solution is small (2k to 3k) and its execution is fast. Another advantage is that no extra communication is required with the Mobile Operator. All the verifications occur within the mobile phone. There is also no need for any server side infrastructure other than the DRM framework.

Disadvantages

The timing assumptions made are very strong, therefore making this DRM solution easy to break. For instance, if there is a significant drift between the clock in the server where the MiDlet is instrumented and the user's mobile phone, the MIDlet application can fail to run for the first time. Also, this solution cannot prevent the user from copying the JAR and installing it on another phone before the validity has expired. Another way to break the protection is by changing the time of the phone clock. Finally, if the user has successfully obtained a license for the MIDlet

application, he can overcome the DRM protection by copying the JAR together with the PRS to another phone.

5.2 Server Licensing Approach

In this approach, when the user requests the download of a J2ME application, the server generates a unique license and a key. The license and a hash of the key are stored in the MIDlet application, which is to be sent to the client's device. The server keeps a copy of the license and the key. When the MIDlet application is first executed the following will happen:

1. The MIDlet establishes an HTTP connection with the server and sends the license.
2. The server checks if the received license is valid and answers accordingly. If the license is valid, it sends also the key.
3. If the phone receives the key, it compares the hash of key with the hash stored in the MIDlet. If they are equal, the MIDlet executes.

The purpose of the key is to prevent a malicious user from sending a forged HTTP packet to validate the license. If the server simply sends the same `Ok` message to all MIDlets, it would be easy to forge this message. By using a key, each confirmation is unique. Furthermore, the user has no way of guessing which packet the MIDlet is expecting from its content, since only the hash of the key is stored[6].

This solution requires the presence of a license server, with the following interfaces:

- **With the Provisioning Server:** A standard API interface, allowing the provisioning server to create a license and the corresponding key for a download request.
- **With the client's terminal:** An interface to be used by the client's terminal to validate the license. Internally, the server must be able to generate licenses and keys. It should keep track of them until the client connects to authenticate itself.

Advantages

This solution is stronger than the connectionless approach. An untampered MIDlet can only execute after receiving the corresponding key that is maintained in the server. Therefore, it must validate its license before executing. This solution can also be easily adapted to support super-distribution of MIDlet applications. For this to be achieved, the server must allow the same

[6] One property of the hash algorithms is that it is not feasible to guess the original message from the hash.

license to be validated more than once and must be able to bill the cost of the J2ME application to the owner of the mobile phone that is requesting the license.

Disadvantages

The main disadvantage is that the *honest* user will notice the DRM mechanism. The MIDlet will have a significant delay in the first execution while it establishes an HTTP connection. Also, this connection will probably have to be paid by the user. Finally, the user must have network coverage to install the application. At the server side, it is also necessary to maintain a license server with a full history of users and downloaded applications.

5.3 SMS-based Approach

This approach is similar to the Server Licensing Approach described in Section 5.2. The main difference is that SMS messages are used to validate the MIDlet. This also requires a slightly different license server, which instead of having a Web interface must have an SMS interface. Apart from that, the core functions are similar.

Advantages

Using SMS messages has some advantages over HTTP. First, the license can be sent in a free SMS message, which is more desirable for the end-user. Second, it is easier to support super-distribution, by using MT (Mobile Terminated) messages to charge for the use of the J2ME application.

Disadvantages

This solution has the same disadvantages as the Server Licensing Approach. The time it takes for the registration can also be noticeable, as SMS messages may take a while to reach its destination. Finally, the terminal must support the Messaging API for sending SMS, which is an optional package of MIDP 2.0. Currently, most terminals are still using MIDP 1.0 and even on MIDP 2.0 terminal, the availability of the Messaging API is not always certain, since it is an optional package.

5.4 Analysis of the DRM Solutions

The DRM solutions presented in this chapter are only a sample of what is possible to do using our DRM framework. The development of these three solutions allowed us to validate the advantages and disadvantages of the framework. The interface provided by the framework proved to be adequate: it abstracts the developer from the low-level details – like instrumentation, JAR re-packaging, manifest updating, pre-verification, and so on. It also proved to be easy to use. The interfaces and classes that had to be implemented are well identified and defined, making it easy for the developer. Also, the possibility of bundling external classes with the DRM solution proved to be very important, as it allowed the solutions to be developed using standard coding practices, like splitting the code across multiple classes, using inheritance, and previously developed packages, among others. All the solutions described in this chapter were tested with several off-the-shelf J2ME applications and they have proved to be effective in all those cases.

6. Conclusion

This paper presented a DRM framework to protect J2ME applications. This development was motivated by the limitations of the J2ME platform that do not provide any specific support for DRM techniques. The current specifications (MIDP 1.0 and MIDP 2.0) does not offer as well any support for the standard tools that are usually used to implement copyright protection schemes, like encryption, key management or unique identification of the device.

Our solution tries to overcome this limitation and the platform we have presented in this paper provides flexible support to deploy different DRM solutions. The framework works by instrumenting MIDlet applications with copy-protection code. This can be done to any Java application in its packaged format (JAR), without having access to the source code. This makes the DRM framework transparent to the Content Providers. The framework is also extensible, allowing different DRM solutions to be developed and deployed.

This paper also presented three particular DRM solutions, each one with different trade-offs. One is easy to break, but has very few requirements. The two others are stronger, but they require more resources from the mobile phone and the network. In the future, other DRM solutions will be implemented, exploiting the available resources on each type of mobile handset.

References

[1] J. Gosling, B. Joy, G. Steele, and G. Bracha, "The Java Language Specification". Boston, Massachusetts, EUA:Addison-Wesley, 2000.

[2] Q. L. Reihaneh, R. Safavi-Naini, and N. Sheppard, "Digital Rights Management for Content Distribution," in *Proc of Australasian Information Security Workshop*, (Adelaide, Australia), 2003.

[3] Sun Microsystems Inc., "Connected Limited Device Configuration (CLDC) Specification, V1.1 (JSR-139)." Available at: `http://jcp.org/aboutJava/communityprocess/final/jsr139`.

[4] Sun Microsystems Inc., "Mobile Information Device Profile (MIDP) Specification, V1.0 (JSR-37)." Available at: `http://jcp.org/aboutJava/communityprocess/final/jsr037`.

[5] Sun Microsystems Inc., "Over-The-Air User Initiated Provisioning Recommended Practice." Available at: `http://java.sun.com/j2me/docs/`.

[6] Open Mobile Alliance, "Digital Rights Management Version 1.0, Specification." Available at: `http://www.openmobilealliance.org/documents.asp`.

[7] Nokia, "Nokia's White Paper on DRM." Available at `https://secure.forum.nokia.com/main/1,,040,00.html?fsrParam=1%2D3&fileI%D=2894`.

[8] DRMSecure, "DMDMobile." Available at `http://www.dmdsecure.com/DMDmobile.htm`.

[9] Secure Digital Container (AG), "SDC Java DRM." Available at `http://www.digicont.ch/sdc_java_drm/core_technology/index.html`.

[10] MacroVision™, "How MacroSafe™Works." Available at `http://www.macrovision.com/solutions/video/drm/overview.php3`.

[11] S. Chiba. "Javassist – a reflection-based programming wizard for Java". In Proc. Of OOPSLA'98 Workshop on Reflective Programming in C++ and Java, October 98.

[12] M. Dahm, "Byte code engineering," in *Java-Informations-Tage*, (Düsseldorf, Germany), pp. 267–277, Sept. 1999.

[13] Sun Microsystems Inc., "Wireless Messaging API (WMA) Specification, V1.0 (JSR-120)." Available at: `http://jcp.org/aboutJava/communityprocess/final/jsr120/`.

DRM and Digital Radio Archives

Antti Järvinen, Pekka Gronow
Finnish Broadcasting Company YLE
antti.jarvinen@yle.fi, pekka.gronow@yle.fi

Abstract

Public broadcasters are currently transforming their historical collections from analogue formats to digital domain. One of the big challenges is rights management (RM) which is becoming even more complex issue when file based production systems are used.

As a national broadcasting company YLE has a long experience in managing the rights of audio-visual works used in broadcasting. It is also one of the first broadcasters which have built a digital archive for the permanent collection. We will describe the practice of rights management in the traditional broadcasting environment and anticipate some of the problems related to move to a fully digital IT-based production process.

1. Introduction

Today, rights management is a time consuming process in radio broadcasting, which still requires a lot of human interaction. In this paper, we discuss technical and other requirements for a digital rights management system (DRM), which could be used for rights management in broadcast environment in a case of national broadcaster having sizeable historical collection. We also present current situation and problems, which have to be solved in DRM that it would be applicable for RM in our case.

We will discuss on the virtues and problems of digital file based archives and archive centric broadcast production process. We present current problems related for example to the copy protected CD's and integration of DRM systems in broadcast production environment. These examples show clearly what kind of problems quick release of immature technology can create to daily operations of a broadcaster.

We will point out some aspects, which show that design principles for a DRM system used in broadcasting should be quite different from system applied in consumer market. The main reason for this is that requirements for technology used in broadcast environment does not allow technology that might prevent or stop transmission. Broadcasters also have long tradition in rights management and they have well established tradition in transmission right acquisition from right owners or societies representing them and due to this, security requirements differ from consumer business.

2. Rights management in broadcasting

Before discussing digital rights management, it may be useful to have a look at the problems of rights management in a historical perspective.

In this context we understand "rights" to mean "intellectual property", and in particular the rights related to the use of copyrighted works in a broadcasting environment, specifically in the radio and television transmissions of Yleisradio Oy (YLE).

Broadcasting companies are mass users of protected works, and they acquire the rights required in a number of ways, depending on legislation and business practices. Just to give an example, last year YLE used protected musical works in its transmissions more than 800 000 times. Each transmission was reported in great detail to the rights owners.

Radio and television broadcasts consist of audio-visual works such as films, documentaries, plays, recorded music etc. It is typical that many persons and companies are involved in the creation of a single work, and transmission rights must be obtained from many different sources. Even the broadcasting of one film or record typically requires rights clearance from two independent sources.

YLE and other broadcasters acquire the right to use copyrighted music through a long-established system of collective licensing. In order to transmit a single piece of music, the broadcaster usually needs a contract with two separate organizations. Composers and lyricists in most countries have granted national collecting societies such as Teosto the right to license their works to broadcasters and other users (such as concert promoters). A special provision in copyright law grants Teosto the right to represent even authors who are not members of the organization. In turn, YLE pays an annually negotiated fee for the use of musical works and reports each transmission to the collecting society.

Only in special cases, such as complete operas or musicals, YLE has to negotiate the transmission rights directly with the original authors. On the other hand, the rights of composers and lyricists expire 70 years after their death, and consequently there is a lot of music, which is in the public domain and can be used freely. Public domain works have a significant role in classical music broadcasting, and they must also be identified properly. In the case of recorded music, the rights of the composers and the performers must be identified separately: the performance may be protected although the work performed is in public domain, or vice versa - all combinations are possible.

When YLE broadcasts recorded music, the performers and producers are also entitled to compensation. Legally, this is based on a system of compulsory licensing, as defined in copyright law and international conventions. The owners of these so-called "neighbouring rights" do not have the option of denying the broadcasters the right to transmit, but in this case, too, detailed reporting is required.

In the case of ”live” music the situation is different, for instance in the context of a television show. The composers are again paid through their collecting society, but the musicians performing in the studio have an employment contract with the broadcaster, and they are paid on this basis. However, if the performance is taped and rebroadcast at a later date, or if it is used in other media, intellectual property rights are also involved, and the performers are paid again on a different basis. YLE has collective agreements with translators, free-lance journalists, actors, musicians and other creators and producers, which are usually a combination of employment contract and transfer of rights. All this requires a huge amount of manual (and computer-assisted) work.

2.1. The ISRC code

When a broadcaster invites an author to the studios to read her works, it is directly in contact with the owner of the rights. However, such situations are not typical. Usually broadcasters acquire broadcasting rights indirectly from various sources. The physical carrier (tape, disc) comes from one source and the rights from another. This often leads to problems of identifying the works and their owners properly. It is quite common that there exist several different compositions with the same name. The names of authors may be misspelled. As a consequence, major broadcasting companies have large departments involved with the cataloguing, archiving and reporting broadcast materials.

Over the years, there have been various plans to overcome this problem. In some countries and in some fields of business (for instance, local radio) it may be difficult to get users to supply all relevant information on works used. Why not collect this information at the receiving end? Audio fingerprinting [1] is a technology where a computer matches an incoming broadcast signal to a database of recorded music, and compiles a list of recordings used in broadcasts. This technique is simple enough in "Top Ten" radio where a limited number of well-known recordings are played in rotation. However, last year YLE broadcast about hundred thousand different recordings on its various channels, many of them once only. Audio fingerprinting is not capable to separate different recordings of classical music, when there exist hundreds of different recordings of the same works to choose from. The only way of creating such a database would be to do it while material is transported in the archive database.

If audio-visual works could carry in some way information identifying both the works and their performers, this could potentially result in great savings both for broadcasters and the rights owners. There are already several plans to identify works by codes. One of the best known such schemes is ISBN [2]. The ISBN (International Standard Book Number) is a unique machine-readable identification number, which marks any book unmistakably. This number is defined in ISO Standard 2108 [3]. The number has been in use now for 30 years and it is widely used also in the book-trade and in libraries. 165 countries and territories are officially ISBN members. The ISBN accompanies a publication from its production onwards.

The number consists of ten digits:

- Group identifier
- Publisher identifier
- Title identifier
- Check digit

In 1992 IFPI, the International Federation of the Phonographic Industry introduced a similar code for published sound recordings, ISRC , the International Standard Recording Code [4], [5] (ISO 3901). Its importance is obvious. It is common that recorded music is marketed in many different formats. The same performance may appear on CD single, single-artist CD album and compilation CD, and the recordings may carry different commercial codes (catalogue numbers) in different areas. Yet the rights related to the recording are the same in all cases. On the other hand, and artist may record the same compositions several times in the course of his career; in this case the legal status of the performances is not identical.

The ISRC code follows the ISBN in most respects. In addition to country, publisher and recording, it also has a Year of Reference Element which identifies the year in which the ISRC is allocated to the recording. The code is not visible on the disc, but it is embedded in the subcode area of a compact disc, and it can be easily read with today's CD-ROM technology.

The ISRC code has not had the success it deserves. Some years ago YLE conducted a study [6] of the use of the code, and we noted the following problems:

1) At the time the code was introduced, the equipment needed to read it was not widely available. It was difficult for users to check the codes for eventual errors.

2) The code was not embedded in the audio signal itself, but in the subcode area of the CD. This could be also seen as an advantage, as there is no risk of the code interfering with the audio signal, but it also means that there is no possibility of reading the code from the sound when it is broadcast, or moving it automatically into another carrier when a recording is reissued in another format.

3) Not all record companies were using the code. When the code was used, there were often logical or factual errors in the codes. Reissues were often given new ISRC codes while the logic of the system demands that the original code follows the recording as it is reissued. In several cases recordings which were in the public domain were given new ISRC codes which appeared to indicate that the recordings were still protected.

4) The national branches of IFPI allocated country and company codes, but the allocation of recording codes was left to individual record companies. There was no central organisation, which would document all the recordings published with ISRC codes. Even if we knew the ISRC code of a specific recording, there was no way to obtain documentation related to this recording in machine-readable (or other) form.

It is typical of the recording industry (and other audio-visual industries) that small companies come and go. There are many short-lived companies. As companies go bankrupt or cease operations, the information related to their publications often disappears. This contrasts with standard practice in the book-publishing world, where national ISBN centers collect, compile and publish information on books.

5) We were especially interested in using the ISRC code to report broadcast music to the collecting societies involved. If we could just report the ISRC codes of all recordings played on the radio, and the right owners could match this code to a register of rights owners, this could potentially result in great savings for both parties.

However, it soon became apparent that there had been no co-operation between the various organisations involved in music rights. Gramex, the collecting society representing performers and producers, was tentatively interested in the idea. As part of its activities, Gramex has in any case to compile information on published recording, and it might eventually have developed into a national "ISRC centre". However, Teosto, the society representing the authors of the musical works had no plans to use the code. Even if we could have reported music to Gramex by using the ISRC code, we would have had to report the same music again to Teosto using the old system. This would have resulted in additional costs rather than savings.

The ISRC code illustrates some of the problems involved in digital rights management. The system planning stage is at least as important as the technical solution. The advantage of giving all audio-visual (and other) works unique identifiers is clear enough. However, audio-visual creations are often packages of several works, which complicates the matter. Unless all interested parties agree on the system, it is unlikely to be widely adopted.

3. Requirements for DRM in broadcasting

Today, broadcasters and copyright societies have well established tradition of negotiations in rights usage. All use of the protected items is based on the contracts. From a broadcasters point of view a DRM system should do something useful and hopefully save some money.

So far the main interest of establishing DRM has been related to control and preventing copying of protected material by consumers. In case of broadcaster where rights for usage are based on collective agreements the problem is not related copying of material, which is based on separate agreements or right to do ephemeral recording.

The main reason for a broadcaster to be interested in DRM is how it could help (or automate) reporting transmitted material and how DRM technology could be used to exchange rights information between different IT-systems used for broadcasting. Of course DRM could be used to protect material where broadcaster own rights, but this is a complex issue at least in a case of public broadcaster, whose income is based on TV license fees. Anyhow, this question can be considered as similar case as we see with recording or film industry. Still, the role of public broadcaster may require that it broadcasts information about public safety hazards etc. If the transmission is normally protected this function could not be implemented as effectively as in a case of unprotected transmission. In a case of YLE this is required by the Act of Yleisradio Oy [7].

Current ideas about DRM are mainly focused on how to prevent the usage of copyrighted material. The current tradition in radio broadcasting is based on reporting after transmission. From broadcaster point of view there are no reasons to change current situation.

This is not true in the case of TV where rights are bought separately from each right owner before each transmission.

The other more technical reason to avoid preventive DRM technology is based on the high availability requirements in broadcast service. Technical requirements for broadcast systems are quite different from consumer electronics. For example in a case YLE IT-infrastructure consist of around 5000 workstations and 500 servers, which are mainly running Microsoft operating system. In a current situation where security patches are becoming a weekly habit, testing patches and updates is a critical issue.

Listeners and viewers expect very good availability from broadcasting services, even in the most exceptional conditions. In the case of YLE this is one of the corner-stone of the public service and it is even by the law [7]. During the transformation from analogue production systems to digital ones, the production environment has changed so that soon the only traditional broadcast equipment are microphone and loudspeaker or camera and the display.

3.1. Metadata

Rights information is one part of the metadata. In broadcasting, full metadata set could include many types of information, such as [8]:

- technical metadata (e. g. timecode);
- preservation metadata (quality of copy)
- documentary metadata (manuscript, news text)
- transactional metadata (loan period)

Core metadata standards such as Dublin Core [9] does not cover all this information and still it could be useful to be able to include this information while material is transferred f. ex from one broadcaster to an other. Broadcasters have developed several standards for metadata used in broadcasting, [10].

In broadcasting the amount of the rights information can be also large f. ex in case TV or radio drama. Rights information has one feature that separates it from other metadata types - it is dynamic. Compared to other metadata, which is created when material is catalogued rights are changing due to limited life span of rights and due to business reasons. In a case of music the dynamic nature of rights is not so big issue, because collecting societies operate as rights clearance centers. This is not true in a case of TV and this feature adds a new complex problem in rights management process of the broadcaster.

The main use of metadata is of course to help searching. Due to this metadata has to be stored in database. The most obvious way to lock metadata and essence to each other is create unique identifier such as ISBN number. So far these unique identifiers have been unique to each media archive. This creates a problems while material is transferred from one archive to an other. If we could transmit the metadata as a part of the essence in a standardized manner, this would ease up material exchange considerably. For example, European Broadcasting Union (EBU) has developed a file format MXF [11] for video material exchange in conjugation with Society of Motion Picture and Television Engineers (SMPTE).

It could be another way to solve the material exchange problem by using steganographic methods to transmit metadata as a part of the essence. This could be feasible for example in a case satellite transmissions or current material carriers such as video tapes where possibility to include files is limited. Exchanged material could be also fingerprinted to guarantee that possible illegal copies could be tracked down afterwards. Problems related to privacy issues could be neglected, because parties involved already have business relation and contract with each others. Same technology could be used to guarantee communication security. These systems have to be designed so that they fulfill Kerckhoff's principle [11]. It says that an encryption system should be designed so that the design of the system is public and the only unknown part is the key. This guarantees that decryption can be made by any instance who knows the key. This is a must, if we consider requirements for permanent archiving.

Even if the amount metadata can be quite large it is still nominal compared to the essence. This due to the fact that quality of the material used for broadcasting has to survive post production without quality loss. In practice this mean that the data rate of audio and video material in production systems has to be much higher than the rate of transmitted material. Anyhow post production requirement of the material could create some difficult unpredictable problems due to that the watermark could become audible or visible in post production process, while signal is distorted. The other obvious problem is that watermark could become unrecognizable or it would be accidentally removed. This same problem exists in a case of highly compressed material [13].

3.2. Rights reporting in radio

The history of reporting transmitted music is in a case of YLE as long as the history of company. As a manual operation this task was so large that when IT-systems became mature enough it was obvious to apply them to reporting.

Before computer assisted reporting became possible creation of a catalogue database was needed. Computer assisted reporting was launched in 1990 and this applications is one of the few remaining mainframe applications in YLE.

History of computer assisted rights usage reporting is much longer than computer aided radio (CAR) systems.

These systems became in production in the second half of the 1990's starting from program flow type format channels and radio news. We are currently moving to fully IT-based production in our classical music channel which also broadcasts radio dramas. It is obvious that radio channels having different type of program profiles need different production systems.

Before the launch digital radio archive [14] the music has been digitized or transferred separately in each computer assisted radio production system. Also the metadata has been created separately in each production location. YLE has transmissions from over thirty systems in twenty locations all over country. One of the reasons to do digitization in each location separately has been the capacity WAN connections, but different stations have been keen on keeping their own music profiles also.

For music reporting from CAR-systems we had to develop an in-house application to link existing reporting application to CAR-systems. Due to inconsistency in metadata, the reporting still needs human interaction. Actually, Teosto and Gramex would approve direct reports from widely used RCS Selector [15] music selection program. In our case, however, we see the virtues of having centralized information in

music usage of the company more important than saving some time in certain channels.

We are currently developing a new version of the rights reporting system and the goal is to automate reporting even further. This has become possible due to the launch of a digital archive, which makes it possible to have same metadata in each CAR-system.

3.3. CD copy protection

Today, YLE radio archive acquires around 6000 CD titles annually. We are currently discussing what kind purchase policy we will have in future due to the launch of the digital radio archive.

Currently, two copies of each title are acquired and transfer to archive is done only for those titles that are requested. We consider that we could change the policy so that only one copy would be bought and all material would be transferred. This would already mean extra cost due to the fees of permanent copy creation. Rate of this fee is currently around two times the price of an actual CD.

Other more unpredictable cost is the extra work caused by copy protected CD's. So far, the only sensible way of transferring a copy protected CD is to digitize it from analogue playback signal. Currently, transfer of CD takes around five minutes and it can be easily automated via CD robot. In case of a copy protected CD, length of the digitization is more than actual length of a CD. It also requires manual work for pause insertion and there are no automated systems available for this. This means that the time for digitization is more than ten times longer compared to the digital transfer in current situation. In our case this means significant increase of workforce. The current estimate is that if we digitize all CD's and 50 % of them are copy protected we need around two man years extra workforce.

3.4. System integration and DRM

One of the main advantages of having centralized digital archive is the possibility to have only one location for digitization and metadata cataloguing. One of the big challenges in Digital Radio Archive project was system integration between the archive and existing CAR systems. For a broadcaster the main reason to have digital archive is related to the possibility to speed-up and automate processes and have new type of programs where old material is reused. Of course, for a national broadcaster, the creation of a permanent storage for historical material is required.

Today most ON-AIR systems are traditional client server systems and their most important feature is that they work. We know from our experiences that things like virus scanning can create interesting problems to these systems. To have DRM components installed in these systems does not sound like a good idea.

4. The ideal DRM for broadcaster

An ideal rights management system - digital or otherwise - would require

- wide agreement by all parties involved
- a method of matching identified works to a database or databases containing relevant documentation, which would be available (not necessarily free) to all users

An ideal system should also keep track of material which is in the public domain. For a classical music broadcaster, for instance, the proportion of public domain material is so large that it has a considerable economical significance. Unfortunately not all DRM systems recognize this, and there is a tendency for republishers of public-domain material to claim (specifically or implicitly) rights to material which is, in fact free.

Present agreements in the broadcasting field do not recognize the possibility of right owners giving their material free of charge for broadcasting purposes, for instance for promotional purposes. However, one should also take into account this possibility in the audio-visual field, following the preceedent of "copyleft" software [16].

Creating such a system would require some investment in the necessary infrastructure. However, the advantages to major users are so obvious that it should be possible to induce them to make the necessary investments in return for future savings. The openness of the system and wide availability of information would also function as a safeguard against errors - unintentional or intentional.

If DRM is going to be used in broadcasting it should not create new threats to availability of the production systems. As we see the DRM as a way to ease up reporting of the transmitted material and method to help material exchange. Either of these applications does not require installation of new components to all of the workstations.

From technical point of view an ideal DRM system should be open and it cannot have unknown features, which could create more problems than solve them. DRM technologies like watermarking and fingerprinting would be most useful for a broadcaster if they designed so that their technical features are known. This of course lowers the level of security, but in a case of well established operators this should not be a problem.

From archivists' view point DRM is one technical detail, which is going to change to a new one time to time. DRM system which features are unknown creates serious risks for a permanent archive. These problems are typically mixed operational and technical as in the case of copy protected CD's. Anyhow, we think that one should not archive material permanently in a file format, which technical details are not public.

This could lead the archivist in a very difficult situation f. ex. in a case where technology supplier is in bankruptcy.

5. Conclusions

From a broadcaster's viewpoint, DRM is a complex issue. Some analysts see it as a tool to restrict consumers' rights, but as we have shown, certain type of DRM could be very effective tool to rationalize broadcasters operational procedures. We anticipate that these possible new applications of DRM technology could create so large savings that if suitable DRM technology would be available it would be feasible for a broadcaster at least to try it.

6. References

[1] P. Cano, E. Batlle, H. Mayer, H. Neuschmied, "Robust Sound Modeling for Song Detection in Broadcast Audio", AES 112th Convention (2002), Munich, Germany.

[2] The International ISBN Agency, http://www.isbn-international.org/

[3] ISO 2108:1992 Information and documentation – International standard book number.

[4]The International Standard Recording Code http://www.ifpi.org/isrc/

[5] ISO 3901:2001 Information and documentation – International Standard Recording Code.

[6] Gronow: "The sound archivist looks at ISRC", *IASA Journal*, 1996:8.

[7] the Act on Yleisradio Oy, 1993/1380. See English version http://www.yle.fi/

[8] R. Wright, "Some consideration on using P_META and Dulin Core" *EBU Technical Review*, No. 294, April 2003.

[9] http://dublincore.org

[10] EBU Tech doc 3295 P_META (P_META v1.0).

[11] B. Devlin " MXF- the Material eXchange Format – an introduction", *EBU Technical Review,* No. 291*,* July 2002.

[12] A. Kerckhoffs "La Crypgraphie Militaire" Journal des Sciences Militaires, 9th series, IX (Jan 1993), pp 5-38; (Feb 1883) pp 161-191.

[13] Anderson R., Petitcolas F., "On The Limits of Steganography", IEEE Journal of Se-lected Areas in Communications, 16(4): 474-481, May 1998.

[14] Petäjä M., Frilander J., Gronow P., Järvinen A., Digital Audio Archiving in Public Broadcasting, AES 20th International Conference proceedings, pp. 71-82, October 5-7, 2001, Budapest, Hungary.

[15] http://www.rcsworks.com

[16] http://www.gnu.org/copyleft

FORMALISATION OF DIGITAL RIGHTS MANAGEMENT: A NEGOTIATION SCENARIO

Jaime Delgado, Isabel Gallego, Eva Rodríguez

Universitat Pompeu Fabra (UPF), Departament de Tecnologia. E-08003 Barcelona (Spain)

ABSTRACT

One key issue when dealing with multimedia content systems is taking care of possible Intellectual Property Rights (IPR) management and distribution associated to the content itself. This is applicable to both mobile and fixed access situations. Digital Rights Management (DRM) systems are being developed to solve this issue. Although many of them already exist, their use is not easy, due to, between other reasons, a lack of interoperability of solutions. The current research trend is to formalise the interchange and expression of digital rights information, and the identification of protection tools.

MPEG, through MPEG-21, is already working on this, by developing standards on IPMP (Intellectual Property Management and Protection), REL (Rights Expression langage) and RDD (Rights Data Dictionary).

After introducing these concepts, the paper presents a formalisation we are proposing on the IPR domain, together with a specific DRM scenario, that of negotiation. The paper finalises with the description of the implementations we are carrying out in this area.

1. MPEG and DRM

MPEG, Moving Picture Experts Group [1], is a working group of ISO/IEC in charge of the development of standards for processing and coded representation of digital audio and video. MPEG has produced the following standards: MPEG-1 and MPEG-2 that provide interoperable ways for representing and coding audiovisual content, MPEG-4 that changes coding to an object-based approach, MPEG-7 for description and search of audio visual content using the metadata concept, and MPEG-21 (still under development) that defines a multimedia framework to support their lifecycle.

The aim of MPEG-21 [2] is to define a multimedia framework to enable content creation, production, delivery and consumption in an open market. The content creator and the content consumer are the focal points of this open framework; this fact will benefit the content consumers by providing them the access to a large variety of digital content in an interoperable manner.

The twelve parts of standardisation within the Multimedia Framework defined in MPEG-21 include Vision, Technologies and Strategy (Part 1), Digital Item Declaration (Part 2), Digital Item Identification (Part 3), Digital Item Adaptation (Part 7) and Reference Software (Part 8), File Format (Part 9), Digital Item Processing (Part 10), Evaluation Methods for Persistent Association Technologies (Part 11), Test Bed for MPEG-21 Resource Delivery (Part 12).

The most relevant MPEG-21 parts for our work are:

- Part 4: Intellectual Property Management and Protection (IPMP).
- Part 5: Rights Expression Language (REL).
- Part 6: Rights Data Dictionary (RDD).

The future recommendations for standardisation related with MPEG-21 multimedia framework are: Persistent Association of Identification and Description with Digital Items, Content Handling and Usage, Terminals and Networks, Content Representation and Event Reporting.

The current status of standardisation in the IPR area is to have approved a FCD (Final Committee Draft) of the MPEG-21 Rights Expression Language and its associated Rights Data Dictionary by the end of July 2003.

2. IPMP

Intellectual Property Management and Protection (IPMP) is one specific term of MPEG for Digital Rights Management (DRM).

MPEG-2 [3] contains a few tools for the identification as well as for the protection of content. For identification purposes there is the copyright descriptor that consists of two parts: The copyright identifier which identifies the type of work, and the copyright number that is the unique identifier handed out by an authority. For enabling protection there are similar provisions that can be used to signal whether particular packets can be scrambled, to send messages that will be used in Conditional Access systems, and to identify the Conditional Access System used.

In turn, MPEG-4 defines two parts of technology too, one for the identification of copyright and one to enable its protection. In the identification part, the Intellectual Property Identification Data Set [4] identifies whether the content is protected by a non-standard IPMP system, the type of content, the Registration Authority that hands out unique numbers for the type of content, and supplementary data. In the protection part, while MPEG

does not standardise IPMP systems, it does standardise the MPEG-4 IPMP interface [5], that consists on IPMP-Descriptors, a part of the MPEG-4 object descriptors that describe how an object can be accessed and decoded, and IPMP-Elementary Streams, that can be used to convey IPMP specific data.

Finally, MPEG-21 specifies an interoperable IPMP framework [2] to protect digital items, with more interoperable IPMP systems and tools. This one includes standardised ways of retrieving IPMP tools from remote locations, exchanging messages between IPMP tools and between these tools and the terminal. It also addresses authentication of IPMP tools, and integrates Rights Expressions according to the Rights Data Dictionary and the Rights Expression Language.

Although the last work done has focused on the protection of digital content, there is a lack of IPMP solutions to provide interoperability between devices and providers of content and services. Because of this fact, MPEG-21 tries to provide a framework for the creation of new services that can be used to support new business models and that meet the needs of all members of the value chain. IPMP has a very important role in the creation of these business models and must provide much more functionality than simply focusing on the content protection.

3. RIGHTS EXPRESSION LANGUAGE (REL)

The MPEG-21 Rights Expression Language architecture is based on the XrML2 Core Specification and Standard Extension [6]. The eXtensible rights Markup Language (XrML) is a general purpose language based on XML, used to describe the rights and conditions for using digital resources. The syntax of REL is described and defined using the XML Schema technology, more expressive than DTD technology. The use of XML Schema in REL allows for significant richness and flexibility in its expressiveness and extensibility.

The most important concept in the REL is the License, that is a container of grants; each grant basically consists of four basic entities:

- Principals: A principal identifies the party to which the rights are granted by information unique to that entity. This information could have associated some authentication mechanism by which the principal can prove its identity.
- Rights: Specifies an action or activity or a class of actions that a principal may perform on or using the associated resource. The right element encapsulates information about rights and provides a set of commonly used specific rights, notably rights relating to other rights.
- Resources: A DigitalWork is a sequence of bits that can be a resource. It represents the content to which rights and conditions are being applied. The DigitalWork type is composed of four different kinds of elements: description, metadata, locator, and parts.
- Conditions: Indicates the source secured repository or device to use when exercising a right. The semantic specification of each different particular kind of Condition must indicate the details of the terms, conditions, and obligations that use of the Condition actually imposes.

4. RIGHTS DATA DICTIONARY (RDD)

The MPEG-21 Rights Data Dictionary, based on a <indecs>rdd proposal [7], comprises a set of clear, consistent, structured, integrated and uniquely identified Terms to support the MPEG-21 Rights Expression Language. The RDD is intended to support the transformation of metadata from the terminology of one namespace into that of another namespace in an automated or partially-automated way with the minimum ambiguity or loss of semantic integrity.

The use of the RDD will facilitate the accurate exchange and processing of information between interested parties in the administration of rights and use of Digital Items.

A Term is the basic unit of the RDD structure, it is defined as a semantic element with a defined meaning and RddIdentifier. A Term may have different Names and Descriptions from different Authorities. Standardised TermAttributes are separated into Direct TermAttributes and Indirect TermAttributes, the latter being attributes of the former.

The Context Model specifies the model through which Terms are introduced to the Dictionary and defines its component Terms. The Context Model introduces five Basic Terms: Context (the circumstances in which Acting occurs, Agent (someone o something that Acts), Resource (someone or something involved in a Context, other than an Agent, Time or Place), Time (the temporal parameters of a Context) and Place (the spatial parameters of a Context).

5. FORMALISATION OF THE IPR DOMAIN

In order to implement applications able to handle Intellectual Property Rights (IPR), it is very useful to have a formalisation of the IPR domain. In this way, we can have a complete view of:

- all the metadata that represent the different information;
- the different events and operations of the whole life cycle of multimedia content, when digital rights exist.

To achieve this goal we have developed an IPR Ontology: IPROnto [8].

The semantic approach we have taken seems a more flexible and efficient way of achieving our objectives than the syntactic one followed by other initiatives as the MPEG-21 related ones. Those initiatives focus on a syntactic approach, the formalisation of some XML DTDs and Schemas that define rights expression languages. The semantics of these languages, the meaning of the expressions, is formalised separately into term-definition dictionaries where definitions are given in natural language, solely for human consumption and not easily automatable. Our idea is then to facilitate the automation and interoperability of IPR frameworks by integrating both parts (REL and RDD).

The proposal has been developed starting from previous IPR related work of our group, the DMAG [9], that ranges from security [10] to automatic IPR negotiation using agents [11,12,13] and the application of a semantic approach [14].

IPROnto describes IPR contracts, actors, intellectual property creations, rights, etc. in order to provide a complete metadata framework.

5.1 Static part of IPROnto

We can see the metadata (or static) part of IPROnto as a tree where elements are related from the bottom to the top. This tree is rather complex, but to give a flavour, Figure 1 shows a sample skeleton.

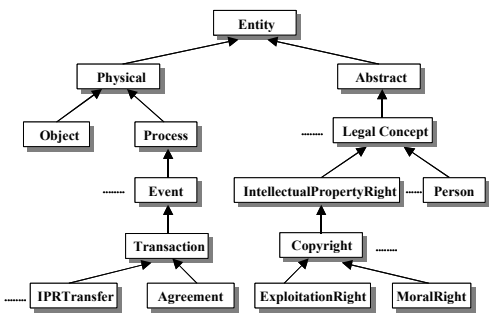


Figure 1. IPROnto key elements for the skeleton

The root of the tree is an Entity, that may be Physical or Abstract. In turn, a Physical entity may be an Object or a Process, being this one more interesting, that might be an Event or a Situation. In the other side of the tree, although several elements may belong to an Abstract entity, only the LegalConcept is presented in the skeleton. Nevertheless, other options, not sketched here, are possible, such as Relation or Quantity. While the LegalConcept might come from a few elements, only two are presented: IntellectualPropertyRights and Person.

The presence of dotted lines in the tree means that other "brother" elements exist but have not been included in the skeleton. All the elements of the tree below Process and LegalConcept, together with all the leaves, are further developed in the ontology.

In this section, we detail, as an example, the Agreement element (see Figure 2), because it is related to the DRM negotiation scenario described in Section 7. The objective of a DRM negotiation is to reach an agreement that must become a Purchase License.

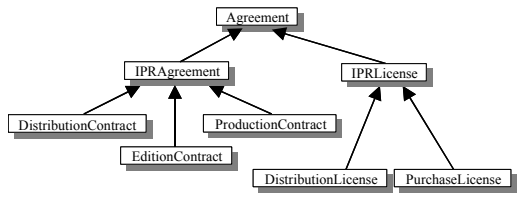


Figure 2. IPROnto Agreement element

An IPR Agreement is an event in which a written or unwritten accord is made between two or more parties. It includes:

1. Distribution Contract: It models an IPR Agreement between a Media Distributor and a Content Provider.
2. Edition Contract: It models an IPR Agreement between a Creator and a Content Provider (Editor).
3. Production Contract: It relates a Creator and Content Provider (Producer). By a Production Contract some Exploitation Rights are transferred to the Producer, for example Reproduction, Distribution, Communication and Translation Rights.
4. Distribution License: A license is an Agreement between two or more parties that does not involve a transfer of rights. The Distribution License is a kind of license in which a Rights Holder authorises a Media Distributor the dissemination of a certain creation in a determined set of conditions.
5. Purchase License: It is a kind of license that appears at the end of the Creation life cycle. It is established between a Customer and the Media Distributor. The Purchase license authorises a determined use of the Creation under certain conditions.

5.2 Dynamic part of IPROnto

Our semantic IPR approach also allows to express a multimedia content life cycle, that is the basis for the active (or dynamic) part of IPROnto. Figure 3 contains a kind of flow diagram relating the different actors involved in some aspects of the content life cycle with all its events (Create, Rights Transfer, Transform, Distribution

Contract, Distribution License, Purchase License and Use).

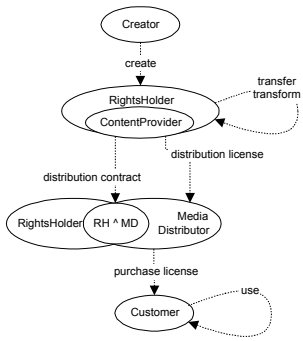


Figure 3. Content life cycle

Next, we give a view of the events that provokes one of the transitions. The events are represented using concepts and relations defined in IPROnto. This is done by means of a graph of nodes, i.e. concepts, and edges, i.e. relations.

The Purchase License Event is the kind of license that appears at the end of the creation life cycle. It is established between a final user, the customer, and the distributor. The license authorises a specific use under certain conditions. Figure 4 shows a graphical view of the event.

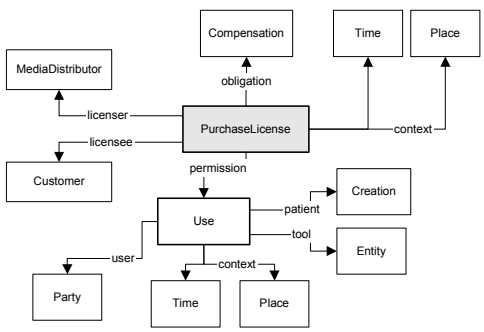


Figure 4. Purchase License Event

6. AN IMPLEMENTATION

In order to fulfill our objective of implementing a rights protection and management system to prototype and check MPEG-21 standards under development, we have started from previous and ongoing implementations we have developed in other projects.

The initial work was in the MARS project, where we first developed a system for e-commerce of video content, in which we initially focused on data localisation aspects following different metadata techniques for its implementation. Then, we added copyright information to that metadata, and we included watermarks in the content itself. The results of this project were our main starting point for the work presented in this paper.

The next step has been to implement an agents and ontology framework in the context of the NewMARS project, an extension of the previous project mentioned before and integrated as a subproject of the AREA2000 project[1]. A specific DRM negotiation scenario, that is described in the next section, has been implemented with mobile agents [15].

Now, we are in the process of adapting the system to the referred MPEG-21 standards, including IPMP (with watermarking based protection tools) and use of the REL and RDD current drafts for rights management.

7. DRM NEGOTIATION

In order to show the functionality of IPROnto, we describe how we implement a specific DRM negotiation scenario, that is taken from the MPEG-21 "Use case scenarios" document [16]. In this scenario, the user is a web designer that has decided to use a specific image in her current web work. She wants to locate a specific version of this image and acquire the necessary digital rights to use it.

The phases of this scenario, according to the NewMARS implementation, are: user interaction, search, negotiation, outcome presentation and control. A sequence diagram of them, except for the background control phase, is shown in Figure 5.

Once the user agent has selected a reference to a provider of the image it is looking for, the negotiation to obtain it begins. The negotiation protocol is obtained from the agent platform, where it has been previously registered.

First, the customer agent issues a call for proposals referred to the desired image. Then, the licensing agent responds with an initial offer if it has the requested content, a refusal otherwise. Given that we are considering a totally automatic scenario, the user agent analyses this offer and decides what to do afterwards. If it does not accept the offer conditions, it can formulate a counter-offer.

The same applies for the licensing agent when it receives the counter-offer. This interchange of counter-

[1] Project supported by the Spanish government (TIC2000-0317-P4-05).

offers continues till any of the parties abandons the negotiation or an agreement arises.

Finally, when any of the parties agrees with the last offer, the other party can also agree and an agreement is reached. An electronic contract is produced, i.e. a RDF/XML document. It contains the agreed conditions and two extra elements pointing to both license consenters. Both parties digitally sign it with XML Signature [18] and the result is a license that authorises the costumer to use the negotiated content under the stated conditions [11].

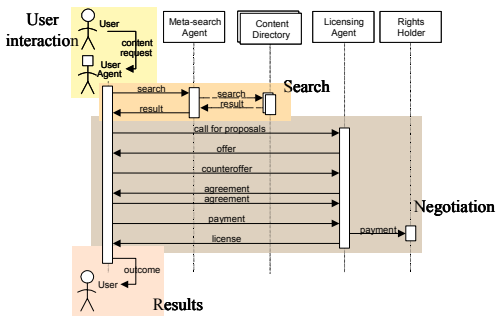


Figure 5. Negotiation Scenario

8. CONCLUSIONS

MPEG is specifying an IPMP for identification and protection of the multimedia content, a REL, based on the XrML, to provide flexible and interoperable mechanisms to support transparent and augmented use of digital resources in a way that protects the digital content, and a RDD that comprises a set of clear, consistent, structured, integrated and uniquely identified Terms to support the REL.

We are proposing a formalisation of the IPR domain by the specification of an ontology, that follows a semantic approach instead of the syntactic one followed by current work. Starting from ongoing implementations, as described in section 6, we are progressing on an integrated system for verification of these standards.

10. REFERENCES

[1] MPEG, http://mpeg.telecomitalialab.com

[2] MPEG-21, http://mpeg.telecomitalialab.com/standards/mpeg-21/

[3] ISO/IEC 13818-1: Generic coding of moving pictures and associated audio: Systems.

[4] ISO/IEC JTC1/SC29/WG11/N1918 Managing Intellectual Property Identification and Protection within MPEG-4

[5] ISO/IEC JTC1/SC29/WG11/N2614 MPEG-4 Intellectual Property Management and Protection Overview and Applications Document

[6] XrML, http://www.xrml.org

[7] <indecs>2rdd Consortium - Rights Data Dictionary, http://xml.coverpages.org/indecs2rdd.html

[8] IPROnto, http://dmag.upf.es/ontologies/ipronto

[9] DMAG (Distributed Multimedia Applications Group), http://dmag.upf.es

[10] J. Delgado, I. Gallego, and X. Perramon, "Broker-Based Secure Negotiation of Intellectual Property Rights". ISC'01, Springer-Verlag, LNCS vol. 2200, pp.486-496, 2001.

[11] J. Delgado, and I. Gallego, "Negotiation of Copyright in E-Commerce of Multimedia Publishing Material", 5th International ICCC/IFIP Conference on Electronic Publishing, IOS Press, pp. 298-306, 2001.

[12] J. Delgado, and I. Gallego, "Distributed Negotiation of Digital Rights", International Conference on Media Futures, Italy, 2001.

[13] I. Gallego, J. Delgado, and R. Garcia, "Use of Mobile Agents for IPR Management and Negotiation", MATA 2000, Springer Verlag, LNCS, vol. 1931, pp. 205-214, 2000.

[14] J. Delgado and I. Gallego, "Standardisation of the Management of Intellectual Property Rights in Multimedia Content", WEDELMUSIC'02, IEEE Computer Society, pp. 125-132, 2002.

[15] I. Gallego, J. Delgado, R. Garcia and R. Gil, "An Architecture for Negotiation with Mobile Agents". MATA'02, Springer Verlag, LNCS, vol. 2521, pp. 21-31, 2002.

[16] MPEG-21 Use case scenarios v2.0, N4330, Section 15.3

[17] XML Signature Syntax and Processing, http://www.w3.org/TR/xmldsig-core

Comparative Study of Digital Rights Management Systems for Music and Text Files

Marc Fetscherin
Institute of Information Systems, University Bern, Switzerland
fetscherin@iwi.unibe.ch

Matthias Schmid
Institute of Information Systems, University Bern, Switzerland
schmid@iwi.unibe.ch

Abstract

Digital Rights Management Systems (DRMS) are seen by content providers as the appropriate tool to, on the one hand, fight piracy and, on the other hand, monetize their assets. Although these systems claim to be very powerful and include multiple protection technologies, there is a lack of understanding about how such systems are currently being implemented and used by content providers.

The aim of this paper is twofold. First, it provides a theoretical basis through which we present shortly the seven core protection technologies of a DRMS. Second, this paper provides empirical evidence that the seven protection technologies outlined in the first section of this paper are the most commonly used technologies. It further evaluates to what extent these technologies are being used within the music and print industry. It concludes that the three main technologies are encryption, password, and payment systems. However, there are some industry differences: the number of protection technologies used, the requirements for a DRMS, the required investment, or the perceived success of DRMS in fighting piracy.

1. Introduction

Digital Rights Management has recently increased in importance. Although there is quite an extensive array of literature available, only a few works provide the high quality content or in-depth analysis about Digital Rights Management. Some authors, such as Rosenblatt, Trippe and Mooney [1], Allan [2], or Pitkänen and Välimäki [3] focus more on the technology, others such as Cope and Freeman [4], Fetscherin [5], or Heil [6], more on the business aspects, and finally there are authors such as Samuelson [7] or Bechtold [8] focusing on the legal implications of such systems.

Very few empirical studies about the current usage of DRM or DRM technologies exist. Although works from researchers such as Felten [12] or Halderman [13] are very insightful as they analyze one type of protection technology, they do not provide an overview or empirical results about what type of protection technologies content providers are currently using for their digital content and why they are using them.

So far, only one study about DRMS usage [9] has been conducted. It was conducted in 2001 and focused neither on the different usages between the music and print industry, nor on the protection technologies used.

This paper wants to close that gap by providing empirical analyses, results, and conclusions about the current usage of DRMS in the music and print industry.

The aim of this paper is twofold. First we want to establish a theoretical basis and explore shortly the seven core protection technologies for a DRMS. Second, this paper provides the first empirical answers to various questions related to DRMS and the protection of digital content such as: Are content providers protecting their digital content? Which technologies are most commonly used? What goal do content providers want to achieve with each protection technology? Are they satisfied with the current protection? Are they going to enforce their protection and why? What do they perceive to be requirements of a Digital Rights Management System? What is the investment to implement a DRMS? What are the protection capabilities of each technology? How confident are content providers that DRM prevents piracy?

2. Digital Rights Management

2.1 Definition

So far, there is no unique or standard definition for Digital Rights Management (DRM). The Association of American Publishers [10] defines it as "the technologies, tools and processes that protect intellectual property during digital content commerce". According to Einhorn [11] "digital rights management entails the operation of a control system that can monitor, regulate, and price each subsequent use of a computer file that contains media content, such as

video, audio, photos, or text." Finally, Gordan [11] defines DRM as "a system of information technology (IT) components and services that strive to distribute and control digital products." There is an overlap in most of these definitions, which all highlight different DRM components [8] and protection technologies such as encryption or watermarking.

2.2 Components and Protection Technologies of DRMS

The role of a DRMS is to protect and manage intellectual property ownership as content travels through the value chain from the content creators to consumers, and even from consumer to consumer (C2C). As previously mentioned, DRMS include different components and underlying protection technologies. While a detailed in-depth description and analysis of all DRMS components and technologies would be beyond the scope of this paper, a short overview of them is provided in Table 1.

Table 1: DRM Components and Protection Technologies [5]

Component	Short Description / Protection Technology
Access and usage control	Controls who has access to the content and how this content is used. Protection technologies: Encryption (e.g., symmetric, asymmetric), passwords.
Protection of authenticity and integrity	Protects the authenticity and integrity of an object. Different types of objects exist such as digital content, rights owner and user. Protection technologies: Watermarks, digital signature, digital fingerprint.
Identification by metadata	Allows the identification of an object by metadata. Different types of objects exist such as digital content, rights owner and user.
Specific hardware and software for end-devices	Includes all hardware and software used by the end-device through which the digital content is being played, viewed, or printed.
Copy detection systems	Search engines which search networks for illegal copies. Protection technologies: search engines (copy detection systems), watermarking.
Payment Systems	Can also be seen as a certain type of protection technology as it requires user registration, or credit card authentication which require also a trust relationship between the content provider and the customer.
Integrated e-commerce systems	DRM systems must also include systems, which support contract negotiation, accounting information and all other sort of information.

According to Table 1, seven core protection technologies can be identified for a DRMS.

- Encryption
- Passwords
- Watermarking
- Digital signature
- Digital fingerprint
- Copy detection systems
- Payment systems

The aim of this paper is to provide an empirical investigation about the current usage of these protection technologies for music and text files. It further wants to explore the extent to which there are similarities or differences in the usage of DRMS by the music and print industry.

3. Methodology

This paper uses multiple case studies as the basis of our research, as the evidence is often considered more compelling, and the overall study is therefore regarded as being more convincing [14], [15]. Furthermore it uses a survey as the main source of evidence. A three step approach was used to select the appropriate content providers using some kind of Digital Rights Management Systems (DRMS).
In the first step, DRM providers were selected by searching through databases and literature. In the second step, the homepages of these DRM providers were examined in order to find content providers using such systems. The third step involved selecting those content providers which belong to the music or print industry (i.e., providing music files and e-books). All of them were contacted by e-mail and/or phone. As of the deadline of this six months study, 9 companies have completed and returned the questionnaire containing more than 20 questions. The return rate of almost 25% is a sign of the high interest of these companies in DRM related research. Despite this relative high return rate, the total number of respondents (N=9) remains rather low. Thus, statistical conclusions cannot be drawn and only descriptive analysis and conclusions are possible. Nevertheless, the analyses in this paper invoke a number of interesting questions and provide directions for future research.

4. Empirical Results

4.1. Participants per Industry

Unfortunately, only 9 content providers have correctly completed and returned the questionnaire. Six are in the music industry and three in the print industry. Note, for all Figures or Tables in this paper, the total number of respondents is always indicated individually.

4.2. Geographical Origin of Participants

The geographical origin of the respondents is split, with two-thirds coming from Europe and one-third from North America. No DRM user within the corresponding industries has been found in Asia, Africa, or South America. Regrettably, content providers from Australia did not return the questionnaire within the required timeframe of this study.

4.3 Delivery Method Used

When asked how their content is delivered, the respondents gave three possible answers: downloading, streaming, or both. The corresponding results are illustrated in Figure 1.

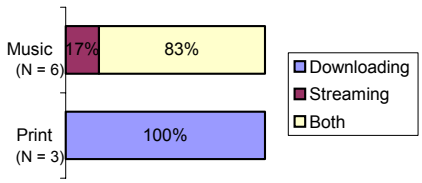


Figure 1: Delivery method used

The majority of respondents from the music industry deliver their content through both downloading and streaming where one is only using. It is not surprising that all respondents from the print industry provide their content only through download.

4.4. File Formats Used

Digital content can be delivered in various file formats. Table 2 summarizes the number of different file formats used for each delivery method per industry.

Table 2: Number of formats used

	Download	Streaming	Total number of different formats
Music	7	6	9
Print	2	-	2

The six respondents from the music industry use a total of nine different file formats for downloading or streaming. All respondents from the print industry use PDF, one of them also employs mpeg. As Table 2 illustrates, the music industry uses many file formats. One possible explanation might be that they are far away from any standard. Another explanation could be that different DRMS support different file formats. On the contrary, the print industry is dominated by mainly one file format which is PDF.

4.5. Protection of Digital Content

The majority of respondents currently protect their digital content (i.e., 78% of all respondents), as shown in Figure 2.

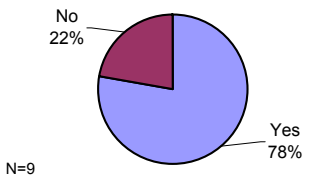


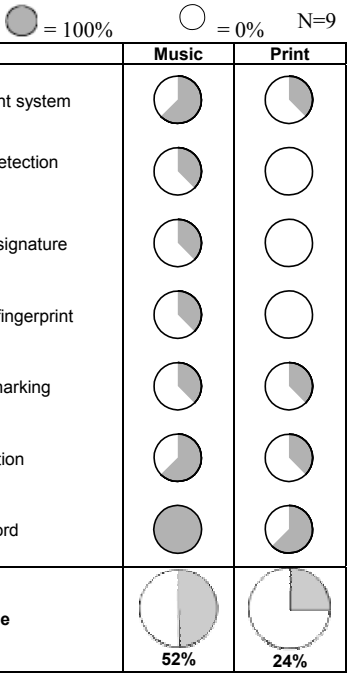
Figure 2: Protecting digital content

There are two respondents – one from each industry - who do not protect their content. Although they are using passwords, they do not perceive it as protection technology. Furthermore, these are the only respondents not currently satisfied with their protection level (see Figure 4).

4.6. Current Protection Technologies Used

According to Table 1 in the previous section, there are seven core DRM protection technologies used today. The first column of Table 3 lists the various protection technologies. The participants have been asked which protection technology they use to protect and/or track their digital content. Table 3 shows the current usage of these technologies by industry. The figures at the bottom of Table 3 represent a theoretical average of the protection level per industry.

Table 3: Protection technologies used by industry



The respondents from the music industry protect their digital content more actively (i.e., average 52%) than those from the print industry (i.e., 24%). It is important to note that each protection technology is used by at least 1/3 of the respondents from the music industry, whereas copy detection systems, digital signatures, and digital fingerprints are not used by the respondents from the print industry.

The three most used protection technologies so far are password, encryption, and payment system. Content providers who do not use a payment system either finance their online offer by paid membership in their offline business or have outsourced the payment process.

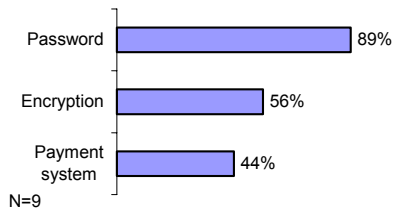


Figure 3: The three most commonly used protection technologies

With the following analysis, we show that each protection technology has its own specific goals to accomplish, independent by industry. Table 4 outlines the various goals content providers want to achieve through the three most commonly used protection technologies (industry independent).

Table 4: Intended goals for top three protection technologies

	Intended goal				
Protection Tech-nologies	Get marketing information	To control access	To improve revenues	To protect against piracy	N=
Password	63%	88%	13%	25%	8
Encryption	20%	80%	20%	80%	5
Payment system	40%	20%	60%	20%	5

The main goal of a password is to control access to digital content (88% of all respondents), whereas 63% of the respondents want to get additional marketing information. By creating a password, the mostly private information is provided through a registration process (e.g., age, gender, income, education). This allows for identification and, at the same time, allows the content provider to get important information to create user profiles. Password is the most suitable technology to obtain marketing information. The main goals of encryption are access control and protection against piracy, both of which were mentioned in 80% of the cases. This protection technology assures that only the customer with the corresponding right (key) has access to the content. After the download/stream of encrypted content, it is not possible to make it accessible to third parties. Finally, the main goal of a payment system is to increase revenues. Although a payment system does not per se increase revenues, it is a prerequisite when attempting to collect revenues.

4.8 Satisfaction with Current Protection

As Figure 4 shows, the majority of the respondents from both industries are satisfied with their current protection.

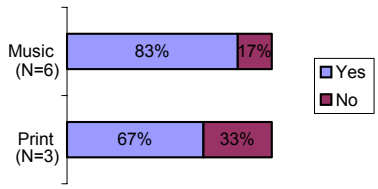


Figure 4: Satisfaction with the current protection

The key arguments for those who are satisfied are that "*consumers do not like complex procedures ... the user experience is still a little clumsy and possibly could be refined further ... cost of litigation is too high*".

Those who are dissatisfied (i.e., one from each industry) are those with the weakest protection within their industry. The dissatisfied declared that the protection, among other reasons, is "*easy to crack ... cumbersome*". Apparently there seems to be a relationship between the number of protection technologies used and the satisfaction level of the respondents. However, due to the small number of respondents, no statistical analysis and tests are possible (e.g., coefficient correlations and significant tests).

4.9 Intention to Enforce Protection

According to Figure 5, most respondents want to enforce their protection in the future. There are some content providers within the music industry which are not sure whether to enforce it or not. The majority of the respondents from the print industry (67%) intend to enforce protection in the future. They do not give any reasons as to why.

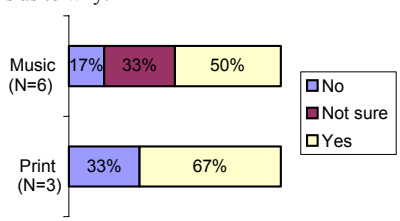


Figure 5: Intention to enforce the protection

At first glance, these results seem to be contradictory to Figure 4. However, one could argue that there is high satisfaction with the current protection, but also a fear of increased piracy in the future. Thus, content providers will need to enforce their protection in order to fight piracy.

Independently of the protection technologies used, content providers will need to either enforce or not enforce their protection in the future. The key arguments to enforce the protection include: "*it is necessary … it is important that we protect our content and convert it into viable business models*". Those who do not yet know, did not provide any reasons or argued that it depends on "*the legal costs to prosecute*". The arguments provided by content providers not to enforce include: "*it is difficult enough to sell protected content … consumers are not yet used to DRM protected content*".

There seems to be a double sword situation: On the one hand, there is a fear of losing control over digital content in the future. On the other hand, too much protection will make the purchasing and usage of purchased digital content cumbersome and will possibly discourage consumers from buying it, therefore disrupting commerce.

4.10 DRM Requirement by Industry

The participants were asked about the requirements they impose for a DRM system. Table 5 outlines the requirements for a DRMS by industry. The various protection technologies are enumerated in the leftmost column of Table 5. The figures at the bottom of Table 5 represent the theoretical average protection level required by industry.

Table 5: Requirement for a DRMS by industry

N=9	Music	Print
Payment system		
Copy detection system		
Digital signature		
Digital fingerprint		
Watermarking		
Encryption		
Password		
Average	**62%**	**43%**

As Table 5 shows, the requirements of a DRM system are perceived differently by each industry.

Only the respondents from the music industry mentioned all protection technologies (i.e., average 62% of the respondents) as being part of a DRMS, with encryption, payment systems, and copy detection systems being the most important.

The respondents from the print industry see all technologies, except for the copy detections system and the digital fingerprint, as requirements for a DRMS. Furthermore, 33% of all respondents are currently using encryption and all of them see it as a "must" for a DRMS. Hence encryption will probably gain in importance in the print industry.

It is not surprising that there is only one protection technology that has been mentioned by all respondents as a core requirement for a DRMS, the technology being encryption. Finally, we can conclude that the average protection level from Table 3 – current protection – has increased in respect to Table 5 – expected protection. Again, this leads us to the conclusion that content providers will enforce their protection in the future.

4.11 Investment Required for the Implementation of a DRMS

Although not all respondents wanted to provide us with their investment amounts, Figure 6 shows the average amount invested (incl. hardware, software, and personnel costs) in a DRMS.

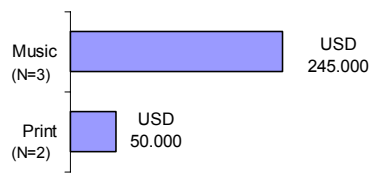


Figure 6: Amount of investments in a DRM system

The respondents from the music industry invested between USD 127.000 and 375.000, whereas those from the print industry invested between USD 37.000 and 67.000. While the respondents from the print industry invested almost five times less than those from the music industry, they perceive the impact of piracy as very low. On the other hand, the respondents from the music industry perceive the impact of piracy to be high (cf. Figure 8).

We could argue that the more content providers perceive piracy as a threat, the more they use various

protection technologies to prevent such piracy (cf. Table 3) and the higher the investment costs (cf. Figure 6). However, due to the lack of data, this proposition cannot be proven statistically.

4.12 Evaluating the Capability of Each Technology to Prevent Piracy

Figure 7 shows the perceived capability of each protection technology to prevent piracy for both industries. The following rating was used: 1=low, 3= medium, 5=high protection capability.
The order in Figure 7 is based on the estimated average of both industries. Thus, the most powerful protection technology used to prevent piracy is encryption, whereas the least effective one is the payment system. The respondents of both industries allocate similar protection capabilities to each technology, with one exception: the print industry respondents allocate a higher protection capability to each technology than the respondents from the music industry. We could argue that the print industry perceived piracy as less of a problem (see Figure 14), yet they are more confident in the existing protection technologies to prevent piracy and therefore are also more optimistic about the potential of DRMS to reduce piracy (see Figure 15). Again, this can unfortunately not be proven statistically and further studies are necessary to prove it.

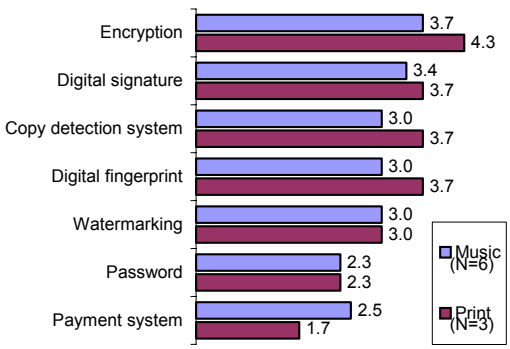


Figure 7: Perception of protection capability of each technology by industry

Although encryption is seen as the core DRM technology (see Table 5), with the highest capability to fight piracy (see Figure 7), it is not used by all respondents at this point (see Table 3). Thus, we can conclude that encryption will gain in importance in the near future. Also, although digital signature, digital fingerprint and copy detection systems are used less frequently than watermarking or passwords (see Table 3), they are seen as being more effective in preventing piracy. Again, one could conclude that these technologies will probably gain in importance in the future and will be used more often by content providers.

4.13 Perceived Impact of Piracy

Each respondent estimated the impact of piracy on his industry *and* on his company. Since the impact on industry and company was estimated to be almost identical by all respondents, the data in Figure 8 represents the average of both estimations. The following ranking was used: 0=no impact, 1=low impact, and 2=high impact.

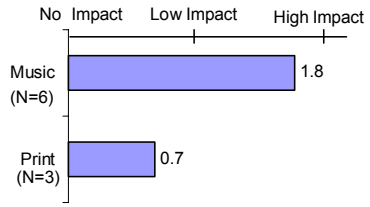


Figure 8: Perceived impact of piracy

The respondents of both industries perceive the impact of piracy on their industry and company differently. These results also reflect the common "mood" in today's media and official reports [16], [17], [18].

4.14 Potential of DRMS to Reduce Piracy

Half of the respondents from the music industry are not convinced that DRM systems will be able to reduce piracy, whereas all respondents from the print industry are (see Figure 9).

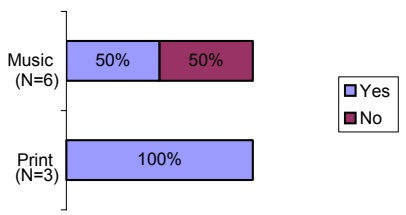


Figure 9:Potential of DRM systems to reduce piracy

The respondents from the music industry that are convinced of the success of DRM, currently use five out of seven protection technologies. Those who do not believe that DRMS will reduce piracy are currently using two or three protection technologies. The majority of them do not use encryption. We can conclude that those content providers from the music industry which use encryption or many different protection technologies are convinced of the success of DRM systems while the non-encryption-users are not. However, this conclusion can not be proven statistically.

The key arguments for the skeptics include: "*the entire media business is changing ... the old conglomerates will die ... most files are out there somewhere ... someone will always find a way to break security*". The optimists say that "*with education on the effects of piracy and an easy way to download music at an affordable price we will be able to convert a good amount of people who engage in piracy to legitimate services ... in a nutshell you have to pay for music - a DRM will ensure that. The music will be priced accordingly then.*"

The respondents from the print industry also are convinced of the success of DRM systems and argue that *"current systems are much harder to hack than those of two or three years ago ... we expect this improvement to continue ... at the same time, we are more proactive in seeking out pirates and taking action against them*."

5. Conclusion

This paper has outlined the seven core protection technologies as being part of a Digital Rights Management System. Furthermore, we have empirically shown that these are also the most used in practice. The top three technologies are encryption, password, and payment system. However, there are also differences in their usage between the music and print industry.

Although most content providers protect their digital content, each industry uses various subsets and combinations of these protection technologies. The music industry utilizes all of them, whereas the print industry only uses password, encryption, watermarking, and payment systems. However, the attempted goals of each protection technology are the same for both industries. Most respondents are satisfied with the current protection. Those who are not satisfied also use the fewest technologies. Those satisfied use encryption whereas all those not satisfied do not use it. Apparently there is a relationship between the number of protection technologies used and the satisfaction level about the protection. Surprisingly and despite the high current satisfaction level with their protection, the majority of all respondents still want to enforce it in the future.

This seems to be a contradictory situation for content providers. On one hand, there is a fear of losing control over digital content which demands more protection in the future. On the other hand, too much protection will make the purchasing and usage of digital content cumbersome and could potentially discourage consumers from buying it, therefore disrupting commerce. The goal must be to have the right technology at the right place for the right product – as strong as needed but as weak as possible.

The requirements of a DRMS are perceived differently by industry. The music industry makes use of all technologies. There is one protection technology which has been mentioned by all respondents - encryption. This is probably due to the fact that encryption is seen by all respondents as the most effective technology to fight piracy. Looking at the investment needed for DRMS, the respondents from the music industry spent five times more than those from the print industry.

Finally, half of the respondents from the music industry are not convinced of the success of DRM systems to prevent piracy, whereas all respondents from the print industry are convinced it will succeed.

As previously mentioned, our analyses are based on a relatively small number of samples and regrettably no statistical analysis and tests are possible. Therefore, further empirical investigations are required in that field in order to test and to statistically prove our conclusions. It is necessary to completely understand which protection technologies are being used, why they are used, how they are used, are consumers willing to accept such protection technologies, to what extent, and finally which strategy (i.e., high vs. low protection) leads to better business results for each industry.

References

[1] B. Rosenblatt, B. Trippe, and S. Mooney, *Digital Rights Management - Business and Technology*. New York: M&T Books, 2002.

[2] A. Allan, "Digital Rights Management (DRM) Software: Perspectives," Gartner Group, Research 2001.

[3] O. Pitkänen and M. Välimäki, Towards A Digital Rights Management Framework, 2001 URL: http://www.hiit.fi/de/drm_framework_iec2000.pdf.

[4] B. Cope and R. Freeman, *Digital Rights Management & Content Development Technology Drivers across the Book Production Supply Chain, from Creator to Consumer*. Altona: Common Ground Publishing, 2001.

[5] M. Fetscherin, "Present State and Emerging Scenarios of Digital Rights Management Systems," *The International Journal on Media Management*, vol. 4, pp. 164 - 171, 2002.

[6] T. Heil, "Digital Rights Management - Konzepte und deren Eignung zum Interessenausgliech im Business-to-Consumer Bereich," Master Thesis at the *Fachhochschule Wiesbaden*. Wiesbaden, 2002, pp. 85.

[7] P. Samuelson, "DRM {and, or, vs.} THE LAW," presented at Workshop on Proactive DRM, Berkley, California, 2003.

[8] S. Bechtold, Vom Urheber- zum Informationsrecht, Implikationen des Digital Rights Management. München: Beck, 2002.

[9] Seybold Research, "Digital Rights Management: Usage,Attitudes and Profile of Users". Foster City: Seybold Seminar & Publications, Foster City 2001.

[10] Association of American Publishers, Digital Rights Management for Ebooks: Publisher Requirements, 2000 URL: http://www.publishers.org/home/drm.pdf.

[11] L. Gordon, The Internet Marketplace and Digital Rights Management, 2001 URL: http://www.itl.nist.gov/div895/docs/GLyonDRMWhitepaper.pdf.

[12] E. Felten, S. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, and D. Wallach, "Reading Between the Lines: Lessons from the SDMI Challenge," presented at Proceedings of the 10th USENIX Security Symposium, Washington, DC, 2001.

[13] J. A. Halderman, Evaluating New Copy-Prevention Techniques for Audio CD, 2002 URL: http://www.cs.princeton.edu/~jhalderm/papers/drm2002.pdf.

[14] R. Yin, *Case Study Research - Design and Methods*, 2nd ed. London: SAGE Publications, 1994.

[15] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students - 2nd Edition*. Harlow: Pearson Education Limited, 2000.

[16] IFPI, IFPI Music Piracy Report, 2003 URL: http://www.ifpi.org/site-content/library/piracy2002.pdf.

[17] J. Wijk, Dealing With Piracy. Intellectual Asset Management In Music And Software, 2002 URL: http://www.eur.nl/WebDOC/doc/erim/erimrs20020930173203.pdf.

[18] L. Takeyama, Piracy, Asymmetric Information, and Product Quality Revelation, 2002 URL: http://www.serci.org/2002/takeyama.pdf.

Contracts and Digital Content

Tobias Regner
University of Bristol
and CMPO
tobias.regner@bristol.ac.uk

Abstract

The paper analyses efficient contracts for digital content, focusing on the music industry. It contributes to the quest for an efficient IPR environment for information goods adding to the literature on copyright. Moreover, it adds an interesting application to the field of behavioural economics.

The model is set in a contract theory framework with the copyright holder being the principal and a consumer as the agent. We offer three contract cases for analysis: a) strong copy protection, b) a low price to compete with copies, c) voluntary reciprocal contributions.

Insights from the economics of information and behavioural economics – information goods have public goods properties; social preferences are significant among individuals – are applied to test the values of a strict copyright enforcement in the digital age.

We find that implicit contracts based on fair, reciprocal behaviour may achieve a first-best allocation of information goods, while explicit contracts are limited to second-best results.

1. Introduction

Copyright law originated in the 18th century[1] and it is regarded as an important cornerstone of successful intellectual property protection. However, recently it has been criticised strongly. Particularly its extensions to now over 100 years granted by U.S. Congress – and followed by other legislations - are widely seen as simply bad policy that misses out on the original intention of promoting the "progress of science and of useful arts" by granting a temporary monopoly.

Moreover, modern information and communication technology makes it increasingly difficult to actually protect the copyright of a digital good. Illicit copies of music files reach billions per year and there seems to be no way to stop the peer-to-peer file trading with reasonable means.

Doubts about its appropriate design and its enforceability bring up the question whether copyright law is still an adequate governance system for intellectual property rights in the digital age. Are there alternative ways of providing information goods, a more efficient eco-system for ideas than the one copyright law offers?

Modified copyright structures[2] give content creators more options compared to the strict copyright law. But why should content creators use them instead of strong protection of their rights, essentially giving up on something that has been granted to them by law?

Once created, the reproduction (or copying) of an information good does not cost any additional resources. Its distribution is also virtually costless. Therefore, marginal costs of information goods are practically zero. They have public goods characteristics: they are non-rival and non-excludable.

However, Pareto-efficient pricing according to p=MC requires an alternative way of rewarding artists for their work, so that at least their basic reservation costs are covered. Otherwise, there would be no motivation to create in the first place.

Social preferences based on fair and reciprocal behaviour might offer such an alternative. The financial reward for the artists is based on a sufficiently high number of fair-minded consumers who contribute voluntarily (if they enjoy the product). The "contract" between the artist and the consumers of his products relies on a trust-based relationship. In fact, fairness and reciprocity might be regarded as the enforcement device of a deliberately left open contract.

Such voluntary contributions for information goods can in fact be observed in reality (a study of shareware software (Takeyama (1994a)), own preliminary research of the voluntary contributions for digital newsletter articles) and can be theoretically explained by social preferences models. Moreover, lab experiments confirm this behaviour in general (Fehr and Schmidt (2000), Charness and Rabin (2002) and several more).

We apply these two insights – information goods have public goods properties; social preferences are significant among individuals – to test the values of a strict copyright enforcement in the digital age. Our tool of analysis is contract theory. Instead of the standard Principal-Agent-situation with a firm and a worker in the labour market, our model features an

[1] It was first enacted in England with the statute of Anne (1709) and then in 1787 as part of the U.S. constitution.

[2] The Creative Commons license, for instance. See: www.creativecommons.org

artist and consumers in information goods markets. We examine three different contract scenarios under information asymmetry and analyse the respective social welfare implications and private investment incentives.

The goals of the paper are twofold. It aims to contribute to the quest for an efficient IPR environment for information goods. We attempt this from a contract theory perspective. Moreover, we want to add an interesting application to the field of behavioural economics.

Our main finding is that implicit or endogenous incomplete contracts may achieve a first-best allocation of information goods, while explicit contracts are limited to second-best results.

Our paper contributes to the literature on copyright. This strand of economic research started with the first formal analysis of copyright by Plant (1934) who in fact rejected the case for copyright mainly on the grounds of a sufficient first mover advantage to establish the product. Landes and Posner (1989) and Besen and Kirby (1989) are main papers with a general welfare approach. Other important works deal with specific aspects of copying. Liebowitz (1985) established the concept of indirect appropriability, Takeyama (1994b) analyses positive network effects from unauthorised copies and Varian (2000) examines the sharing of information goods. Watt (2000) offers an excellent survey of the literature as a whole.

We particularly consider the welfare effects of copyright for digital content. One recent paper – Yoon (2002) – specifies the optimal level of copyright protection in the light of widespread digital copies. However, they do not take maintenance costs of the copyright system into account as we do.

The rest of the paper is organised as follows. Section 2 explains the economic context of the paper. It gives a brief overview of the three strands of the economic theory we relate to: information economics, behavioural economics and contract theory. Section 3 sets up the basic model and derives the main results. We present some open aspects for later versions of the paper and for future research in section 4 and section 5 concludes.

2. Economic Context

2.1 Information Goods and Welfare Economics

Much has been written about the New Economy and the revolutionary effects of information technology on the economy. Much has also been put in perspective by serious accounts of the implications like Shapiro and Varian (1999), for example. However, one thing that indeed is about to change on the way to an informational society is the emergence of a number of goods – information goods[3] – that did rarely exist before. Computer software, digital music or e-books for instance are products of the informational society and their attribute of zero marginal costs of reproduction gives them public goods properties.[4] The use of one digital copy does not diminish the value of any other digital copy. Moreover, potential users might hardly be excluded from consumption.[5]

Generally welfare economics calls for perfect market competition as this achieves optimal allocation of resources, however under certain hypotheses. These assumptions can – by and large – be expected to hold for many products of our economy. This is particularly true – and especially relevant in our case – for very homogenous products like books or music CDs. However, the transition from ordinary goods to information goods affects these basic assumptions. The appropriability of digital goods is seriously in question and they cannot be regarded as private goods anymore.

If we then ask the classic question of welfare economics again for digital goods, the answer will not be so clearly in favour of perfect market competition. Arrow (1962) analysed the welfare implications related to the production of knowledge. He shows that a free enterprise economy will under-invest in research, because the product can be appropriated only to a limited extent. The price set by the market will exceed the socially optimal one of zero marginal costs, one that would make everybody benefit from the research. He concludes that for optimal allocation to invention some organisation not governed by the profit-and-loss criteria – an alternative to the free market - needs to fund research.

Until recently research and its production of knowledge used to be the only commodity that matched the characteristics of an information good, of course being in fact the quintessential information

3 We will also call them digital goods or weightless goods as in related literature, but will focus on the term information goods. Following Quah (2003) they are distinguished from other goods by five characteristics: information goods are non-rival, infinitely expansible, discrete, aspatial and recombinant. More examples include videogames, DNA sequences, news, recipes, sports scores, visual images.

4 Non-rival and non-excludable.

5 Peer-to-peer file sharing networks provide the online community with a huge amount of files for free (among them copyrighted music and movie files). The case of Napster is well-known. However, offshoots that emerged after its demise work without a central file server and also exchange a great number of legal files. Recently a court ruled in favour of two online services and for the first time against the Recording Industry Association of America, recognising the legality of P2P services in a way. See Richtel (2003).

good. As described earlier the New Economy introduces some products either entirely new as software or transformed from ordinary goods like digital music or e-books; all of them are information goods, though.

It is important to stress again the difference between ordinary goods and information goods in terms of the property rights governance here. While our property rights system is designed for ordinary goods – correctly and with a lot of success – information goods require a more nuanced property rights environment to encourage a socially efficient allocation. This system change appears particularly difficult to understand for ordinary goods that have been flourishing under ordinary property rights, but metamorphosed into information goods in the New Economy.

A number of information goods are already being given away for free: E-books, open source software or computer shareware. However, this can generally be explained with positive promotional network effects that increase revenue indirectly and/or a production that is primarily for personal use.

Some authors explicitly offer their e-books for free. The rationale here is a positive word of mouth effect (a network externality) that increases the actual sales of the real book, the ordinary good. The promotion effect is significant and free downloads are massive.[6] What makes this work is the quality difference between the e-book and an ordinary book. They can be regarded as complements, because the reading experience of a real book is so much better than reading the e-book on a screen. People with a high enough quality preference will buy the real book after getting to know it as a free e-book. Voluntary donations are not really intended here as they would bypass the publisher who is required for book production.

Voluntary contributions of code to open source software are intrinsically motivated. Non-academic literature mentions entertainment, challenge and social ties as the main motivation for programmers (Torvalds (2001)). Economically it can be explained with peer recognition concerns and potential lucrative jobs in the future if the coding is successful. (Lerner and Tirole (2002))

Most computer shareware is programmed out of personal motivation: working out a better way for a simple specific software problem the coder encountered. Giving the software away for free supports the public domain with no additional costs. Takeyama (1994a) presents an empirical study of the shareware industry. The software is distributed under a voluntary payment scheme. The main finding of the paper is that the distribution of returns has a positive expected value even when development costs (time) are considered. Therefore, potential voluntary contributions can make it worthwhile to program shareware.

These reasons do not particularly apply to music products. The marginal quality difference between conventional music products (CDs) and the information good music (MP3s) makes them rather substitutes and not complements (as e-books and real books). Positive network effects of free digital music can not be expected to have a significant positive effect on traditional sales, at least not in the long run. Moreover, making music is rather aimed at entertaining other people. It is not mainly for a personal purpose as computer shareware often is (initially).

However, voluntary contributions from consumers like in the case of shareware might provide an alternative reward system to justify giving away music for free.

2.2 Social Preferences

Social preferences explain economic behaviour moving away from the self-interest hypothesis of neoclassic economics. This departure is based on the results of a vast number of experiments conducted in recent years. (see the survey of Fehr and Schmidt (1999)) However, the concept of social preferences goes back to the very beginning of modern economics – in fact, literally even beyond that. Adam Smith already stressed the importance of other-regarding preferences in his "Theory of Moral Sentiments".

Without a doubt economic motivation by self-interest does play a major role. The self-interest hypothesis can accurately explain economic behaviour in many areas. Predictions are particularly fine the more competitive markets are and the more homogenous goods are. This is also confirmed by experiments (Smith (1962)). On the other hand, many economic transactions are not about standardised goods and they are not taking place in a competitive market environment. The more personal the exchange is, the more other-regarding behaviour matters (see Smith (1998) and also Fehr and Schmidt (1999)).

Therefore, social preferences "assume people are self-interested, but are also concerned about the payoffs of others." (Charness and Rabin (2002))

Several formal models have been developed recently to describe the role of fairness and reciprocity. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) both use inequity aversion to model other-regarding behaviour. Models of intention-based reciprocity like Rabin (1993) focus on the intentions of other agents and its impact on behaviour. Social preferences in Charness and Rabin (2002) combine existing theories of fairness and reciprocity and contain three different motivations: an indifference aversion component (agents want to reduce differences between their and others' payoffs), concerns for social welfare (agents like to increase

6 Cory Doctorow's novel "Down and Out in the Magic Kingdom" at http://www.craphound.com/down/

social surplus not just their private one) and a reciprocity part (a desire to raise or lower others' payoffs depending on how nice or not these behaved).

This is the approach we adopt for our model.

The data of Charness and Rabin (2002) comes from 29 different games with 467 participants, making 1697 decisions. Their main goal is to get a better understanding of social motivations and its different types in order to improve formal models that explain social preferences. From the statistical analysis of the experimental results they conclude that all three types are significant, however to a different extent. Social-welfare preferences appear to be the most dominant factor, followed by reciprocity and then difference aversion. While we do not want to discuss specific details of their experiments, one of the results deserves particular spotlight in the context of our paper.

In game Barc7 player A can forgo a (750,0) outcome to give player B the choice between (750,400) and (400,400). Only 6% of the B's choose (400,400) here, while 30% of B's choose this option following either no move or a nasty move of player A. CR conclude the reason for this might be a very strong form of positive reciprocity compared with difference aversion. They conjecture that agents who have just been treated very kindly will not take Pareto-damaging action just to equalize payoffs. They also stress the resemblance to real world situations of this particular game. Although this is the result of just one game and more research needs to be conducted, the relevance of this result to our setting is interesting as will be showed later.

2.3 Contract Theory Framework

Our model framework is based on general contract theory, with the copyright holder of an information good as the principal and a consumer as the agent. Standard contract theory (as in models for the labour market with a firm/manager as principal and a worker as the agent) deals with information asymmetry. The action the agent takes (e.g. effort) usually affects the output, but cannot be contracted on. The output, which is determined by effort and some randomness, is used to write a contract to create incentives and make the agent exert optimal effort. In our setting there is no production function with a randomness term involved. Information asymmetry causes non-contractibility of the payment, not of an effort. This makes our principal agent situation somewhat more straightforward. The principal contracts directly on the action of the agent - if he is able to observe and verify the action, that is.

The simple relationship between copyright holder and consumer is based on the principal contracting the agent to make a payment in exchange for the utility of consuming the music. We will see that this contractual relationship is very trivial for ordinary music goods, but far from that for information goods of music.

Moreover, we integrate insights from the incomplete contracts theory in our framework. We compare explicit contracts that specify all aspects of the relationship with implicit contracts that are much less defined. These endogenous incomplete contracts may outperform explicit contracts in combination with reciprocally fair behaviour of agents. This is based on strong experimental and theoretical evidence from Fehr and Schmidt (2000).[7] They apply their model of inequity aversion to an experiment featuring a manager as principal and a worker as agent. Contrary to the prediction of the self-interest hypothesis implicit contracts are offered by the principal and reach a higher effort level than explicit contracts.

3 The Model

The model describes the relationship between a copyright holder and a consumer from the perspective of contract theory. We consider the transition process from a traditional music industry with ordinary goods to a music industry in the New Economy featuring information goods. Therefore, we distinguish between four different contract scenarios.

The music market with ordinary goods allows for complete contracts. The product – a CD – is standardised and the market relatively competitive. The transaction process of getting the product and paying for it is observable and enforceable. Naturally, contracts are explicit.

In the digital world with information goods this transaction process becomes difficult to observe and we move to an incomplete contracts world. Principals can a) continue to write explicit contracts and monitor to enforce them or b) reduce the price in the explicit contract to compete with pirated copies or c) offer implicit contracts that encourage reciprocal behaviour and voluntary contributions.

3.1 The Question of First-Best

Before analysing the four contract variations of the model, we want to focus attention for a moment on the general benchmark of a first best world.

In standard contract theory there exists a certain level of agent action (effort of a worker, for instance) that maximises total surplus. If information is symmetric, complete contracts can be written and the first best can be obtained. Under information asymmetry though, agency costs arise and the optimal incomplete contract induces the agent to exert effort on a second best level only. This logic naturally applies to ordinary goods of the traditional music

[7] A more detailed description and analysis of the experiment can be found in Fehr, Klein, Schmidt (2001)

industry. Similar to the trade off between incentives and risk that reduces the effort of a worker or the costly monitoring scheme that keeps effort at a certain level, the costs of the product would increase, if the payment transaction were not observable and action to enforce paying had to be taken.

However, the first best criterion in markets for information goods is different. Remember that an additional copy can be produced at negligible costs; the marginal costs are zero. In the first best world the price would equal marginal costs as this maximises total surplus. We still have to consider the issue of dynamic efficiency – motivation to produce information goods in the first place (when there is no price charged) – but it is already clear that under information asymmetry a positive price cannot lead to a first best allocation of the information good, only to a second best.

Again, complete contracts – imaginable under perfect information, though not realistic – deliver first-best results as they would allow perfect first-degree price discrimination.

3.2 Set up of the model

Our simple principal agent model describes the relationship between a copyright holder H and a consumer C. The pleasure from listening to the music gives the consumer some utility u, the payment to the copyright holder is denoted as p. In complete contracts this payment p is equal to the price the copyright holder sets, whereas it can be zero in incomplete, explicit contracts when piracy occurs. However, pirated copies cause some disutility d to the consumers as they might be of lower quality and bear the risk of a virus attack. Finally, in incomplete, implicit contracts there is no price but a voluntary contribution v that C can make.

Under asymmetric information H has the option of implementing a monitoring scheme, which costs K. This scheme increases the probability of the agent being convicted of copyright infringement from 0 to q. Getting caught as a pirate means a financial/moral damage of f for the agent as a result from government prosecution. Without a monitoring system in place piracy is impossible to observe and the government cannot take action.

In order to focus on the contractual problem we do not introduce a utility function that distinguishes between monetary payoffs and non-monetary (dis-)pleasures. Thus, we transform the non-monetary utilities u, d, f and express them directly in monetary terms. Agents' payoffs are then calculated in monetary terms.

Furthermore, we assume both principal and agent to be risk-neutral. The participation constraint of the agent is: $u \geq p$. The representative consumer we analyse in the contract cases just fulfils this condition. He is the marginal consumer with u=p

The principle has to invest resources (time, money) to create the good. He could spend his time doing something else and therefore we call this investment his reservation costs R.[8]

There are two stages of the model: one for production, one for consumption. The principal incurs the fixed reservation costs in stage 1 and has to decide whether to produce or not. In stage 2 the good is priced and consumed. The pricing is derived from the different contract scenarios. The costs incurred in stage 1 are sunk and will be ignored in the second stage.

3.3 Contract Designs

3.3.1 Ordinary Goods / Complete Contracts. Under perfect, symmetric information in the traditional music industry framework complete contracts can be designed. The analysis under complete contracts is very straightforward and mainly serves for a better understanding of the bigger picture.

In this situation the principal H has some variable costs of production c. Remember that the ordinary good is not costless to reproduce in contrast to the information good. We abstract from occasional shoplifting and assume that the agent's action of paying for the product is perfectly observable. Thus, a complete and enforceable contract can be written.

The condition for the optimal contract is: $p = c$

The market allocation (perfect competition[9]) with explicit contracts delivers first best results for ordinary goods.

3.3.2 Information Goods / Incomplete Contracts. The following three cases describe the music business in the New Economy where the product is an information good. The implications of this transition for the model are twofold: ordinary goods convert to information goods, complete contracts have to be replaced by incomplete contracts.

The principal now faces a situation of asymmetric information. He does not possess the means to observe the payment transaction easily as he used to in the traditional industry. The payment becomes non-contractible, contracts become incomplete.

3.3.2.1 Explicit Contracts.

a) Strong Copy Protection with a Monitoring System

8 In the related literature this term is also known as 'the cost of expression' (Landes and Posner (1989) or the fixed cost of development (Yoon (2002)).

9 We do not consider the complications from oligopolistic pricing in the music business here in order to focus on the contract issue.

Pirated music is widely available in file sharing networks and consumers can download songs for free. The copyright holder cannot contract on the payment. However, the principal can introduce what is known in the literature as a verification technology. He implements a monitoring system that helps to detect consumers who do not pay, but rather use the P2P software. This investment in verification technology makes copying verifiable at probability q = 1/3. To simplify things we assume that this signal (being caught) is perfect and always results in litigation of the agent in court for copyright infringement.[10] This punishment f is exogenous as it is set by legislation. It is supposed to work as a threat only though as it should keep the agent from shirking/pirating. At the optimal price p* the agent chooses to buy the product, since the risk of getting caught when copying is too high for him.

Naturally, it is costly to implement the verification system. The huge traffic of P2P networks needs to be monitored and tracked which is technologically very demanding.[11] Also the identity of online users has to be revealed by the internet service provider which poses some legal complications.[12] We denote the fixed cost of implementing a monitoring system as K.

Payoffs are (using a representative consumer):

$$\Pi_H = p \quad \text{if } p \le q \bullet f + d \qquad \text{and} \qquad \Pi_H = 0 \quad \text{if } p > q \bullet f + d$$
$$\Pi_C = u - p \qquad\qquad \Pi_C = u - q \bullet f - d$$

If the threat of the punishment is meant to work, the principal must set a price lower or equal to the expected damage to the agent. Instead, the agent chooses to copy, when the price he is charged exceeds the risk of getting punished.

The optimal contract therefore features: $p_{monitoring} \le q \bullet f - d$

b) Low Price to compete with Pirated Copies

Another option for the principal is to accept the fact that digital copies of the product are readily available through file-sharing P2P networks. Illicit copying is tolerated and not actively prosecuted. It follows that the principal does not invest in the monitoring system.

Although pirated copies are for free, they do cause some costs for the consumers. Notice that the quality of consumption is equal no matter if it is a direct copy or pirate copy. It is the transaction cost that is different, though. The quality of the downloaded music file cannot be verified before and it might be a bad recording. As a result the user might want to get another pirate copy. This is time-consuming and his inconvenience increases. The downloader also runs the risk of getting a file that is infected with a virus and which might in turn damage his computer. Moreover, one could also think of moral burdens that come with something not exactly approved by society.

We aggregate these transaction costs in the disutility from copying d, a constant.

All these costs for the consumer appear if the product comes from piracy, they do not if the product comes directly from the principal. Thus, a reasonable strategy for the principal would be to take advantage of this cost difference and offer the product for a very low price that matches the consumer's disutility from copying – as long as this still covers his reservation costs. The pricing should be so attractive that buying the high quality product is more convenient than getting a low quality copy for free.

The copyright holder cannot charge more than the monetary equivalent of the disutility from copying. Otherwise, the consumer will opt to pirate music instead of buying it legally.

Payoffs are:

$$\Pi_H = p \quad \text{if } p \le d \qquad \text{and} \qquad \Pi_H = 0 \quad \text{if } p > d$$
$$\Pi_C = u - p \qquad\qquad \Pi_C = u - d$$

The optimal contract is defined as: $p_{low} \le d$

This explicit contract gives the copyright holder a profit of d. The consumer gets a utility of u – d, which is equal to his reservation utility. Surplus basically shifts to the consumer.

3.3.2.2 Implicit Contracts

10 The first direct legal action against individuals was a lawsuit of the Recording Industry Association of America against four college students who were running "mini-Napsters" or online directories on their computers, facilitating file sharing for fellow students on the university network. They settled and paid between 13,000 and 17,000$. See Harmon (2003).

11 The music industry is very active to develop electronic countermeasures against online piracy; some of them legal, some illegal. See Sorkin (2003) and also Wired (2003)

12 The Recording Industry Association of America is in a legal battle with Verizon – a major internet service provider. It claims recent legislation obligates Verizon to reveal the names of customers if they are suspected of infringement. Verizon argues the law violates free-speech and due-process rights protected by the Constitution. See New York Times (2003).

c) Voluntary, reciprocal Contributions

Finally, the principal can offer the product for free relying on enough voluntary contributions out of consumers' social preferences that cover or exceed his reservation costs. It seems important to stress again that only because of the particular characteristics of information goods he has this choice. This could not work with ordinary goods involved since giving these away is costly, but not giving away information goods.

In contrast to an explicit contract a deliberately left-open contract leaves room for fair and reciprocal behaviour between the agents.

The fact that the principal offers the product for free – despite other options – is regarded as kind behaviour in the eyes of the consumer. A fair-minded consumer – one with social preferences – will recognise and appreciate the effort of the principal and will reciprocate. He contributes voluntarily. Obviously, he will only give a fraction of his actual utility from the song and he will certainly not contribute if he finds out he does not like the music at all. On the other hand, a selfish consumer does not care about the income of the principal nor about any kind behaviour towards him. He does not contribute and free rides.

In the literature of behavioural economics usually a ratio of 60% selfish to 40% fair-minded individuals is assumed (Fehr and Schmidt (1999), Charness and Rabin (2002)) and we adopt this measure. However, certain experiments suggest that reciprocal behaviour of individuals is even stronger when the amount of effort involved in the relationship (known as "earned property rights" (Fahr and Irlenbusch (2000), V. Smith (1998) among others) is taken into account. Moreover, the social, personal transaction between the artist and a consumer instead of an impersonal market exchange with a record label matters, if the copyright holder is the artist.

Again, transaction costs as in file sharing use do not play a role when the principal makes the product freely available on his web site. The direct download from the site of the copyright holder or a licensed intermediary is quick and of high quality.

In order to incorporate social preferences we use a simple two-person model from Charness and Rabin (2002), which we slightly adjusted to allow for positive reciprocity. This model is very conceptual and crude, but it captures the main elements and permits simple applications. More sophisticated formal models exist, but they are very complex and not yet suitable to explain experimental evidence or to be used in applications. For our purposes the simple version appears to be sufficient.

V describes the utility of an agent and is defined as follows:

$$V_C(\Pi_C, \Pi_H) = (\rho \bullet r + \sigma \bullet s + \theta \bullet t) \bullet \Pi_H + (1 - \rho \bullet r - \sigma \bullet s - \theta \bullet t) \bullet \Pi_C$$

where:

$r = 1$ if $\Pi_C \geq \Pi_H$ and $r = 0$ otherwise

$s = 1$ if $\Pi_C < \Pi_H$ and $s = 0$ otherwise

$t = 1$ if H behaved nicely and $t = 0$ otherwise

$0 < \sigma < \rho \leq 1$ is the parameter condition for social welfare preferences with $\sigma \leq 1/2$ to have C not be more concerned about H than about himself.

$\theta > 0$ as the measure for reciprocity

$$\Pi_C = u - v \quad \text{and} \quad \Pi_H = v$$

Furthermore we assume the ratio of fair-minded consumers α to be 40%. The voluntary contribution v of the agent is a fraction g of his actual utility from the music. It is determined by his self-interestedness (ρ or σ) and his tendency to reciprocate (θ). Concerns for the overall welfare and actions of the principal that affect the social welfare (the Pareto-damaging implementation of a verification technology) are therefore integrated in the agents' preferences. To make the model more straightforward, yet not less realistic, we focus the analysis on a single, representative consumer. He contributes with a probability of α.

In order to allow for a discrete not binary choice of v we have to endogenise the contribution decision of the agent. We assume the fraction g of u to be set by the social preferences parameters ρ and θ. It follows that:

$$v = (\rho + \theta) \bullet u$$

Naturally, the implicit contract does not feature a price. However, fairness and reciprocity might be regarded as an enforcement device of the endogenous incomplete contract. No potential efficient consumer is deprived of a benefit with the price equal to marginal cost. Voluntary contributions to the principal can exceed his reservation costs and motivate him to offer his products for free.

The optimal contract is: $p_{implicit} = 0$

DIAGRAM WITH ALL PAYOFFS

3.4 Stage 2: Consumption

We model the demand of consumers in a very basic way, similar to Besen and Kirby (1989) or Yoon

(2002). Consumer's valuations are uniformly distributed over the interval [0,10]. Prices are also confined by this interval. Based on the contract design analysis above they are ranked in the following order:

$$0 = p_{implicit} < p_{low} < p_{monitoring} \leq 10$$

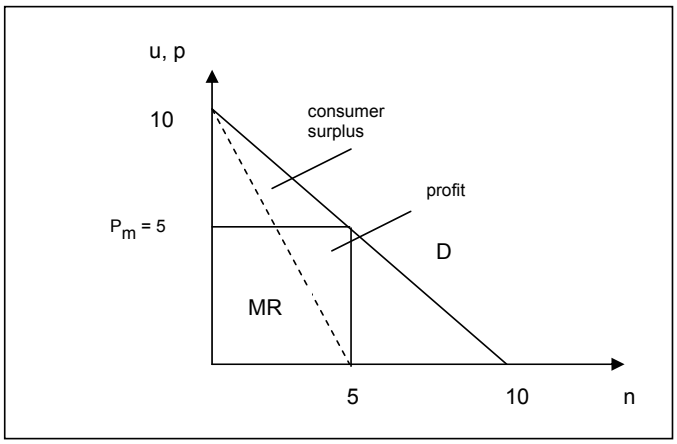


Figure 1: Supply and demand under strong copy protection

Remember that any fixed costs incurred in stage 1 (the reservation cost R or the monitoring cost K) are sunk now and that marginal costs are negligibly low.

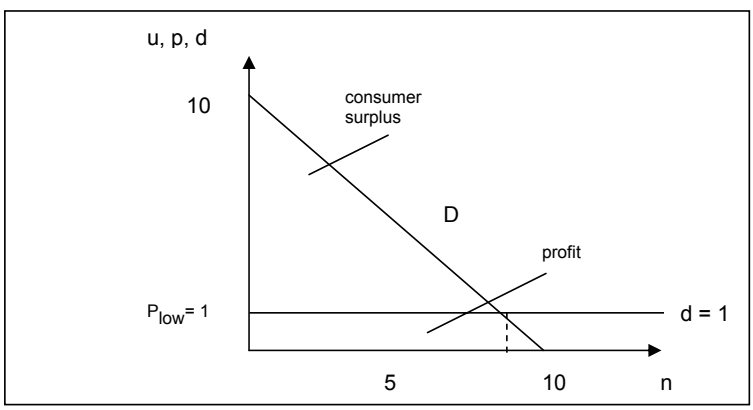


Figure 2: Supply and demand with a very low price

a) Strong Copy Protection

The condition of $p_{monitoring} \leq q \bullet f - d$ defines the price in the strong copy protection scenario. We assume the expected threat of court litigation to be smaller than 10 and larger than d.

The principal uses a monopolistic pricing policy to maximise the profit. However, he might be forced to lower the price in order to fulfil the contract condition (this is not yet implemented).

If the threat of punishment is not a binding condition for the price, then the principal can set the monopoly price of 5. He maximises revenue and his profit is 25. Consumer surplus is 12.5 and the usual deadweight loss results.

b) Competition with pirated Copies

If the principal decides to compete with copies obtainable in P2P networks, his price cannot exceed the disutility agents experience from copying: $p_{low} \leq d$

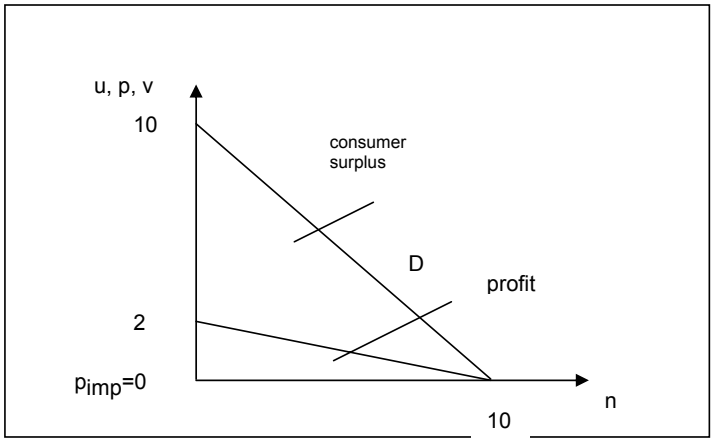
As mentioned before we assume these transaction costs (the virus risk, moral issues, inconvenience from downloading) to be constant across consumers. It is easy to see that in the monopolistic environment for realistically small values of d the profit maximising price the principal chooses will equal d. We assume the disutility from copying to be 1.

Being forced to set a very low price the profit of the principal significantly shrinks. Notice however that no monitoring costs are incurred. With our linear, uniformly distributed demand a price of 1 results in a revenue of 9. The consumer surplus is 40.5. Only a few potential consumers are kept from a beneficial trade.

c) Voluntary Reciprocal Contributions

There is no price charged in the implicit contract scenario. Revenue for the principal comes from voluntary contributions by agents. These payments are determined by the social preferences parameters and the ratio of fair-minded consumers, but also depend on the actual utility each consumer gets.

Figure 3: Supply and demand with voluntary contributions



With a price of zero no consumer is excluded from the benefit of the product. Consumer surplus is maximised, however a certain fraction α of consumers returns $\rho + \theta$ of their utility u to the copyright holder because of their social preferences. Again, we transform α into the probability of every consumer to reciprocate. Voluntary contributions (for common parameter values of $\alpha = 0.4$ and $\rho + \theta = 0.5$) amount to 10 and are the principal's profit. The remaining consumer surplus is 40.

This shows that for certain parameter values consistent with social preferences sufficient income for the principal can be generated. This happens despite the fact that the fraction of purely selfish agents does not contribute. Clearly, this outcome depends on the specific choice of parameters and the precise situation we model has not yet been covered and analysed in experiments. But we feel the related evidence from studies in behavioural economics is encouraging enough and the implications for digital markets are substantial.

We already mentioned one specific game of Charness and Rabin (2002) that delivered surprising results and calls into question the prevailing doubts regarding positive reciprocity shown in conventional games studied. Another of their games, Berk 14, appears to be an even closer fit for the simplified payoffs of our model and underlines the significance of positive reciprocity: Player A chooses between a (800,0) outcome and giving player B the choice between outcomes (0,800) and (400,400). 55% of the B's make the balanced choice here, while only 22% of B's choose this option in the controlled version (a pure dictator game) without a move of player A.

However, it supports the intuition of our scenario. The principal has the choice between two outcomes. One gives him some surplus although limited because of the monitoring costs and leaves not much for the agent. The second lets the agent decide between two options. He can "cheat" and abuse the trust (no payoff for the principal, everything for the agent) or he can share the benefit with the principal by contributing voluntarily. Following the kind first move of the principal (he forgoes trying to enforce an explicit, welfare-reducing contract) the agent is more likely to act reciprocally.

3.5 Stage 1: Production

The detailed implications on the production decision – the private incentives for the principal – will be included in a later version of the paper.

However, it is clear that the principal's profit can exceed his fixed costs at production stage (the reservation cost R and in case a) R plus the monitoring cost K) to motivate investment in all three scenarios.

4 Open Aspects

Several aspects of this early version of the paper have not been covered properly yet.

The integration of a bonus like in Fehr, Klein and Schmidt (2001) makes the contract even more incomplete and thus leaves more room for reciprocity (from both sides). A bonus in the context of the music business could be exclusive access to concerts or backstage, special merchandising for consumers who did contribute. The bonus cannot be specified ex ante, though.

An important question is, who the copyright holder actually is. Is it the record label like in the traditional music industry or the artist? The implications on the fair behaviour of agents (an impersonal market trade or a social personal exchange) and also on the definition of the reservation cost are significant. Related research on the efficient ownership structure in the music industry in Regner (2003) and Clemons and Lang (2003) point out that artists should own copyrights in the digital age.

Moreover, the contribution of the agent might lead to an overall welfare increase. With enough contributions the artist continues to be creative, which is good for the agent and society when future creative works are taken into account. This would increase the payoffs for principal and agent.

Finally, the social welfare analysis has to be generalised and the decisions at the production stage – the private incentives to invest – have to be covered for the different contract cases.

5 Conclusion

The paper aims to integrate the peculiarities of digital information into the social preferences framework. It describes a contractual model based on incomplete contracts theory that provides an alternative way to offer information goods – more efficiently we conclude. Some key parts of the paper are yet to be analysed in this preliminary version. It

opens up an interesting field for future research. Modified experiments to test for social preferences in digital age contexts would be a logical next step.

6 References

[1] Arrow, K. (1962). "Economic Welfare and the Allocation of Resources for Invention", in: Nelson, R. ed., "The Rate of Inventive Activity: Economic and Social Factors", Princeton University Press, Princeton

[2] Besen, S. M. and S. N. Kirby (1989). "Private Copying, Appropriability, and Optimal Copying Royalties." Journal of Law & Economics 32(2): 255-280.

[3] Bolton, G. E. and A. Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition." American Economic Review 90(1): 166-193.

[4] Charness, G. and M. Rabin (2002). "Understanding social preferences with simple tests." Quarterly Journal of Economics 117(3): 817-869.

[5] Clemons, E. and K. Lang (2003). "The Decoupling of Value Creation fromRevenue: A Strategic Analysis of the Markets for Pure Information
[6] Goods" Journal of Information Technology & Management, 4, 259–287.

[7] Fahr, R. and B. Irlenbusch (2000). "Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment." Economics Letters 66(3): 275-282.

[8] Fehr, E., Klein, A. and K. M. Schmidt (2001). "Fairness, Incentives and Contractual Incompleteness", University of Munich working paper

[9] Fehr, E. and K. M. Schmidt (1999). "A theory of fairness, competition, and cooperation." Quarterly Journal of Economics 114(3): 817-868.

[10] Fehr, E. and K. M. Schmidt (2000). "Fairness, incentives, and contractual choices." European Economic Review 44(4-6): 1057-1068.

[11] Harmon, A. (2003). "Suit Settled For Students Downloading Music Online", online at (last accessed 13.05.2003): http://www.nytimes.com/2003/05/02/national/02STUD.html?th

[12] Landes, W. M. and R. A. Posner (1989). "An Economic-Analysis of Copyright Law." Journal of Legal Studies 18(2): 325-363.

[13] Lerner, J. and J. Tirole (2002). "Some simple economics of open source." Journal of Industrial Economics 50(2): 197-234.

[14] Liebowitz, S. J. (1985). "Copying and Indirect Appropriability - Photocopying of Journals." Journal of Political Economy 93(5): 945-957.

[15] New York Times (2003). "Court Says Verizon Must Identify Downloaders", online at (last accessed 13.05.2003): http://query.nytimes.com/gst/abstract.html?res=F00F14F93C590C768EDDAD0894DB404482

[16] Plant, A. (1934). "The Economic Aspects of Copyright in Books" Economica 1: 167-195

[17] Quah, D. (2003). "Digital Goods and the New Economy", Centre for Economic Performance, discussion paper 563

[18] Regner, T. (2003), "Innovation of Music", in "The Economics of Copyright: Developments in Research and Analysis", edited by Richard Watt, chapter 6, pages 104 to 117, Cheltenham, Edward Elgar Publishing

[19] Rabin, M. (1993). "Incorporating Fairness into Game-Theory and Economics." American Economic Review 83(5): 1281-1302.

[20] Richtel, M. (2003). "Entertainment Industry Loses in Net Case", online at (last accessed 13.05.2003): http://www.nytimes.com/2003/04/26/technology/26MUSI.html?th

[21] Shapiro, C. and H. Varian (1999). "Information Rules", Harvard Business
[22] School Press, Boston.

[23] Smith, A. (1759, reprinted 1976). "The Theory of Moral Sentiments" in: Liberty Classics, edited by D. D. Raphael and A. L. Mactie, Liberty Press, Indianapolis

[24] Smith, V. L. (1962). "An Experimental Study of Competitive Market Behaviour" Journal of Political Economy 70, 111-137

[25] Smith, V. L. (1998). "The Two Faces of Adam Smith." Southern Economic Journal 65(1): 2-19.

[26] Sorkin, A. R. (2003). " Software Bullet Is Sought to Kill Musical Piracy", online at (last accessed 13.05.2003): http://www.nytimes.com/2003/05/04/business/04MUSI.html?pagewanted=1&th

[27] Takeyama, L. N. (1994a). "The Shareware Industry: Some Stylized Facts and Estimates of Rates of Return" Econ. Innov. New Techn., Vol. 3: 161-174

[28] Takeyama, L. N. (1994b). "The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities." Journal of Industrial Economics 42(2): 155-166.

[29] Torvalds, L. (2001). "What makes Hackers tick? A.k.a. Linus's Law" in Himanen, P. "The Hacker Ethic", Vintage, London

[30] Varian, H. R. (2000). "Buying, sharing and renting information goods." Journal of Industrial Economics 48(4): 473-488.

[31] Watt, R. (2000). "Copyright and Economic Theory: Friends or Foes? ", Edward Elgar Publishing, Cheltenham

[32] Wired (2003). "Under Cover" and "Dirty Dozen", February 2003 issue

[33] Yoon, K. (2002). "The optimal level of copyright protection." Information Economics and Policy 14(3): 327-348.

Viral contracts or unenforceable documents? Contractual validity of copyleft licenses

Andres Guadamuz-Gonzalez
University of Edinburgh
a.guadamuz@ed.ac.uk

Abstract

This paper attempts to ask the question of whether copyleft Free Software licences constitute valid legal contracts, in particular with regards to the fact that it may create obligations through a distribution chain. There is increasing interest about the license model expressed in popular documents such as the General Public License (GPL), but not enough work has been done in asking perhaps the most important question of all: are these contracts enforceable? Is there really a viral transmission of obligations? To do this the GPL license will be analysed to try to determine whether or not the terms included are contractually valid.

1. Introduction

The issue of non-proprietary software licenses – such as the Free Software (FS) and Open Source Software (OSS) license models – is gaining interest in legal circles, a development that must be welcomed taking into consideration that the phenomenon of open source/free software licensing was initiated with almost no intervention from legal scholars, leaving the legal profession once again to play catch-up in the fast-paced computer world.

Non-proprietary software licenses pose some interesting questions from a traditional contractual law perspective because they create what some authors have defined as a viral contract, a contract that is to be transmitted through a distribution chain. The question must be asked of whether the obligations arising from the initial license are to be considered enforceable, or if any of these contractual terms should be suspect, particularly in jurisdictions where unfair contractual terms are strongly regulated. Surprisingly, these licences are yet to generate any court rulings, so a full study of the eventual validity or invalidity of the contractual copyleft clauses must be subject to an analysis by the academic community, something which has not been forthcoming. The present work will attempt to redress this trend by looking at the contractual validity of the FS licensing (in particular copyleft licenses) as opposed to the OSS model, which is less restrictive and whose contractual clauses are much less likely to generate judicial revision. The author is aware that this may prove difficult in a work of such limited length, and because of the lack of judicial review of the licenses, but the main objective of the paper is to start a much needed debate in this area.

2. Non-proprietary software

2.1. Free Software

It has become increasingly common to read and hear the term open source applied to all types of software developed under a free distribution of the programme's source code.[1] It is important to stress that it is technically incorrect to refer to all of these models of software development as either Open Source (OS) or Free Software (FS), which are the two main types of non-proprietary software, but not the only ones by far.

In general, there are some philosophical differences between both terms. In the strictest sense, the FS concept is centred on the concepts and philosophies of developing programs and distributing them freely [1]. This is not the place to provide a detailed description of the birth of the FS model, [2] but suffice it to say that FS is not new. It has been noted that software sharing is "*as old as computers, just as sharing of recipes is as old as cooking*" [3]. It is vital to note that the meaning of the word "free" in FS does not mean free as in having no price, but rather free as in "freedom" [4]. Stallman defines free software as having the following four characteristics:

- The freedom to run the program.
- The freedom to study how the program works by giving access to the source code.
- The freedom to redistribute copies.
- The freedom to improve the program and release those improvements to the public.[5]

As understood by the proponents of free software, programmers and other developers can charge for the software if it is their desire to do so, but the same underlying freedom behind the software must exist either it is acquired for a monetary fee or if it is not. The user must still be able to have all of the freedoms described, with access to the source code as the most basic requisite [6]. The Free Software Foundation (FSF) goes as far as stating that:

"The freedom to use a program means the freedom for any kind of person or organization to use it on any kind of computer system, for any kind of overall job, and

[1] Source code is the programming statements in a programming language that exists before the program is compiled into an executable application.

without being required to communicate subsequently with the developer or any other specific entity."[7]
This freedom is kept by the adoption of a restrictive licensing model that makes use of existing copyright legislation to protect the source code from proprietary software developers who want to copy it, adapt it and include it in their own programmes. This licensing model will be explained in more detail later.
Open Source is closely related to the Free Software development, but it does contain a different emphasis on the freedoms involved. The term open source was coined during a strategy meeting in February 1998 in Palo Alto California by a group of software developers with links to the Linux operating system [8]. The group met to plan a new strategy in response to the groundbreaking announcement by Netscape that they would be opening their operations and providing the source code of the popular Netscape Internet browser to the public. Netscape decided to do this prompted by fierce competition by Microsoft [9]. They believed that this gesture would give them a precious opportunity to sell the open source software development approach to the corporate world [10].
The need to create a new term to define this viewpoint had become evident because, until then, the prevalent way to describe all output produced by the non-proprietary approach was by using the expression "free software", based mostly on the FS philosophy described. It was apparent to many software developers that this movement had a tarnished reputation in the business world as a result of the more radical ideas held by people linked to the FSF.
In the widest sense, open source is the opposite of "closed source", the traditional proprietary approach to software development in the commercial world. Closed source is software *"in which the customer gets a sealed block of bits which cannot be examined, modified, or evolved."* [11] The main idea behind open source is to provide software for which the source is available for examination, modification and peer-review. The official definition of open source came out of the original meeting, and was based on the Debian Free Software Guidelines, a licensing model that accompanies the Debian GNU/Linux system, a Linux distribution [12]. This has generated an Open Source Definition (OSD), which includes a recommended set of clauses that an OSS license should contain [13].
There are several similarities between OSS and FS licenses. In fact, some OSS licenses have been deemed to be compatible with FS principles, and vice versa [14]. Nevertheless, there are several differences between the FS and OSS philosophies. The main difference is the fact that OSS does not impose in its licenses obligations for derivative software to be kept free – such as the case of copyleft licenses that will be explained later – a practice that has been deemed too restrictive and commercially-unfriendly by its proponents. One of the many complaints that FS advocates make of the open source philosophy is that it is not strong enough in trying to keep software free, and that it simply allows anybody to name their software "open source" even if it is not [15]. This is something that has been partially acknowledged by OSS proponents, which is why they have created the Open Source Initiative (OSI) certification. This certification is given to those licenses that follow the open source definition and provides a certification to inform the public that the software is indeed open source [16]. There are many different OSI certified licenses [17], and it is important to point out that this list includes all sorts of FS licenses that comply with their definitions and guidelines.
Regardless of which definition one prefers, it has become important to use a term that encompasses all sorts of definitions within this development model. The author prefers the use of the phrase *non-proprietary* as an umbrella term that refers to the different sub-categories encompassed by this movement. Another acceptable term is "Libre Software" – now in use by the Information Society Directorate General of the European Commission[18] – as the Spanish word 'libre' does not have the same meaning as its equivalent in English, and encompasses better the philosophy behind non-proprietary development systems. Another valid way of describing this is to refer to Free and Open Source Software (FOSS). The distinction may seem academic, but it is important because the use of each of these terms presupposes a specific development philosophy behind the software. The author also believes that the use of either these three terms is better than to use either FS or OSS of the proposed terms because they encompass all different types of philosophies and distributions, ranging from commercial variations of the non-proprietary model to those that are offered freely to the public.

2.2. Copyleft licensing

From the many different types of FS recognised by most non-proprietary proponents, the most popular type of FS distribution is by means of copyleft licensing, with surveys estimating more than 70% of this type of software uses copyleft licenses as their main contractual mechanism [19]. Copyleft is Free Software with a twist; it maintains the general freedoms awarded to users of free software, but by acquiring a copyleft program, the user has to agree to a license agreement that states that that the software will not be used to develop proprietary commercial applications derived from it [20]. The FSF has a specific definition of what a commercial program is for the purposes of copyleft. According to them, a proprietary program is one that is "*software that is not free or semi-free. Its use, redistribution or modification is prohibited, or requires you to ask for permission, or is restricted so much that you effectively can't do it freely.*" [21]
Copyleft was created from a perceived need to protect the fruits of non-proprietary development. After several

years of producing computer programs with a sharing mentality and offering the code to the public, it became evident that some companies had started using this output in a parasitical fashion, obtaining the source code, tweaking it and selling it as commercial proprietary software with very low production costs [22]. Copyleft became the contractual solution to stop companies profiting from non-proprietary products by distributing software that must remain free.

For GNU software, the recommended contract to use is the General Public License (GPL), which is a standard contract that makes sure that the software code is passed on, but anyone who redistributes the software – with or without changes – must pass along the freedom to further copy and change it. This places a burden to the person transferring the software; the burden is that the software must remain "free", as defined by the FSF and the GPL. This is different from just placing software in the public domain because the work remains copyrighted [23].

The GPL is the main exponent of the legal framework that sustains the copyleft system. It reads as a mixture of a legal contract and an ideological manifesto. The preamble to the work states clearly some of the most common beliefs of Free Software and the non-proprietary approach, with several admonitions about the meaning of the word "free". The main point is that, as has mentioned before, the source code must be made available to the users. The preamble states:

"For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights."[24]

The license specifies that this is achieved by two means: by protecting the software by means of copyright; and by providing the users with a license that gives them the freedom to use and modify the software in any way they see fit. The main body of the license reiterates these ideas. Section 1 for example states:

"1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program."[25]

The section also states that the user can make monetary charges when passing the copy, which is also consistent with the general Free Software characteristic that does not discriminate against commercial software.

Many of the provisions of the GPL can be found in other non-proprietary licenses, including several OSS ones. What makes the GPL unique is the section 2(b), as this is where the restrictions against using the software to create commercial software are specified. The section reads:

"2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions: […] b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License."[26]

What this means is that any software developed by using the open source code of the copyleft programme must not charge for the derivative product, and most importantly, must ensure that the GPL is transferred to further users of the derivative software. This type of license has been aptly named a "viral contract" by Professor Radin, defining them as "*contracts whose obligations purport to 'run' to successor of immediate parties*" [27]. These contracts would then spread in a viral form, as the licensee must make the terms of the original agreement part of any subsequent license they will perform because that obligation is part of the contract, and those subsequent licensees will have to impose the same contractual terms in further licenses that they perform, *ad perpetuam*.

The restrictions imposed by copyleft would seem to go against some of the principles of Free Software by the viral imposition of restrictions and obligations, which denies the very freedom of doing what one desires with the software – and the FS proponents should face the fact that this may very well include the freedom to profit from the subsequent use of the code. The use of non-proprietary software to create a proprietary or "closed source" software may be morally suspect, but one cannot elevate freedom to the highest pedestal and begrudge those who will use that freedom for purposes that are philosophically and politically adverse to those of the creator of the program.

Another conundrum that must be understood is the distinction between contractual enforceability and copyright protection awarded to computer programs. It could be said that copyleft licenses create a double-pronged protection of the software. On one hand it poses contractual restrictions in the shape of a license, in particular by the contractual enforceability of the GPL license and its clauses. On the other hand, works protected by copyleft use copyright protection to be able to make this license enforceable. This certainly creates a very interesting relationship between the predominant nature of copyright, which is directed towards the protection and regulation of ownership, and a system that seems to advocate the exact opposite. The irony that such a contrary system requires copyright to survive cannot possibly be lost, and it is something that Stallman and many copyleft advocates have trouble answering, even though the web sites belonging to the Free Software advocates are filled with essays that

criticise copyright and intellectual property [28]. Regardless of these problems, the restrictions imposed by copyleft have a good number of outspoken defenders set on furthering the copyleft model regardless of any opposition [29].

3. Validity of the GPL license

The viral nature of copyleft licenses has generated a considerable amount of interest in circles that transcend software development. The idea of sharing materials is not new, and has been made more evident by the chaotic and sometimes anarchic nature of the internet [30]. However, shared materials tend to suffer from the possibility of third parties that use the freely acquired information to turn them into proprietary works. That is why many different organisations are turning to the copyleft model to protect works that are being freely shared online. One such project is the OpenContent License (OPL), a collaborative effort that sets a similar license to the GPL, ensuring that shared works will continue to remain free to subsequent users [31]. In the area of biotechnology, there have been some suggestions that the copyleft model could be used to protect the public results of the human genome race being placed in the public domain by several researchers, something that has been suggested by a leading member of the Human Genome Consortium, although the idea has never been implemented [32].

But the problem is that the actual validity of the licenses, and in particular of the copyleft clauses, has never been tested during its relatively short history. There have been no court cases against non-compliance with a copyleft license, and the few incidents that have arisen have been dealt swiftly with cease-and-desist letters to those parties suspected of producing proprietary software [33]. Despite this apparent success, there appears to be enough ground to at least consider this issue from a contract law perspective in at least two different fronts: unfair contractual terms and the passing of obligations and rights to third parties. The copyright aspect of the protection of GPL works will be analysed as well.

3.1 Unfair contractual term

The first concern for the consideration of the validity of the copyleft clauses must be to ask if they must be read as being unfair. Most jurisdictions have different public policy restrictions to contractual terms, the most common being restrictions against terms that will give away basic human rights [34], but beyond these basically recognised principles, the range of restricted or excluded terms varies from one jurisdiction to another [35]. It is the wide variation in this area of contract law that the European Union felt the need to harmonise the different approaches to unfair terms across the EU. Consumers in member states are now subject to a wide-ranging regime designed to protect consumers from unfair terms in a variety of circumstances in which they are presented with pre-formulated standard contracts, thanks to the Unfair Terms in Consumer Contracts Directive (the Directive) [36], which specifies what an unfair contractual term is, and sets a number of considerations by which clauses will be analysed to test for unfairness. The directive also provides a non-exhaustive list of some terms that will be considered unfair, none of which applies directly to copyleft licenses.

The GPL contains several different clauses that may be considered in light of the existing unfair terms legislation.. The first question will be regarding whether the licensee of some GPL software should be considered a consumer as understood by the definition provided by Art. 2(b) of the Directive, which states that a consumer will be any natural person who "*is acting for purposes which are outside his trade, business or profession*". This is a very broad definition of consumer, and even though the wording of the Directive would seem to exclude legal persons, it should be underlined that courts have generally taken a very broad interpretation as to what a consumer is, even to include companies [37]. The common interpretation of this requirement will be that the person entering into a standard contract, such as a software license, will be considered to be a consumer if they are not signing the contract as the regular course of dealing in that business. It would be fair to assume that if a software firm develops a software programme and licenses it to another software firm using the GPL, the licensee firm will probably not be considered a consumer for the purposes of the Directive. On the other hand, an individual consumer who has acquired some copyleft licensed software could possibly make a strong case arguing that he is signing the license as a consumer. This is of course a general interpretation, and the circumstances of each contract must be individually determined on a case-by-case basis.

Assuming that the license is considered to be a consumer contract as described, there is still a need to determine whether the term itself is unfair. Art. 3(1) of the Directive specifies that:

"A contractual term which has not been individually negotiated shall be regarded as unfair if, contrary to the requirement of good faith, it causes a significant imbalance in the parties' rights and obligations arising under the contract, to the detriment of the consumer."

A term will be considered to not have been negotiated individually if it has been drafted in advance and the consumer did not have a say in the terms of the final contract [38]. This definition is at the heart of any contractual dispute that may arise by the application of the Directive, and its interpretation is the one that offers more problems as it can be considered as using a very open-ended requirement, such as the often nebulous expression 'good faith'. In the UK, the test for unfairness as expressed by the Directive has been established by *Director General of Fair Trading v.*

First National Bank plc [39]. According to this ruling, the consumer must prove that there has been bad faith on the part of the undertaking in the drafting of the contract, that there is a significant imbalance to the obligations and powers of the parties, and that such imbalance must be detrimental to the consumer. The court in this ruling specified that good faith would be present if the contract was signed with fair and open dealing. Openness means that the term must be clear, legible and not contain hidden pitfalls; and fair dealing would have to be understood that the supplier should not take advantage of the other party's relatively weak position. It is important to note as well that McKendrick suggests that the concept of "good faith" should be understood in accordance to Civil Law principles [40], and as such many different aspects must be taken into consideration, for example the gravity of the imbalance, the social position of the parties and the way in which the term in question came into existence [41].

Analysing the copyleft clause with the requirements presented by this ruling, one could say that there appears to be an imbalance in the obligations of the parties as the licensee will have to use the GPL license and cannot profit from derivative works. This imbalance could also be assumed to be detrimental to the consumer as it is imposing the responsibility of not being able to use the work in whatever way it is desired. However, one must say that this is precisely the same type of imbalance that exists in copyright-based software licenses.

The main question will be in trying to determine if there has been good faith by the drafter of the license. This is more difficult to ascertain given the test of good faith presented above. In the case of GPL licenses, the test does not appear to be met by copyleft licenses. The copyleft clause is clear enough, it does not contain hidden pitfalls and the software owner is not taking advantage of the relatively weak position either, as the consumer is always free not to use the software if he so desires, and is even free to look for similar software that does not use copyleft licenses.

Based on this brief analysis of the copyleft contract term and the existing European unfair contract legislation, it would seem that the GPL copyleft clause is valid, as there are too many uncertainties as to whether or not a court would interpret this clause in favour of a licensee on the basis of the existence of good faith. It must also be assumed that the copyleft clause will be valid as it does not fall into any of the specified unfair terms provided in the Annex to the Directive. However, the question must remain open until the first case testing the validity of this type of license comes up. Given the amount of money involved in software development, it is likely that at some point copyleft will indeed receive some judicial review.

3.2. Passing obligations and rights to third parties

Another interesting legal issue that arises when considering the validity of GPL clauses is the problem of passing obligations to third parties. The legality of this practice is usually covered under the English contract law concept of the privity of contracts, of which there are two rules, one for passing obligations and one for passing benefits.

The first rule exists under traditional privity doctrine, where "*a third party cannot be subjected to a burden by a contract to which he is not a party.*" [42] This general principle is still in effect in most jurisdictions and responds to the reasonable assumption of security by not allowing parties to place contractual obligations that they are not aware of. Wherever this practice is permitted, it is usually well regulated [43]. The question must be asked of whether the GPL constitutes the imposition of an obligation to third parties. The initial response would be negative, as the imposition of the clause is done on a one-to-one basis. If one does not agree with the copyleft clause, then it is only logical that one should not use the software; and certainly one should not use it to create a derivative product.

This argument is deceivingly straightforward, but there are still other considerations to be made. It is common in competition law to have rules against imposing obligations through a distribution chain which may impose anti-competitive restrictions on the recipient; this is evident in the strict regulation and implementation of competition law in the area of licensing and vertical agreements [44]. Even though copyleft licenses do not impose obligations to third parties as the license is passed to a single licensee at the time, it is less clear whether such restrictions could be considered anti-competitive in accordance to European competition rules, as it could be found that the imposition of the copyleft clause, even if done on a one-to-one basis, could be found to be anti-competitive.

If the passing of obligations is generally not accepted in contract law, what happens to the passing of benefits? There used to be a second controversial privity rule in English law which did not allow a third party to benefit from the contract, but this has been recently modified in England [45]. It is important to point out that this second privity rule did not exist in Civil law jurisdictions [46], where third-party rights (known in Scotland as *jus quaesitum tertio*), has been an integral part of contract law [47].

The relevance of third-party rights to copyleft results in the question of whether the originator of a programme licensed under the GPL may sue a licensee who is located further down the software distribution chain for breach of contract. Assuming that A is the software creator and B is the copyleft licensee, and B licenses the software to C using the GPL, could A sue C for contractual breach if C does not comply with the copyleft clause? Contractually speaking, one would have to assume that for A to successfully sue C; A must have a third-party right arising from the contract

between B and C, which appears to be a completely invalid proposition.

The possible applicability of third-party rights to copyleft can be better understood in the famous Scottish case of *Beta Computers v Adobe Systems* [48]. In this case, Beta Computers provided a copy of software authored by a third party called Informix, for which they had a license. The court in this case found that Informix – although not part of the contract between Adobe and Beta – had a third party right. This position has been adequately criticised by MacQueen, who says that when the subject of a software transaction is a licensing agreement, third-party rights cannot possibly apply as a license grants rights by the third party, it does not create rights to the third party, which is the doctrinal requirement of third-party rights [49]. There cannot be much doubt that in the case of copyleft licenses, the author's rights arise from the license itself and the contractual provisions contained within. It will be seen next whether the author could sue under copyright providing the code has been copied without a license, but it would be more difficult to state that the author could sue for a broken contractual term contained in the license. The most secure way to maintain the validity of the copyleft clause would then be to maintain the contractual chain on a one-to-one basis, eliminating the possibility of involving third parties, even if the third party is the author.

3.3. Copyright Infringement

The analysis above would seem to indicate that the author or owner of a work that has been licensed using copyleft will find it difficult to sue subsequent users of the software down a distribution chain for contract breach. Yet, the question still remains on whether the author can sue for copyright infringement. The answer to this is much more straightforward than the contractual analysis.

Using the same example cited, let's assume that A is the software owner and B is the copyleft licensee, and that B licenses the software to C using the GPL. C modifies the software and releases a proprietary version of it by closing the source code to subsequent users. Could A sue C for copyright infringement? The answer would seem to be positive, as copyright is less preoccupied with who licensed the software to C, but the emphasis would be whether or not C is committing actions that would be considered as infringing A's copyright. The question then would become one of infringement and originality, possibly hindering on the question of whether or not C has done enough work to the original source code to be considered an original work.

This is a much better explored area of copyright law. Computer software is awarded copyright protection if it is considered an original work. The question of originality has been long discussed by the courts, but there is agreement that an original work is one that demonstrates the use of skill and labour by the author, in short, that "*that it should originate from the author*"[50]. Even though the originality requirement states that the work should not be copied in its entirety, courts have recognised that certain amount of copying is acceptable. For example, copying of the drawing of existing designs has been deemed to be original in some instances [51]. When copying exists, the copying must fulfil the long standing qualitative test to determine whether copying has been substantial [52].

In computer software, the courts have been following the general qualitative test in cases of copying from another work. In both *Richardson Computers v Flanders* [53] and *Ibcos v Barclays* [54] the courts found that if there had been any copying from a protected original work, that there had to be an analysis of whether such copying had been substantial. It is important to stress that the test is for qualitative copying, not quantitative. There will be some consideration about the quantity of the work copied [55], but even if this is minimal it may result that the copying may be deemed to be substantial. This is evident in the case of *Cantor v Tradition* [56], where copying of original source code took place from former employees of a financial services company. In this case expert witnesses found that only 2% of the original source code had been copied, accounting for only 2,952 lines of code out 77,000 [57]. The lines of code were deemed to be of importance for some modules in the resulting software, but the copying was not considered substantial to grant the infringement case. Nevertheless, the fact that some of the copying was even considered in the ruling must send signals to potential copiers of non-proprietary software about their chances in court.

Given the state of the rulings in software copyright infringement, it appears that if a copyright author or owner can prove to a court that a proprietary copy of their original software has been infringed, then it will not matter just how they obtained the software, and it will certainly not matter if they are further down in a chain of distribution. If a programmer uses substantial sections of code belonging to a copyleft program, that programmer will still be subject to legal action by the author. There may also be a question about moral rights, but these considerations fall outside of the scope of the present article [58].

5. Conclusion

An initial look at the problem of the validity of copyleft licenses seems to provide a positive response to this novel and ingenious software distribution model. There are some unanswered questions, in particular with regards to privity of contracts, but as long as the contractual chain is kept at the most simple relationship between licensor and licensee, the validity of the copyleft clause appears to be sound. Software authors interested in making sure that their works are distributed to the largest number of people without fear of commercial interests placing a fence over their works

should definitely consider the GPL model as a successful example, but some reservations may still be healthy until the first copyleft licenses are tested in court.

6. References

[1] T. Stanco, *We are the New Guardians of the World*. 16 May, 2001. @: http://lwn.net/daily/guardians.php3

[2] There are several works that achieve this, see: J. Naughton, *A Brief History of the Future,* London: Weidenfeld & Nicholson, 1999, pp.172-174; H.E. Pearson, "Open Source: The Death of Proprietary Systems?" *Computer Law & Security Report*, 16(3), 2000, pp.151-156; and R. Stallman, *The GNU Project*. 1998, last updated 24/10/2001. @: http://www.gnu.org/gnu/thegnuproject.html

[3] Stallman, *The GNU Project*. Op cit.

[4] Or as it is often stated in OS and FS circles, free must be understood as in freedom, not as in beer.

[5] R. Stallman, *The Free Software Definition*, 1996, last updated 10/17/2001. @: http://www.fsf.org/philosophy/free-sw.html

[6] R. Stallman, *Selling Free Software*, 1996, last updated 08/08/2001. @: http://www.fsf.org/philosophy/selling.html

[7] Stallman, *The Free Software Definition*, Op cit.

[8] Open Source Initiative. *History of the OSI*. 2001. @: http://www.opensource.org/docs/history.html

[9] It may even be said that Microsoft's competitive tactics against Netscape were excessive and even predatory, and they prompted the anti-trust case brought by the US Department of Justice against Microsoft. A roadmap to the case can be found here: http://www.stern.nyu.edu/networks/ms/top.html

[10] Open Source Initiative. *History of the OSI*. Op cit.

[11] E. Raymond, *Keeping an open mind*, March 1999. @: http://tuxedo.org/~esr/writings/openmind.html

[12] The guidelines can be found here: http://www.debian.org/social_contract.html#guidelines

[13] The OSD can be found here: http://www.opensource.org/docs/definition.php

[14] For examples of these, see: Free Software Foundation. *Various Licenses and Comments about Them*, 1999, last updated 2003/06/15. @: http://www.fsf.org/licenses/license-list.html

[15] R. Stallman, *Why "Free Software"' is better than `"Open Source",* 1998, last updated 20/08/2001. @: http://www.fsf.org/philosophy/free-software-for-freedom.html

[16] Open Source Initiative. *OSI Certification Mark and Program,* April 30, 2001. @: http://www.opensource.org/docs/certification_mark.html

[17] At the moment there are a total of 43 OSI certified licenses.

[18] Working group on Libre Software. *Free Software / Open Source: Information Society Opportunities for Europe?* April 2000. @: http://eu.conecta.it/paper/paper.html

[19] O'Sullivan M. "Making Copyright Ambidextrous: An Expose of Copyleft", *The Journal of Information, Law and Technology (JILT)* 2002 (3). @ http://elj.warwick.ac.uk/jilt/02-3/osullivan.html

[20] R. Stallman, *What is copyleft?* 1996, last modified 05/11/2001. @: http://www.fsf.org/copyleft/copyleft.html

[21] Ibid.

[22] R. Stallman. *Copyleft: Pragmatic Idealism*. 1998, last updated August 26, 2002. @: http://www.fsf.org/philosophy/pragmatic.html

[23] Lambert, P. "Copyleft and Copyright: the Legal Issues", *Seminar on Copyleft and Opens Source Software: History, Applications and Legal Issues*. Tuesday 6th February, 2001. University College Dublin, pp.10-12

[24] Free Software Foundation. *GNU General Public License*. Last modified July 15, 2001. @: http://www.fsf.org/licenses/gpl.html

[25] Ibid.

[26] Ibid.

[27] M. J. Radin, "Humans, Computers, and Binding Commitment", 75 *Ind. L.J.* 1125, Fall 2000.

[28] For example, see: Free Software Foundation, *Reevaluating Copyright: The Public Must Prevail*, 1996, last updated January 8, 2001. @: http://www.fsf.org/philosophy/reevaluating-copyright.html

[29] For one such defender, see: E. Moglen, "Anarchism Triumphant", *First Monday*, Vol. 4 No. 8 - August 2, 1999. @: http://www.firstmonday.org/issues/issue4_8/moglen/index.html

[30] For more on this subject, see: A. Guadamuz, "The New Sharing ethic in Cyberspace", *Journal of World Intellectual Property*, Vol. 5 No. 1, January 2002, pp.129-139.

[31] The license can be found here http://www.opencontent.org/opl.shtml. Other interesting copyleft licenses include the Design Science License, the Open Audio License and even Open Cola,

the world's first copyleft fizzy drink. See: G. Lawton, "The Great Giveaway", *New Scientist*. @: http://www.newscientist.com/hottopics/copyleft/copyleftart.jsp

[32] J. Sulston, "Intellectual Property and the Human Genome", *Global intellectual Property Rights*, Drahos and Mayne eds. London: Palgrave, 2002, p.561-73.

[33] G. Moody, *Rebel Code*. London: Penguin Books, 2001, p.313.

[34] Radin, op cit.

[35] In the UK for example, the Unfair Contract Term Act 1977 (UCTA) contains an exhaustive list of unfair terms, which include exclusion, limitation and indemnity clauses.

[36] Council Directive of 5 April 1993 93/13/EEC on unfair terms in consumer contracts, O.J. No. L95/29, 21.4.1993.

[37] Most recently in the UK one can find examples of this in *SAM Business Systems Limited v Hedley & Co.* [2002] EWHC 2733.There are several older examples of this, such as *R&B Customs Brokers Ltd v United Dominions Trust Ltd* [1988] 1 WLR 321; Cass. Civ. 1re, 28 April 1987. Most notably for software purposes are *St Albans City & District Council v International Computers Ltd* [1996] 4 All ER 481.

[38] Directive 93/13/EEC, Art. 3(2).

[39] *Director General of Fair Trading v First National Bank Plc,* [2001] UKHL 52; [2002] 1 A.C. 481.

[40] E. McKendrick, *Contract Law*, Fourth Edition, Basingstoke: Palgrave, 2000, p.369.

[41] Some of these principles can be seen in several continental cases, such as *Saladin/HBU*, Hoge Raad, NJ 1967.261 (G.J. Scholten). For a more complete work on the subject of good faith in Civil Law, see: R. Zimmermann and S. Whittaker (eds), *Good faith in European contract law*, Cambridge: Cambridge University Press, 2000.

[42] McKendrick, Op cit; p.133.

[43] Radin notes for example that some cases that occur in competition law, or public policy issues, see: Radin, Op cit, p.135.

[44] For example, there are restrictions in the area of competition is in the area of technology transfer licensing, where impositions of this nature are blacklisted. See: Commission Regulation 240/96/EC on the application of Article 85(3) of the EC Treaty to certain categories of technology transfer agreements, OJ 1996. L 31/2. For more on vertical restraints, see: M. Furse, *Competition Law of the UK & EC*, London: Blackstone Press, 1999, pp.104-112.

[45] This was done by The Contracts (Rights of Third Parties) Act 1999.

[46] And in mixed legal systems such as Scotland.

[47] In France for example, privity of contract is qualified by Art. 1121 of the *Code Civil*, which allows third party rights. In Germany, Art. 328 of the *Burgerliches Gesetzbuch* allows for the performance of rights by third parties. Another example can be found in Art. 2.115 of the Principles of European Contract Law, see: European Commission on Contract Law, *Principles of European Contract Law: Part 1: Performance, Non-performance, Remedies*, ed. O. Lando and H. Beale, 1995.

[48] 1996 SCLR 587.

[49] For an excellent attack to this ruling, see: H.L. MacQueen, "Software Transactions and Contract Law" *Law and the Internet: Regulating Cyberspace*, Edwards and Waelde (eds), Oxford: Hart Publishing, 1997.

[50] *University of London Press Ltd. v. University Tutorial Press Ltd.* [1916] 2 Ch. 601.

[51] For examples of this see: The Duriron Company Inc v Hugh Jennings & Co Ltd [1984] FSR 1; and *Interlego v Tyco Industries* [1989] AC 217; [1988] 3 All ER 949.

[52] Existing in common law since *Bleistein v Donaldson Lithography Co*, 188 US 239, 250 (1903).

[53] *John Richardson Computers Ltd v Flanders and Chemtec Ltd* [1993] FSR 497.

[54] *Ibcos ComputersLtd v Barclays Mercantile Highland Finance* [1994] FSR 275.

[55] For which software may result helpful in analysing the number of lines of code copied. Software such as MOSS: http://www.cs.berkeley.edu/%7Eaiken/moss.html

[56] *Cantor Fitzgerald International v Tradition (UK) Ltd* [1999] Masons CLR 157.

[57] Lloyd, *Information Technology Law*, 3rd Edition, London: Butterworths, p.411.

[58] For a good look at moral rights and OSS, see: A. Metzger ad T. Jaeger, "Open Source Software and German Copyright Law", IIC Vol. 32, 2001, pp.52-74.

Open Source Managagement

Norm-constructionism, structural order and DRM features in Open Source

Working Paper by **Kristoffer Schollin**,

CIP - Center for Intellectual Property studies, Gothenburg Department of Law

kristoffer.schollin@law.gu.se

Abstract

Starting from the political schism between Open Source and Free Software, this paper takes a look at the recent debate about DRM features in the Linux operating system. The Development-process of complex, enabling software, so-called Platform Technologies, is then coupled with the lore of Social Constructionism, particularly as it has been used in the field of Law. After an overview of the copyright regulation in Europe, three kind of regulatory modes are outlined: Implicit regulation, Institutional regulation and Virtual Regulation. The Constructionist view is then adapted to the software development field and the debate between Open Source pragmatists and Free Software activists is seen in partially new light. Finally, a suggestion is made from the lessons of Norm-Constructionism regarding the future attitude of Open Software development.

1. Introduction

"The principle of law seems to mediate between the principle of morality and that of democracy. But it is not entirely clear how the latter two principles are related."
- Jürgen Habermas

The 1960s and 1970s saw the birth of a new class of property: **software**. Computer market dominant IBM, in a preemptive action against a grand US Department of Justice antitrust suit, had decided to unbundle its software from its hardware and started charging money for its software separately. The important question whether this birth was primarily a result of IBM's decision or whether this was actually a development that would have taken off anyway, due to other factors such as the rising complexity and cost of software development, is a serious and very interesting one, albeit a question that will not be answered here. It is sometimes theorized that the 1956 US government consent decree forbidding the giant AT&T from engaging in commercial computing activities prolonged the view of software as free and that the development of software as property would actually have started sooner, had it not been for this.

Regardless of the grounds for this trend, it spawned a response, a counter-reaction among those that disliked the development for varying reasons. The situation where software was free that was taken for granted, was no longer. In effect, this trend forced the formulation of an alternative. It required an answer to a question that had hitherto been left unaddressed: "Why should software be free?". As the modern legend goes, Richard Stallman, in 1979 employed at MIT, was moved to a fundamental belief in free software by a jamming Xerox laser printer. When requesting the source code for the printer drivers from Xerox because he wished to rectify the problem he found that his request was denied. Seeing this as the start of a "ban on a cooperating community" Stallman in that moment also found his belief in free software. In 1982 Stallman went on to start development of the GNU project, a suite of software components aimed at providing a free Unix-like operating system. This start was followed by establishment of the Free Software Foundation, in 1985. Such was the gravitational pull of FSF and Stallman that Eric S. Raymond describes him in his influential paper A Brief History of Hackerdom[1] as

[1] http://www.catb.org/~esr/writings/hacker-history. Last visited 030628.

largely defining the public ideology of hacker culture for more than a decade. Thus the Free Software movement was (re?)born.

The power that words can hold is sometimes simply amazing. In the late nineteen-nineties, the use of the phrase "Free Software" was more and more being replaced with the use of "Open Source software". The need for an alternative to the Free Software concept had supposedly been born out of many reasons, the need for contractual flexibility being one. It was felt in some circles that the structure of the GNU general public license, the GPL, was too stifling and rigid. Another reason was that the word "Free" was to often misunderstood as non-commercial, or even anti-commercial, something that would doom the Free software movement to be ignored by the "people in suits" (Bruce Perens), no matter how "technologically excellent stuff" they managed to develop. Many felt that not only was there a need for a new word to describe the movement, there was a need for a entirely new language as well, one that wasn't so tied in with Stallman's openly political approach, one that would not engender suspicion among the corporate headquarters. So a faction of Free Software people donned the tux, and became Open Source. The strategy was a market success, as Open Source would eventually outshine its older sibling by many magnitudes, at least in matters of fame and citations.

Are the aims of Open Source and Free Software different? Or are they just different brands for basically the same kind of movement? This paper starts by presupposing that they are at least useful in designating the political camp from the apolitical one within the Free software/Open Source movement. However, as the story of whether to include DRM and TCPA capabilities in Linux unfolds we will learn that the goals and motivations are so diverse that no simple explanatory model will suffice.

To start with, we find that there is at least a large part of the Open Source movement that are explicitly apolitical, feeling that politics and ideology are, mostly for reasons of legitimacy, best left alone. In this sense there is a true difference vis-à-vis the openly ideology-driven Free Software movement and those that dub themselves Pragmatics. To these pragmatists, instead of ideology, the raison d'être of the Open Source movement is explained as many varying factors. The most common ones being different variations on the theme that the Open Source way of organizing software development yields better, more reliable, more effective software. Another angle is that Open Source development needs no "raison", that there is no need to discuss the purpose of it and that a theory of such a purpose is undesirable, because the individual developers are doing it for their own impenetrable pleasure, "just for fun". In fact Linus Torvalds openly takes the apolitical approach, in the recent debate on Digital rights management software and Linux, he bluntly states that Linux is an Operating System, not a political movement.[2]

This paper treats treat software development in general terms, and finds that, at least on the OS level, it is a process of regulatory development. Viewed as such it shares its basic premises with such fields as that of Law where one is constantly faced questions of what goals should be achieved and what means are suitable to achieve those goals. There are many factors that decide to what extent this analogy is viable. The complexity level plays a deciding role here, where the capabilities and purpose of the software are highly decisive when it comes to extending this analogy. This paper reasons around particular features that covers Operating Systems where we find market-relevant attributes (such as market-entry barriers, lock-in and lock-out effects) as well as technology enabling attributes that serves as platforms for other kinds of software, with constrains and affordances built into the infrastructure. These concepts are further analyzed and analyzed from a legal scholarly perspective.

To claim that laws and legal phenomena are social constructs is not particularly controversial in this day and age. There are precious few that hold the belief that regulations or the Law share ontological status with rocks and gravity. Axel Hägerström cleared the way for the school of thought known as **Social Contructionism** back in 1939 when he established that the legal system is nothing more than "a social machinery where the cogs consist of men and women."

This paper deals with three kinds of facts about the world. Brute facts, Social facts and Virtual facts. These facts underpin three corresponding modes of regulation called Implicit regulation, Institutional regulation and Virtual regulation. While brute facts rest on the facticity of the natural world, social facts and virtual facts become objectified through a process of **reification**, achieving an ontological status

---

[2] http://marc.theaimsgroup.com/?l=linux-kernel&m=105119647419011&w=2. Last visited 030627. This kind of statement is of course in keeping with Mr. Torvalds' usual attitude taken in these kinds of dealings. The "pragmatist" approach, as I will call it here, is also the one he took in the KDE versus GNOME desktop environments. Because of his huge influence, it is of course tempting to speculate on the motivations behind Mr. Torvalds' consistent apolitical, pragmatist attitude in his capacity as Linux general. In my opinion however, this would be a futile example of near-Kremlinology, as well as demeaning to Mr Torvalds himself.

independent of immediate human activity, a certain permanence, through a collective acceptance of their reality. Though they rely on some very different factors to achieve this permanence, they nonetheless share some important ones.

Complex software creations are virtual constructions. Following the insights of social constructionism, there is a need for being on our guard and to question what is positive (describing) and what is normative. Bruno Latour and others point out that it is by definition impossible to describe objects such as firms, property rights, and software movements without participating in a normative process where our impact is in direct opposite relation to the strength of permanence of the social/virtual object at hand and in direct relation to our involvement and status within the structure.

The current debate on DRM systems within the Linux movement was sparked by Linus Torvalds on the Linux Kernel list, but the debate was rather lukewarm and matter-of-fact. This paper seeks to clarify how different kinds of DRM-capabilities means very different things for any complex techno-regulatory system such as an operating system and what the normative ingredients consist of and how the responsibility of the developers programmers and managers within a software development network plays out. At the core of the issue is the fact that DRM capacities rely on **Platform Control** and include ways of locating power about what kind of software should be able to run on a hardware system to a certification body of some kind. If we ignore any factual use, or abuse of such capabilities and just analyze the power-exchange in itself, we have to answer the question of whether this kind of decision making power is political or not, and thus whether it is possible to take an apolitical stance with regards to this development. The different motives and goals at internal management level as well as the market level and at the social level are examined from a **norm-constructionist** point of view and an attempt at shedding new light on the schism between Free software and Open source is made through an analysis wherein the author attempts to see the problem as a bigger quest for structural order in society.

Incidentally, the name "Linux" will be used instead of the GNU/Linux. This is not done in order to take sides in any kind of debate on the origins of an essential development in the software industry. I am a firm believer in using names as what Saul Kripke calls "Rigid Designators"[3] and not as descriptors. The history of the GNU project and the development of the Linux kernel remains the same, no matter how convoluted a name you use. In short, "Linux" as a name, flies better than Gnu/Linux, it is simply a better name, albeit a less-accurate descriptor.

2. Institutional and Implicit Regulation in DRM

Turning first to the issue of "content", the information that is sought to be protected through the use of Digital Rights Management. Broadly speaking, there are two kinds of regulation that have been viewed to apply to information-content: institutional and implicit regulation.[4]

The **implicit regulation** is what we start with. It is given and defined by the medium on which the information exists or the environment in which the users exist. It sets certain affordances and constraints on the use of the information. For example, a story in a book is easy to read and carry around. It is also easy to give the book to someone else so that person can also read the story. These are affordances of the Implicit regulation provided by book technology, in the present state of our society. A book does not, however, make it particularly easy to read the story simultaneously with another person, or to make a copy of the story so that you can both read the story while sitting opposite each other. These latter are constraints of the book medium. Together with laws of nature, implicit regulation is an amalgamate of technology level, economy and overall technology access in our society, what can be called "the socio-technical infrastructure". Though the legislator not only tries to monitor but also to actually steer the development of this infrastructure, it is one of the premises of this paper's definition that this factor is to a large extent out of the hands of the direct human policy-setting attempts.

What is in this paper termed "**Institutional regulation**" tries to improve upon the Implicit regulation by its own set of constraints and affordances. By way of **laws**, it stipulates for example that certain monopoly rights should be given automatically to authors upon the creation of a piece of work, or to inventors upon filing for a patent application. It also stipulates limitations in these monopoly rights for certain uses for the benefit of society in order to create a better world than the alternative, where implicit regulation reigns supreme.

[3] Kripke, Saul: "Naming and Necessity": Blackwell Publishers, Oxford: 1993.

[4] Rosenblatt, Trippe, Mooney, pvii: "Digital Rights Management – Business and Technology". M&T Books, NY, 2002.

There is a sub-class of institutional regulation that is dubbed "transaction regulation". These are institutionally-based constraints and affordances that arise when a transaction has taken place, such as the right to give your copy of a book to a friend. But just the same, these are institutionally based phenomena, and thus not necessary to treat any differently than other regulation of the same class.

Institutional regulation comes "on top" of implicit regulation. This means that it is forced to adapt to changes in socio-technical infrastructure, in order to be useful in the way that it was originally imagined.

3. Copyright Regulation in Europe

Copyright is the institution that stipulates that authors, composers, and artistic creative individuals in general have certain rights in the works that their creative endeavor gives rise to. In Sweden, these rights are chiefly to be found in the Copyright Act (1960:729). The kind of works regulated therein can come in the form of, for example, novels, magazine articles, sonnets, musical lyrics, computer programs, music, theatrical works, movies and what-have-you. As you can see, the Copyright Act has a fairly wide scope of human activity to regulate.

These rights are given to the creator upon creation of the work and, for the time being, they remain until 70 years have passed, from the death of the creator. Thus these rights not only guarantee a level of control to the creator, but sometimes a handsome revenue to his or her scions,

The copyrights are two-fold: they are divided between the **economic copyrights** and the **moral copyrights**.

The economic rights consist of the rights to make copies of the work as well as the rights to make the work available to the public. An artistic work is made available to the public when it is performed, exhibited or when copies of it are released to the public.

The moral rights give the creator two major rights. First, the right to be named as creator whenever the work is used. The details of this provision differ between artistic fields according to accepted practices. Secondly, the creator has the right to oppose the use of the work in such a way or in such an environment that it would violate the "artistic integrity" of the work or in other ways sully the artistic integrity of the creator.

Given the luxury of historical perspective, it seems clear that the reasons for the creation of the institutional rights were manifold. But there is a general consensus that their continued existence hinges on them as a means to an end. The end being in this case to raise the incentive for the creation of new artistic works, to the benefit of all society. This is exemplified in the preamble of the new European Copyright Directive (2001/29/EC):

> *(4) A harmonised legal framework on copyright and related rights, through increased legal certainty and while providing for a high level of protection of intellectual property, will foster substantial investment in creativity and innovation, including network infrastructure, and lead in turn to growth and increased competitiveness of European industry, both in the area of content provision and information technology and more generally across a wide range of industrial and cultural sectors. This will safeguard employment and encourage new job creation.*

The economic copyrights were created with a lot of **built-in exemptions**, to prevent them from being counter-effective in their goal of furthering the level of creativity. The awarding of positions of exclusivity within a field always carries the potential of destructive consequences that prevent the diffusion of culture and even erects hindrances to the creation of new works. Examples of exemptions are the right to make copies for private/non-commercial use (corresponding in part to the fair-use doctrine of the US), the right for libraries to use the work extensively, the right to quote the work, the right to make satire of the work, and so on …

The Swedish Copyright Act also includes provisions to the effect that so-called "neighboring rights" are also protected. These rights mean that performers such as singers, musicians and actors, are given a right to their performance of an artistic work. The producers of records and movies, together with radio- and television broadcast companies are given neighboring rights with regard to their records, movies and programs just as photographers and database creators are. These rights typically have a longevity of 50 years from the time when the work was first created.

The neighboring rights come with basically the same kind of limitations that was mentioned before.

Swedish copyright regulation is primarily relevant for Swedish creative works, just as Finnish copyright pertains to Finnish creations. Copyright is, like most legislation, a national concern. However, through ratification of several different international conventions in the area of copyright, these rights are also very much valid in other countries. The ratifying countries have, among other provisions, accepted to protect foreign copyrights with equal fervor as their own national copyrights. Particularly within the European Union there is now an ever-closer view regarding the copyright regulation.

Regarding regular copyrights there is the Berne Convention for the Protection of Literary and Artistic Works wherein the contracting states have accepted to

protect foreign copyrights at an equal level as domestic copyrights. So far, June 2003, 150 states have ratified the Berne Convention.

The neighboring rights are given protection under the Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations (the "Rome Convention"), the Convention for the Protection of Producers of Phonograms Against Unauthorized Duplication of Their Phonograms (the "Phonogram Convention) as well as a third convention; the European Agreement on the Protection of Television Broadcasts.[5]

Outside of the above-quoted Directive 2001/29, there have been five earlier European Directives within the copyright are that have served to further the common European ground in copyright. They are (1) the Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programmes, (2) the Council Directive 92/100/EEC of 19 November 1992 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, (3) Directive 93/83/EEC of 27 September 1993 on the coordination of certain rules concerning copyright and rights related to copyright applicable to satellite broadcasting and cable retransmission, (4) Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights, and (5) Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.[6]

The consequence of violating copyright regulation most commonly known to the public is that of liability for damages. However, in some cases, fines and even jail for up to two years can follow, according to the Swedish Copyright Act.[7] These latter provisions are only applicable as long as the violation is outside of the borders of private/non-commercial use.[8]

4. Brute facts, institutional facts, virtual facts

So then, what does it mean when I say such things as "This is my copyright, you can't copy my story"? Is it the same kind of statement as "This painting here is impossible to copy, because the pigments are so old and complex"? Do those two statements have comparable ontological status?

Of course they don't. The latter statement, about the painting, is a statement about the implicit regulation of this world whereas the first statement is not. Statements about implicit regulation are statements where the truth or falsity is based on what John Searle[9] calls **brute facts**.

Brute facts and implicit regulation are ontologically different from institutional facts and institutional regulation because they don't depend on anyone but themselves for their existence, whereas institutional fact do not possess an existence of their own, being instead dependent on common human capacities of holding beliefs. The institutions, being ultimately social constructions of humans, act as a "platform" that makes the existence of institutional regulation possible. Legal concepts belong solidly in the camp of institutional phenomena.

The attitude that results from the division of the implicit and the institutional is called Social Constructionism and has given risen to a wide variety of sociological and philosophical theories. Influential contributors to the groundswell of literature are, to name a few, Ludwig Wittgenstein, Karl Marx, Friedrich Nietsczsche, Michel Foucault, Jürgen Habermas, Laques Derrida, Peter Berger, Bruno Latour and Niklas Luhmann. To say that legal phenomena are social constructions are, to theses theorists, like calling the kettle black.

Just as the implicit regulation makes some things possible and some things impossible, what I called "affordances and constraints" earlier, so does also institutional regulation mirror this effect. Some such regulation merely regulate already existing activities, such as the rules that regulate the driving of cars, even though driving is an activity that existed well before there were institutional regulation for it. But other regulation actually enables activities that were not possible before. Copyright regulation enables the trading and selling of licenses for example. Before copyright grew to maturity, asking person X for permission to tell a story that has elements in common with a story he came up with just wasn't done, just as pushing around wooden statues on checkered boards wasn't done until the rules for chess were created. John Searle calls the regulation that creates new affordances "constitutional rules" and separate them from "regulative rules", rules that merely sets boundaries for already existing activities. Some readers will have noticed by now that the implicit regulation, by its very

[5] These texts are readily available at http://www.wipo.int/treaties/ip/. Last visited 060329.

[6] http://www.europa.eu.int/scadplus/leg/en/s06020.htm. Last visited 060316.

[7] 53 §, 57 §

[8] 53 §, second paragraph

[9] Searle, John R, "The Construction of Social Reality". Penguin books, 1996.

definititon, only holds constitutive elements, never regulative.

Very often it transpires that an institutional phenomenon becomes so well defined, so irreplaceable and taken for granted that people start to treat it as if it had an existence of its own, an intrinsic existence, like brute facts do. This process is called **reification**, after the latin "re" for thing, so it really means "objectify".

Just as implicit regulation acts as a platform that enable and disables activities and the constitutional element of institutional regulation perform the same function for other activities, so does the technological infrastructure that is a platform technology enable and disable activities.

Platform technologies is a concept that in the context of this paper is used to denote technologies that act as the constitutive element of other activities and that incorporates its own set of constraints and affordances to regulate behavior in that activity. As such, the DVD-video system is a platform, just the same as MS Windows is a platform, and Linux is a platform.

Platform technologies enable a third mode of regulation: **virtual regulation**. A kind of hybrid between implicit regulation and institutional regulation, virtual regulation are dependent on human creations for its existence and are not outside of policy control. But these phenomena do not altogether lack an existence of their own, being anchored in the technology that is able to sustain itself, at least for a limited period of time. Virtual regulation is the embodiment of institutional rules and makes possible statements such as: "*You cannot fast forward through these commercials that precede the film."* The example is taken from the DVD platform. Observe that the "cannot" is used both in the normative and the positive (descriptive), at the same time.

As mentioned before, implicit regulation, being outside of human control, take precedence before institutional regulation. That is not the case with virtual regulation. It evolves together with institutional rules. *In many areas today, particularly that of DRM, the two forms of regulation, virtual and institutional, interplay and substitute in an ever-increasing complexity.*

Since both these forms of regulation are man-made, they are both classified as "structural phenomen", consisting of structural concepts.

5. Platform Control

Because the implicit component of virtual regulation falls under human control, it is also threatened by the breakdown of this control and thus the failure of the virtual regulation. What is required is called "Platform Control".

To have perfect platform control is to have supremacy over the virtual regulation that underpins a certain field of activity. Taking the watching of DVD movies as an example, platform control aims to prevent the possibility of uttering statements such as: "*This European DVD can be played in Japan, if only you use the player manufactured by X instead*", or a statement like: "*This movie can be played on in japan with the same features of DVD movies if you get it on the equally good and prevalent format Y instead.*" The existence of such utterances in Sweden today, are a sure sign that the DVD platform control is breaking down.

Platform control can be subdivided into **internal and external platform control**. The internal control is dependent on common vision among the controlling parties as well as the implementation of effective tools of enforcement. The external control focuses on cultivating the exclusivity position to eradicate or prevent the existence of competing and compatible platforms. To this end is often leveraged various network economy effects, such as: returns to scale through coordination on standard, incompatibility engineering, Sunk costs and application barriers to entry. These effects are described thoroughly in Tim Bresnahan's 2001 paper "The Economics of the Microsoft Case".[10]

The kind of DRM discussed in this paper is dependent on platform control for its existence. The golden graal of platform control would be a system like the proposed Next generation Secure Computer Base – NGSCB and the visions of the Trusted Computer Platform Alliance - TCPA where a level of uniformity and enforcement unrivalled by any legal system is possible.[11]

6. The Breakdown of Control of the DVD Platform

Is this kind of control needed for DRM maintainable on a platform like Linux? Of course it is, Linux, like

---

[10] www.stanford.edu/~tbres/Microsoft/The_Economics_of_The_Microsoft_Case.pdf. Last visited 030630.

[11] The paper at hand does unfortunately leave no possibility of delving deeper into the capacities of these systems. See the following sites, for a start: www.microsoft.com/ngscb, www.cl.cam.ac.uk/~rja14/tcpa-faq.html, www.trustedpc.org, www.againsttcpa.com.

many other Open Source projects, has a hierarchal command structure that aids internal platform control, and its dependence on networking interoperability makes for excellent supervisory and enforcement capacities. Some might object that because of the open code of Linux and the dispersed developer network, there will always be tools and loopholes created to make circumvention of a DRM structure a possibility. Yes, that is of course always a possibility, but as Lawrence Lessig has eloquently shown, it is a classical fallacy to confuse the non-existence of perfect control with the non-existence of effective control.[12] Locks can be picked, but that does not mean locks are useless. Murder is a crime, but that's does not mean that some murderers do not go unpunished, but neither does it mean that murder legislation is useless. In essence: the fact that a few technologically savvy individuals might be able to circumvent a system of virtual regulation has very little real effect when it comes to the question of how that system affects society as a whole, just as the fact that a small circle of rich people can mitigate or avoid the effect of institutional regulation has very little impact on how most peoples lives are affected by the legal code and enforcement system.

The now famous DVD-Jon of Norway was party to creating a DVD-protection circumventing software, DeCSS, but that feat was not responsible for the break-down of the DVD platform. Lacking any substantial research in the field, I would still venture that less than five percent of the population in Sweden has benefited from the existence of DeCSS. However, most people in Sweden can still ignore the virtual regulation of the DVD system known as the region-encoding.

The region encoding is piece of virtual regulation that was put into the DVD platform to make it possible to sector the global market in nine zones. This, in order to prevent a disc bought in, for example China, to play successfully on a DVD player bought in Sweden. This piece of virtual regulation is a serious attempt at combating the free-trade forces that have tried to remove barriers to global exchange of goods and service.

The DVD platform lost internal control partly because the system lacked an effective system of remote monitoring and enforcement but mostly it did so because the partners of the DVD platform could not maintain solidarity towards each other, sometimes not even within the own company. Some of the parties of the DVD consortium are mostly in the business of selling media content, while others are more in the business of selling the hardware that played the discs. The content-providers benefited from the region-control since they could prevent parallel imports between markets, but the hardware-manufacturers did not have this incentive. This resulted in a state where one part of the organization would sit down and create virtual regulation constraints, such as the region encoding, while another part of the system was busy building in backdoors for the easy removal of these constraints in order to gain a consumer advantage vis-à-vis their "partners". Once the ball started rolling, it became an accepted fact among Swedish consumers that you did not have to accept region-control since some brands could be easily modified at the store. Of course these brands sold better than the other brands. This kind of backstabbing within the consortium, the loss of internal platform controll, has resulted in the almost complete failure of the DVD region-constraint in Sweden.

7. Social Constructionism and Law

As we pointed out before, institutional phenomena have no intrinsic existence. The Scandinavian countries, particularly Sweden and Denmark, were highly prevalent in clearing the way for these trains of thought for the field of law. There, the idea of law as a metaphysical, independently existing, entity was harshly criticized. While Axel Hägerström is the most famous, Wilhelm Lundstedt was probably the more radical. He relentlessly criticized this idea as not only untrue, but also as dangerous to society because it provided a "veil" under which the legal professionals and legal scholars could realize their own values in a guise of objectivity and impartiality.[13] These are exactly the dangers of the process of reification described above, and not only a danger, but also a natural and unavoidable consequence of the hidden reification of institutional facts.

Scandinavian Realism viewed the legal concepts as **constructions**, artificial phenomena created in order to satisfy a multitude of factors such as commonly held notions of justice, class-related interests, fear of social unrest to just name a few. Hägerström and Lundstedt, can aptly be characterized as belonging to a social constructionist school of thought. Both argued for seeing the legal constructs behind the veil and to always take the real social ramifications into consideration.

The view held by the Scandinavian Realists is blandly non-controversial and scathingly controversial,

[12] Lessig, Lawrence, p57: "Code and other laws of cyberspace", Basic Books; NY, 1999.

[13] Petrusson, Ulf and Glavå, Mats: "Illusionen om rätten! – juristprofessionen och ansvaret för Rättskonstruktionerna." From "Erkjennelse och engasjement" – Minnesseminar for David Roland Doublet, Fagbokförlaget 2002.

both at the same time. It is widely accepted that law and other forms of non-implicit regulation are social constructions, and that these construction have come to be because of human needs and strife. On the other hand, speaking too loudly about this tends to raise the fear that the constructions will lose their legitimacy, the very thing that underpins their efficacy. This, what has been characterized as the Grand Dilemma of the social constructionist, makes the theory still threatening. There is something troubling about it that makes it difficult to handle.

John Searle describes this as the apparent self-referentiality of social concepts. And we humans tend to see self-reference as a vicious circle rather than as a benign one.

Niklas Luhmann is often viewed as one of the most influential social constructionist theoreticians and his theories go by the name of "epistemic constructionism".[14] He characterizes the field of law as a self-referencing system wherein the controlling actors decide upon policy exclusively according to self-generated knowledge and purport to possess the ability to separate between internal, relevant information as opposed to external, non-relevant information. In reality, this closed attitude applies only on the normative plane (i.e. as a "façade") since the system is actually cognitively open. By cognitively open is meant that there is a continued flow of impressions and influence from the outside that are taken into internal consideration no matter what the official party-line purports. Luhmann describes law as an "information processing system", that from the binary code communicates information legally-internally and legally-externally.[15] This is the very essence of social constructionism. That while treating a social phenomenon as a thing with an independent existence, the act of describing the phenomenon is also the act of normatively asserting what it is, thereby partaking in its construction. Luhmann shows in his theories how the lack of awareness about the social constructionist process results in a legal community that normatively claim to be doing one thing while in reality the are doing something altogether different. The legal profession is mostly unaware of the real cognitive processes and does not strive to understand the process of social constructionism.

But when it comes to the controversial question of what this theory means for the legitimacy of law, Luhmann, like so many others, backs down. To protect the legitimacy of the legal system the closed normative attitude should be maintained. Only legal theoreticians should devote their time to constructionist theories, while legal practitioners and other legal scholars should ignore these findings.[16]

The school of thought known as norm-constructionism tries to provide a workable solution to this obvious enigma: how to maintain an open attitude about the underlying institutional processes without wrecking the institutions themselves in the process?

8. Normative Closure, Cognitive Openness in Open Source

So much for the legal community and the institutional regulation. How about the Operating Systems design communities and the virtual regulation required for effective DRM?

Firstly, note that there is no implicit regulation on a digital platform, it's role is fulfilled by the constraints and affordances made possible by the virtual regulation. Ergo, the old image of institutional regulation as having to relate and adjust to implicit regulation does not apply here. Virtual regulation and institutional regulation develop side by side.

Secondly, operating systems are, like all platform technologies, pieces of virtual regulation. They bring with them a set of constraints and affordances. In the area of DRM, there is no doubt that centrally controlled operating systems with hidden code, like Mac OS or MS Windows, are more suitable to DRM than for example Linux. There is also no doubt that participating in a software project on such a scale that it attempts to rival MS Windows is at heart a social constructionist project.

To call someone "just an engineer" and meaning someone who stands apart from the political sphere does then neither have to be meant as a belittling term, nor as a source of virtue, it is simply nonsensical in the context of a project like Linux. When a company like Microsoft chooses to include or to exclude certain features from their Windows software and its integrated applications, they do this on the basis of carefully orchestrated policies that involve visions

[14] Bertilsson, Margareta: "Socialkonstruktivisme: Et erkendelsesociologisk perspektiv". From Margaretha Järvinen og Margareta Bertilsson (red): Socialkonstruktivisme. Bidrag till en kritisk diskussion. Copenhagen: Hans Reitzels Forlag, 1998.

[15] Luhmann, Niklas: "The Self-reproduction of Law and its Limits." From Gunther Teubner (red): "Dilemmas of Law in the Welfare State." Berlin: Walter de Gruyter & Co., 1985.

[16] Petrusson, Ulf and Glavå, Mats: "Illusionen om rätten! – juristprofessionen och ansvaret för Rättskonstruktionerna." From "Erkjennelse och engasjement" – Minnesseminar for David Roland Doublet, Fagbokförlaget 2002.

about what kind of market and society are desired in five-ten years from now. Linux, despite its open development structure, is fundamentally the same kind of process.

In their capacity of co-creators of virtual regulation, the individuals behind the design of that regulation gain a responsibility with respect to the societal effects of that regulation. This responsibility follows directly from the participation in a virtual regulatory process and its individual level is related to the amount of impact each participant have on the constraints and affordances of the system. Being a responsibility with respect to societal effects, this responsibility is of course a responsibility that befalls any one involved in a social constructionist activity.

In this sense, the participants in a movement like Linux face a similar dilemma to that outlined for lawyers above. Cognitively, there is a continued flow of impressions and influence from the outside world, such as political and market actors' demand for features, or lack of features coupled with interests of the public domain etc. Normatively however, the pragmatist approach is taken, denying the interaction with society and the built-in social constructionist agenda of the Open Source movement.

9. Open Source and Legitimacy

On the Freeware Summit in Palo Alto on 7 April 1998 the term Open Source replaced some of what had previously been known as Free Software.[17] This was a drive to start a more business-friendly approach that came about not only because of problems of duality with the word "free" in English, but also because of a need to distance themselves from the movement of the openly political Richard M Stallman.

The feeling that the openly political approach could hurt the growth and possibilities of the Open Source software seems deeply ingrained in the movement. Some people have even gone so far as to term themselves "Oppenheimers" without implying any kind of derogatory meaning whatsoever. In this sense, an "Oppenheimer" is someone who builds a tool, without taking any kind of responsibility for it's use. Obviously hydrogen bombs don't kill people, people kill people …

The much sought-after legitimacy is gained by normatively denying the political aspects of creating software and instead focusing on "building the best OS there is". Normatively, what is the best OS is something that can be understood inside the movement, by the code-writers, much like lawyers claim to be able to figure out what is legally right or wrong.

Because of its development-structure, Open Source projects are less normatively closed than other kinds of software development, but some of the projects are becoming increasingly closed. Possibly in an attempt to woe big business. This paper makes no comment on whether this strategy is a successful one, since that still remains to be seen.

10. Open Source management

Open source management is a term used to try to adapt the lessons of the success of Open Source software together with an approach developed at the Centre for Intellectual Property studies call Norm Constructionism. This approach was developed partly in an attempt at solving the dilemma of the social constructionists that we have discussed above. Norm constructionism is an entrepreneurial approach as well as a theoretical approach. The following closing paragraphs are a very brief overview of the norm-constructionist project.

Concisely, the norm constructionist approach is based on two tenets. The first one holds that institutional phenomena can be classified as either structural tools or structural building bricks. If, for example, the concept of trade secret as it is expressed in a piece of legislation, is used to claim some piece of information as a specific trade secret, then that concept is a structural tool. The claimed trade secret is now a structural building brick.

Concepts are developed in the course of court procedure, during corporate business luncheons and in the academy, as well as many other places. (Virtual concepts are developed by Linux programmers, for example.) The concepts are tools that enable entrepreneurs to perform normative claims that will, if successful, have normative consequences.

The second tenet is that structural phenomena do not have any existence in themselves, and cannot be described without also being influenced. As a consequence of this insight, a displacement of what is normative and what is descriptive follows. Attempts to describe structural phenomena include, in the end, being normative, to a lesser or larger degree. Entrepreneurs, lawyers and academics are used to taking advantage of this possibility. When a lawyer describes an intellectual property she uses her **normative space**, often in the interests of her client. The normative space within a concept is limited when the concept is very commonly known and not-

[17] Moody, Glyn, p168: "Rebel Code – Inside Linux and the Open Source revolution": Perseus Publishing, 2002.

particularly complex, just as the normative space is wide-stretched within concepts that are vague and complex. The normative space is also wider within concept that exist in areas where the implicit regulation is changing rapidly, making their function less understood.

Marketing managers who describe trademarks, concepts, services etc. and analysts who describe companies, lines of business, currencies, shares and markets have a corresponding normative space, wherein they are free to influence the anatomy of these concepts. Theorists, too, use this space when they analyze and describe contract terms, companies, the efficiency of markets etc. All the claims result in norm experiences. When, for example, software companies claim that they can patent their software, the normative space is used to modify the concept of a patent, and in the end we have software patents. Even software developers have a normative space when constructing features, or leaving them out. The boundaries of normative space are set depending on how strong the consensus is within a given field regarding such internally-accepted values as speed, stability, interoperability, scalability etc.

Regarding the enigma of social constructionism, the norm-constructionist approach advocates a two-fold approach to the dilemma of how to reconcile this social constructionist view with preserving the useful characteristics of the system. Very briefly, the approach requires the social-constructionist insight to be coupled with an insight of the usefulness of the institutional concepts and constructs. From this intermingling, a loyalty toward the structure should flow: a normative approach where using constructionist insights to trash the system for short-term benefits are refrained from due to long-term needs. On one hand, the workings of the system are discussed in all their squalid splendor and on the other hand the long-term usefulness of the system is normatively discussed. The success of the Open Source development model does much to raise hopes that this might actually not be impossible for legal practitioners and scholars.

For the future of Linux and DRM, the norm-constructionist approach means that an open and sincere discussion of the social implications of a piece of code must be acknowledged as a valid topic of discussion in kernel development, alongside such values as speed, stability and interoperability and that the role of the software developer is not altogether different from that of a lawyer. At the same time the progress and adoption of Linux within the corporate environment must be preserved. This could either mean voluntarily keeping a low profile or raising the general level of awareness that platform technologies are irrevocably related to societal policy considerations.

I wish to end this paper by paraphrasing Jürgen Habermas:

"*The principle of OS development seems to mediate between the principle of technology and that of democracy. But it is not entirely clear how the latter two principles are related.*"

DigiRight: Network of Excellence for a Framework for Policy, Privacy, Security, Trust and Risk Management for Digital Rights Management

H. Abie[1], J. Bing[2], B. Blobel[3], J. Delgado[4], B. Foyn[1], S. Karnouskos[5], P. Pharow[3], O. Pitkänen[6], and D. Tzovaras[7]

[1]*Norwegian Computing Center, Norway, {habtamu.abie, bent.foyn}@nr.no*

[2]*Norwegian Research Center for Computers and Law, Norway,* jon.bing@jus.uio.no

[3]*University of Magdeburg, Germany, {Bernd.Blobel,* Peter.Pharow}*@medizin.uni-magdeburg.de*

[4]*DMAG-TECN, Universitat Pompeu Fabra, Spain, jaime.delgado@upf.edu*

[5]*Fraunhofer Institute FOKUS, Germany,* Stamatis.Karnouskos@fokus.fraunhofer.de

[6]*Helsinki Institute for Information Technology (HIIT), Finland,* olli.pitkanen@hiit.fi

[7]*ITI, Center for research and Technology Hellas, Greece,* Dimitrios.Tzovaras@iti.gr

Abstract

In today's digital world there is an enormous and increasing amount of digital content. In the future world of ambient intelligence, digital content will be ubiquitous and people will interact with it in all areas of their lives, a situation that presents new challenges in the area of Digital Rights Management (DRM). While valuable information products need protection from theft and prying eyes, access to information and the ability to contribute to information products and to share information within communities are also essential to all citizens of the information society. The needs for security and privacy are predominant in such situations. All of this is making DRM crucial. Therefore, we proposed to establish a Network of Excellence for a Framework for Policy, Privacy, Security, Trust and Risk Management for DRM, DigiRight, which will consist of experts from various disciplines and will conduct and guide on-going and future high quality research.

1. Introduction

Today's wired and wireless digital world has yielded a massive and increasing amount of digital content. Indeed, in the future world characterized by ambient intelligence, digital content will be ubiquitous, and people will interact with it in all spheres of their personal life, social activities and work, even in situations where they may not realize it. All this presents us with new kinds of challenge in the area of DRM.

Information and communications technologies (ICT) provide us not only with evermore powerful means to develop and distribute information products, but also with means to copy-protect data and restrict its availability. On the one hand valuable information products need protection from theft and prying eyes. On the other hand, access to information and the ability to contribute to information products as well as to share information within communities, are essential to all citizens of the information society. While efficient business methods require collecting detailed information on transactions, business partners and customers, the need for privacy of all stakeholders must also be respected. The amount of sensitive information that must be securely stored, shared, or distributed within and between organizations is also rapidly increasing. Striking the balance between the appropriate level of security and the protection of user privacy and enabling users to control how personal identifying information is to be stored, distributed, and used, is crucial. All of this is making DRM crucial.

Digital Policy Management (DPM) is becoming a discipline in its own right, whose concern is the design, analysis, implementation, deployment and use of efficient and secure technology that handles digital information in accordance with the relevant rules and policies. These policies are based on the security requirements of digital information, which in turn are based on rigorous analysis of risks, its vulnerability, and threats to it. Thus, since the improvement in the implementation of policy depends on an improved risk management process, any DRM research must give full attention to the improvement of risk management process, and risk assessment methodologies. Consequently, security, trust and privacy policies must be developed and integrated into the DPM-enabled DRM system (DRMS). Furthermore, seamless interoperability of DRM solutions across fixed and wireless networks and infrastructures need to be addressed.

Therefore, we need to establish a Network of Excellence (NoE) [1] for a Research Framework for Policy, Privacy, Security, Trust and Risk Management for DRM, viz DigiRight [2]. It will consist of individual experts from various research institutes and organizations having expertise in the fields of technology, law, business, social science, ethics, policy-making, and security. As the issue is very complex, an NoE is needed in order to conduct and guide on-going and future high quality research. The description of

DigiRight's relevance and potential impact may be found in [3]. The ubiquity of digital content means that DRM concerns almost everyone, from authors and publishers, to consumers, libraries, schools and educational institutions, infrastructure providers, hardware and software manufacturers [4], and governments and standard bodies. Therefore, any DRM related research must take into account both the complexity of disciplines and the concerns of the various stakeholders.

This paper describes the DigiRight NoE, which will meet these requirements and has been submitted under the Sixth Framework Programme for the first IST Call. Section 2 and 3 describe the DigiRight objectives and integrated DRM research framework, respectively. Section 4 describes the scenario methodology for making the goals operative, and the plan to establishing a virtual DRM research Center, and section 5 concludes.

2. The DigiRight objectives

The overall goal of DigiRight is to develop a synergy Research Framework for Policy, Privacy, Security, Trust and Risk Management for Digital Rights Management with an ultimate goal of establishing a virtual DRM research Center. The purpose of the DigiRight research Framework is to

1. integrate the traditionally separated DRM research communities across Europe (both at national and regional level) in the fields of technology, business, law, ethics and social science (all of which are important operative factors in the uptake of DRM), and to structure the way DRM research is carried out in the research community and amongst practitioners by networking together teams of experts in these fields;
2. stimulate joint scientific research projects to gain insights into the fundamental issues and challenges associated with future DRM systems, exchange of research personnel, harmonization of DRM technologies and solutions, and learning programs at the European level;
3. create a self-sustainable set of knowledge-spreading activities through liaison with end-user communities, industries, standard bodies and governmental organizations, and a solid two-way technology transfer between the industries, standard bodies, and governments;

The final goal is to establish a virtual DRM research center with the aim to develop solutions, guidelines and standards to protect, manage access rights (including the evolution, emergence and negotiation of the new rights of the e/m-society) to, control usage of, and distribute trustably tangible and intangible digital assets without risking users' privacy, and hence to stimulate the development and use of European digital content on the global networks promoting the linguistic diversity in the Information Society. In particular, we shall address the new challenges presented by new broadband access networks and mobile telephony, thus enabling content providers and technology companies to publish information on any Internet platform, from the web to wireless devices, to Internet appliances and broadband television. Through all this, we aim to build customers' trust and confidence so that the Intellectual Property Rights (IPR) business will flourish on a global scale.

3. DigiRight: An integrated DRM research framework

The primary feature, which assures a coherent integration, is the well-defined collective goal, which can be simply stated as DRM. The topic itself, DRM, is an extremely motivating goal for researchers and an attractive product for the public. However, the research necessary to achieve this goal is, by its nature, highly complicated and diverse, and can thus not be conducted without steps being taken to integrate it and bringing together relevant, complementary researchers. The necessity of a well-coordinated large and diverse research group to achieve this goal strongly discouraged researchers for a long time.

DigiRight will therefore network experts in the different disciplines necessary for a holistic view and understanding of DRM. For each discipline a task force has been created, a task force of experts within each discipline, who will be responsible for on-going and future high quality research into those aspects of the discipline concerned, which are relevant to DRM. The task forces will co-operate with each other on joint research using common concepts, methodologies and tools that will be developed and synthesized from components taken from jurisprudence, the social sciences, business theory and economics, and science and technology. This integration of interdisciplinary approaches and ensuing technologies will provide the Network with a common background and basis for combined research, and facilitate the exploitation of the synergy of the various projects, areas of expertise and stakeholders. Intellectual property (IP) asset creation, IP asset capture, IP asset management, and IP asset usage [5] control and tracking will be handled effectively as common domain platform services. Standards will be developed to allow interoperability so as not to force DRM users to encode their works in proprietary formats or systems.

DigiRight will concentrate mainly on technology. In this connection it is important to note that the object is not merely to develop and implement DRM technology, but also to ensure that it is widely used. This will require a reliable and secure infrastructure, and will depend on users' (citizens, businesses, communities) trust and confidence in the technology which provides them with controls fine-tuned for the balance of, on the one hand, privacy and security, and, on the other,

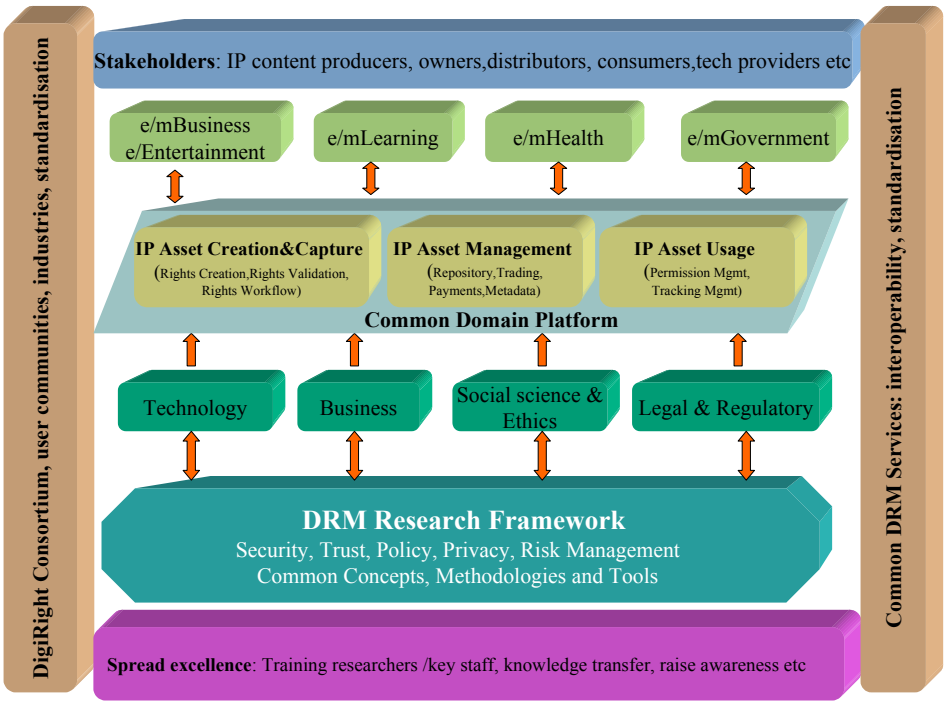


Figure 1 – DigiRight research framework

accessibility and usability that they need. It will also require correct attention to be paid to privacy, policy, security, trust and risk management in DRM, and must be addressed from technological, business, legal, ethical and societal perspectives. Figure 1 depicts the DigiRight DRM research framework with some of its major components.

3.1. Technology

The overall objective of the Technology Task Force is to contribute to common DRM research methodology, integrating and spread of excellence activities from the technology perspective. This will involve among other things:

- identify and analyze the relevant technological challenges and solutions to DRM application scenarios in question;
- bring forward existing and lacking knowledge in technology;
- describe the technology requirements, solutions, and obstacles;

The Technology Task Force will concentrate on the following seven central aspects: privacy, policy, security, trust management, risk management, protection mechanisms, and information representation semantics.

Privacy enhancing technologies: The need for privacy is predominant in any core business. The next generation of DRM will cover the description, identification, trading, protection, monitoring and tracking of all forms of rights of usage over both tangible and intangible assets, and would manage rights holders relationships [5]. The ability of this next-generation DRMS to track and monitor will lead to a need for more efficient mechanisms for the protection of personal privacy, protection that the DRMS itself must ensure. Although there are those who claim that this is a red herring on the basis that such privacy is protected and guaranteed by law, it should be pointed out that unscrupulous manufacturers and individuals may be technically capable of violating privacy undetected and therefore unpunished.

The aim of this activity is to investigate approaches to protecting the privacy of individuals, groups, and even companies and governments, and strike the balance between tracking usage and user privacy, and enable consumers to control how personally identifying information is obtained and used [6, 7, 8]. Essential challenges are:

- Personal information privacy: What personal information can be shared with whom;
- Digital assets privacy: Whether digital assets can be exchanged without anyone else seeing them;
- Anonymity: Whether and how one can send messages anonymously, and whether this should be permitted or is desirable;

- Anonymity vs. Accountability: How accountability and anonymity can be balanced to allow user control as much as possible, community norms when users' desires conflict, and government regulation when the norms of the communities differ [9];
- Provide controls fine-tuned for the balance of, on one hand, privacy and security, and, on the other, accessibility and usability that users need;

Digital Policy Management (DPM): DPM's concern is the design, analysis, implementation, deployment and use of efficient and secure technology that handle digital information in accordance with the relevant rules and policies. Brose et al. [10] have also proposed a systematic approach to integrating security policy design into the system development process. The aim of this DPM activity is to investigate different trust and privacy policies that must be developed and integrated into the DPM-enabled DRMS. This digital policy can for example be embedded in a mobile software component, which may provide services helping authenticate and authorize use of the digital content and regulate what the user is allowed to do with the content. For the DRM policy part of the NoE, an architecture [11] is proposed in which the intellectual property rights owners (e.g. content providers) are associated to a broker that is in charge of exploiting (selling) their content rights, and, once those are sold, of controlling that the rights are respected; i.e., no illegal copies are circulating on the Internet.

Security architecture and infrastructure services: The problem of protecting digital information from unauthorized distribution is the concern of many rights holders, content providers and distributors. The function of this activity will be the investigation of DRM-enabling security architecture and infrastructure services as a basis for DRM applications. The aim of the security infrastructure is to enable valid users to create, distribute, store, manipulate and communicate information objects across organizational boundaries with the required level of security [12].

In order to achieve DRM solutions that are interoperable and standard-based as well as applicable in different domains, a common infrastructure platform for the DRM technology and enabling basic security services is required at both the application level and infrastructure level. Openness and interoperability lead to a seamless inter-connection and co-operation of security services. Communication security services comprise strong mutual authentication and accountability of principals involved, integrity, confidentiality and availability of communicated information as well as some notary services. Application security services concern accountability, authorization and access control regarding data and functions, integrity, availability, confidentiality of information recorded, processed and stored as well as some notary services and audit. DigiRight will address content qualities that can be managed semi-automatically properties such as integrity, confidentiality, authenticity, and trustworthiness in DRM. Specific challenges include:

- Research on the application of cryptographic technologies / Public Key Infrastructure (PKI) for the IPR protection;
- DRM-enabling security infrastructure as a basis for DRM applications;
- Design, analysis, and implementation of an advanced architecture and related security protocols for a distributed DRM in seamless environments;
- Integration of Biometrics and Smart Cards for DRM applications;

Trust Management: Trust is an essential factor in any business-transaction systems including DRM systems. To wit providers need to establish trust and confidence in their products and services, and consumers need to protect their privacy and information and assess the trustworthiness of their providers. Lack of trust in the ability of DRM infrastructure to protect IPR is a significant barrier to growth in the IPR business transactions. Usage Tracking is essential for providing trust for content providers. At the same time, the user must be able to trust that a service will not violate his/her privacy, and be sure that the service quality is the agreed upon one. Understanding user concerns related to trust and confidence has a key role in the work of DigiRight. In addition, DigiRight will engage in standard setting operations, which help to define a DRM architecture, which meets the security and dependability concerns of the users. Thus it is essential to facilitate the cross-disciplinary investigation of fundamental issues underpinning trust models by bringing together expertise from technology oriented sciences, law, philosophy and social sciences. Activities include:

- Develop formal social cognitive theories of trust and reputation, and explore the role of reputation in the evolution of altruism and co-operation in human societies;
- Apply the trust models to agent societies [13];
- Test theory-driven hypotheses about the effects of different types of reputation systems by means of simulation–based and natural experiments, also in view of optimizing existing online reputation reporting systems;
- Facilitate the emergence of widely acceptable trust management processes for open DRM systems and applications;
- Explore the role of attitudes towards a DRM-based transaction, which is defined as the overall evaluation of the desirability of a DRM-based transaction with an agent. The aim is to develop a trust model that will help each user to judge whether authenticity and provenance evidence of the transaction make a digital content sufficiently trustworthy;

- Model and simulate human factors regarding trust and security to understand the real background of the trust phenomenon;

Risk Management: Risk management holds the key to security: A security policy is necessary to support the security infrastructure required for the secure movement of sensitive information across and within national boundaries [15]. To ensure the secure operation of this kind of infrastructure, it is necessary to have some well-founded practice for the identification of security risks as well as the application of appropriate controls to manage risks. The risk management process provides a framework for identifying risks and deciding what to do about them. Risk management is not a task to be completed and shelved. It is an ongoing process (with well-defined steps [16, 17]) that, once understood, should be integrated into all aspects of an organization's management.

Trust management also relies on risk management: quantification of trust based on systematic methods for threat identification and risk analysis may offer better evaluations of DRM transaction. Risk in the digital environment is typically influenced by the organisational structure and circumstances [18] that affect human interaction (situational trust), beliefs and inclinations (human centric trust), and confidence on technology infrastructure in place (computer centric trust). Risk management thus allows us to combine risk with trust in order to form a security policy [18]. Furthermore, DRM and content distribution industry related companies would require risk management strategies and tools to protect vital assets. The application of risk management disciplines will help identify, assess and control risks relevant to the distribution of digital content. Sound risk management will help create a sense of confidence and safety about an operation. In an environment where the threat of unnecessary risk is reduced, services can be more creatively provided to clients and better results can be achieved, hence company/institution safety and security.

Consequently, the essential challenges are:

- Building appropriate balance between trust, privacy, policy and risk management for DRM with a balanced legal framework that takes account of the change in the academic, political, economical and socio-cultural model while at the same time safeguarding fundamental rights, freedoms, fair-use, and private-use in the digital world.
- Future possible risks related to information in digital form must be managed in advance in several ways [19].
- Research regarding risks and threats specifically related with digital rights management, in order to enhance the risk management procedure and ensure its completeness and research to manage risks involved in participating in DRM transactions thereby building trust in those transactions.
- Risk management methodologies for IPR protection development – especially the creation of knowledge bases with specific risks and control for addressing the risks.
- Research on DRM scenarios to qualitatively and quantitatively support appropriate decision making for minimization of risks, based on system dynamics based modeling and simulation.

Protection mechanisms: watermarking, encryption and fingerprinting - technical solutions are required to restore some control over the identification of original content, the monitoring and tracking of the use, and the management of distribution/communication channels. There are techniques to identify original content such as hash codes in digital files, watermarks in images and hidden sound codes in music files, and encryption to secure communication and distribution. This activity will investigate protection techniques including:

- Watermarking (1D / 2D / 3D multimedia data), combining watermarking with indexing;
- IPR protection of data between Internet and mobile telecommunications systems, using encryption and watermarking;
- Accountability mechanisms. Accountability is a more challenging goal for distributed or peer-to-peer systems or networks with a transient population of users, where it is hard to identify user identities and obtain information about their past behavior in order to predict their future performance;
- Reputation mechanisms. The notion of reputation can be employed in a variety of mechanisms as a means of providing fairness and balanced use of resources;

Information representation semantics: In order to improve the management of rights in the digital environment (DRM), there is a need for a common language for DRM representation in the open and global framework provided by the Web. This kind of language is aimed to help building a reliable Web where IPR can be managed in an open, global and adaptable form, so people can share, sell, buy, etc. content subject to DRM, depending on their needs. A semantic approach seems a more flexible and efficient way of achieving these activities than a syntactic one.

Using metadata for referencing multimedia material is becoming more and more usual. This allows better ways of discovering and locating this material published in the Internet. Several initiatives for establishing standards for metadata models are being carried out at the moment, but each focuses on their own requirements when defining metadata attributes, their possible values and the relation between them. For someone who wants to seek and buy information (multimedia content in general) in different environments, this is a real problem, because he/she has

to face different metadata sets, and so, must have different tools in order to deal with them. A DRM ontology can put into practice this approach, endowing agents with more complete background knowledge, which allows them to work quite autonomously.

The idea of this NoE is to facilitate the automation and interoperability of DRM frameworks integrating both parts, called Rights Expression Language and Rights Data Dictionary. This can be accomplished using ontologies. They can provide the required definitions of the rights expression language terms in a machine-readable form. Thus, from the automatic processing point of view, a more complete vision of the application domain is available and more sophisticated processes can be carried out.

3.2. Business processes and models

Connector in the field of business processes and models is the detailed analysis of all involved acting parts. On the one hand there are the rights holders, which are a heterogeneous group with acting parts such as authors, agencies, and publishing houses, which follow different aims and are connected on to each other in complex relationships. On the other hand, the target markets are also highly heterogeneous. In this area of tension varying business models are formed, which are distinguishable by achievement and revenue. According to the Oxford dictionary process is a method of producing goods in a factory by treating raw materials. A business model [20] is a description of how a company intends to create value in the marketplace. It includes unique combination of products, services, image, and distribution that a company carries forward, and the underlying organization of people and the operational infrastructure that they use to accomplish their work.

The objective of the business models task force is to be able to analyse and study business models' aspects of the scenarios in question. This activity should identify relevant research and results for the selected scenarios in order to bring forward existing and lacking knowledge. The product of this task force should be a report with analyses of what could be done from the business models' side of view to realise the scenarios, and where the major obstacles are believed to be. The main research challenges to be addressed in this activity are negotiation, contracting, and production processes, publication, and data models.

Negotiating: The negotiation protocol, that it is part of the "Service Request" phase in an e-commerce model, has three sub-phases: Initial offer, co-operative contract production, and payment. In the Contract production sub-phase, the most complex and important one, there are several alternatives over which to work. First, the selling entity initiates the protocol with an initial proposal of digital rights conditions, normally taken from a pre-defined subset. After that, the buying entity has three alternatives: Accepting the offer, making a counter-offer and rejecting the offer. After the initial proposal, the negotiating entities elaborate the contract, using the negotiation protocol, from the sequence of offers and counter-offers until a final agreement is reached, forming then the final electronic contract.

Contracting: By DRM negotiation we mean the process in which, at purchase time, the buyer of some multimedia content and the rights owner (or representative) negotiate the conditions (concerning rights) in which that material is sold. This process, run through a protocol with some interchange of information, is equivalent to creating an electronic contract. It could be also seen as a joint editing of a structured document (the contract), following pre-specified alternative rules. The electronic contract, that should be electronically signed, has two parts:

- Mandatory part: It contains the minimum information necessary to formalize an electronic contract.
- Optional part: It contains optional information related to any kind of contract.

Production processes, publication, data models: Publishing houses and media companies are developing the opportunities of expanding their own competitive position with the aid of innovative products and services, and for acquiring entirely new business segments. At the same time, they are confronted with a lack of systematic processes and methods, which bear in mind issues of DRM. Such processes are essential above all to develop successful products and services. The aim is to allow publishing houses and media companies to prepare and design content for publication in a manner, which is manageable by typical midsize companies. This demand results from changing possibilities of data storage and the big expectations in the field of media products.

The question is therefore, how production processes and DRM can be integrated in this complex field of media production. For that reason a model for reference processes and a model of production have to be developed. These models consider co-operation within publishing houses as well as co-operation between companies; they should allow multiple uses of content through standardized asset management and support the use of integrated information systems along the production processes.

3.3. Legal and regulatory, private and public policies

The objective of this activity is to analyze and study legal and societal aspects of the DRM scenarios in question. The Task Force should identify relevant research and results for the selected scenarios in order to bring forward existing and lacking knowledge. The most important research challenges in the area of legal, regulatory, policy and societal aspects are the following four central aspects.

Data protection: The task will be to identify IPR in the terms of elementary actions which require the consent of a right holder, i.e. to "copy", to "public performance", to "systematically access and extract elements from a data base", *etc.,* [13,14]. There are also fundamental unsolved issues related to IPR in new kinds of information products. Within this task, we are going to integrate the participants' excellence in understanding which intellectual property rights are applicable to different information products and which parts of the products are protected. For data protection, one will have to identify in which way to obtain a relevant consent from a data subject, or alternatives in obtaining the right for the processing of the personal data involved. This will especially be a challenge in the health care sector, where the data will be of sensitive nature, but is also of growing significance in the telecom sector. Though coordinated by the data protection directive, different national statues have implemented the provisions rather differently, especially with respect to sensitive data, of which processing in many jurisdictions is subject to license from a national data protection authority. Therefore, the inter-legal issues (jurisdiction and choice of law) have to be included.

Content policies: Content policies are developed on the basis of the recent directive coordinating national copyright and related rights. "Content" is a facile term covering a variety of material in different legal categories, copyrighted material, material subject to neighboring rights, especially the rights of performing artists, producers and database builders. Content is usually the part of an information product without which the product has no value. The other parts, like metadata or programs, however, may add value to the content. It is not possible to precisely define the concept of content. As there can be tremendously many kinds of information products, also content can differ a lot. It can be nevertheless described as the actual payload of the information product. For example, a computer program as such can be an information product. On the other hand, as a part of a multimedia product, it does not necessarily need to be something without which the product has no value, but is merely a value-adding auxiliary part. Therefore a program may or may not be content. It should be noted that not only commercial publishers produce information products or content, but using modern information technology it will become more common that authors themselves distribute their works and the end-users, on the other hand, contribute to the content. Often the subject for trade is not content, but the legal position related to the content, allowing the purchaser to exploit the content according to terms specified in a license, which also will include remuneration.

Ethical aspects: Legal rules may not be sufficient for business models to operate, but will have to be bolstered by more restrictive ethical rights. Especially for data protection, one should make explicit the trade practices. The identification of human individuals is one of the most difficult ethical issues. Technically, it is difficult to reliably relate any physical identification to a human being. However, that is a small problem compared to legal and ethical issues related to privacy, anonymity, and identity. In general, everybody should be able to remain anonymous and to keep privacy. On the other hand, a human being may act in a large number of roles. A person at work, at home, at leisure activities and so on has many roles that should be distinguished. For example, usage rights like private use or fair use are often different depending on the role and a license may only cover certain role-based usages. Therefore it is hardly possible to build solutions that in general rely on human beings direct identifications. Instead, most systems need to depend on indirect user identification based on for example device identification.

Consumer rights and expectations: There are latent but growing tensions between the actors involved, especially where DRM may restrict the use of "content" with respect to end user equipment (only authorized DVD-players). An example of consumer protection issues related to DRM is one with rights description languages (e.g. ODRL, XrML). It is possible to describe very complex sets of rules using those powerful and expressive languages. A rights description resembles a computer program. For a human, it can be very difficult to understand what those complex sentences mean. However, when somebody buys an information product, it is essential what rights are licensed or assigned. Even if the customer gets the right data, but does not get the rights needed, the customer does not get what was expected. In accordance with consumer protection laws, it is important to inform a consumer in advance what is to be sold. It must be possible to cancel the transaction if the consumer does not get what was anticipated.

3.4. Societal questions

A balancing act of the rights of the provider or right holder and the end-user must be made in the perspective of the society, where promotion of electronic trade may be a separate policy objective. The European Commission has announced that bringing every European online and into the digital age, creating a digitally literate Europe, and ensuring that the whole process is socially inclusive will be the key objectives in bringing an information society for all the Europeans. This raises important societal aspects on DRM. DRM systems that unnecessarily prevent people from accessing information or increase the digital divide between population groups are not welcome. Instead, future DRM systems should actively help to achieve the above goals.

One of the key issues in the societal area is the rise of user communities. Users themselves contribute to content and share information and resources. A topical

example is gaming communities in which players around the world develop the games and play them together. Another example is open source movement: software engineers without any formal organizations create programs together and distribute them freely. This model will enlarge and cover many walks of life.

3.5. Application domains and stakeholders

There is a lack of communication between application domains. Practitioners in one domain are frequently totally ignorant of the activities of their peers in others, and are quite capable of producing the most exciting results without sharing them, and on occasion, after someone else has produced them without bothering to tell anyone. How often has the wheel been reinvented? This is due to the unfortunate fact that the results are neither disseminated through the right channels nor, more importantly, in a cross disciplines. The same concepts and ideas often apply in many different areas, and those few of us who have managed to abstract these concepts from one domain and apply them to the problems of another have often gained wonderful results.

Therefore DigiRight will pay special attention to inter-domain communication and co-operation as it meets those challenges highlighted by the Commission as top priorities for Europe in the coming years in the following domains **e/m-business, e/m-entertainment, e/m-learning, e/m-health, e/m-government,** and **e/m-generic-services** with the objective of ensuring that all stakeholders including producers, owners, distributors/retailers, users, technology providers enabling the delivery, and hardware and software companies enabling the consumption of intellectual property (IP) content, are all winners. Thus, in DigiRight all domains relevant to the information society will be represented by domain experts among the partners reflecting specific challenges, needs and solutions.

Therefore DigiRight will attempt to address any stakeholder in any business chain. DRM is a key part of the future platform for application and service provision. A DRM architecture that balances the interests of the various stakeholders will be a key enabler of new applications; an ill-balanced architecture is a showstopper.

4. DigiRight: Scenario methodology, integrating process, and a virtual DRM research center

4.1. Scenario methodology – making the goals operative

DigiRight aims at studying future systems that involve many disciplines whose systems do not exist today, so they cannot be observed directly. At first sight, it seems that, for instance, legal challenges related to the systems should be analyzed using the methods of legal science. However, the challenge is about forthcoming issues while legal science mostly uses court cases, statutes, and their preparatory works as its sources and derives theories by analyzing them. Thus it is hardly possible to tell almost anything about the future using conventional methods. Instead, future research provides us with more suitable methods. Especially scenarios are useful when we want to describe how the world will be like. In addition to providing us with adequate research method, scenarios are excellent means to integrate and communicate ideas, views and concepts. The participants will be able to share common understanding and disseminate outcome using clear, explicit scenarios. Scenarios used in other fields of science are typically quite broad. On the other hand, sometimes it is useful to create very small scenarios or use-cases. In this network, we expect the scenarios to be relatively narrow: they will merely describe a possible service that is grounded on participants' research, literature, existing services, and discussions with other experts. However, there may emerge needs to develop also very small or huge scenarios.

We do not claim that any of our scenarios would actually come true. Neither is their actual probability of being realized in the focus of work. Instead, they are to form a picture of possibilities and concerns that may exist in the future. In the network of excellence, scenarios will be used as means of integrating the excellence of various partners, defining research areas, accomplishing actual joint research work, and disseminating the conclusions. The scenarios will be updated and new scenarios will be created as we are making progress.

4.2. DigiRight integrating process

DigiRight aims at developing a synergy research framework whose purpose is to structure the way DRM research is carried out in the research community by networking together teams of experts in the fields of technology, business, law and social sciences. The provision of such a Framework is expected to become a critical instrument for attracting researchers and practitioners to DRM issues. DigiRight needs to address DRM from all sides, identify where there are obstacles to overcome in order to realize services that use DRM. It will achieve its goal through a number of carefully planned activities, which collectively bring a high degree of long lasting integration. Figure 2 depicts the DigiRight integrating process/cycle with the main activities and task forces.

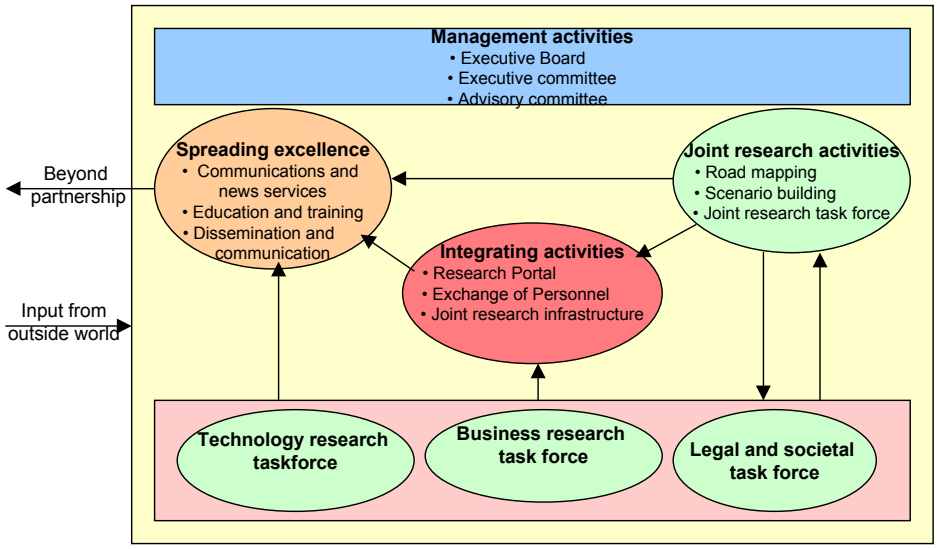


Figure 2 - Integration process

4.3. Establishing a virtual DRM research center

The objective of this activity is to ensure that the Network activities continue to be adequate for DigiRight to approaching a virtual organization that will continue to exist after the cessation of Community funding. This will first of all be a result of all the other activities in DigiRight. The reason for having this activity is to be able to address important questions from this side of view that might not be taken sufficiently into account in the other activities. Examples of important questions are:

- How to ensure that DigiRight becomes the preferred unit of co-operation within DRM research in Europe;
- How to create services that will secure the economic basis for funding when the EC financing is terminated;
- How to ensure sufficient anchoring of DigiRight in international organizations that run conferences, standardization work and other scientific activities and within the most important partners in DRM research;

The ultimate goal of DigiRight is to create one single virtual research organization in DRM issues across Europe in order to co-ordinate DRM research in the future. This virtual organization should span the different traditional borders of research such as technology, legal & regulatory, societal questions, and business processes and models.

5. Conclusions

In this paper we have described DigiRight, a Network of Excellence proposal for a DRM research framework, which aims to

1. integrating the traditionally separated DRM research communities across Europe (both at national and regional level) in the fields of technology, business, law, ethics and social science all of which are vital to understanding the issues related to future DRM and its use;
2. stimulating joint scientific research projects to gain insights into the fundamental issues and challenges associated with future DRM systems;
3. creating a self-sustainable set of knowledge-spreading activities through liaison with end-user communities, industries, standard bodies and governmental organizations;

The DigiRight NoE is an integrated approach to address the **trust** and **confidence** in communication, e/m-business, e/m-entertainment, e/m-learning, e/m-health, e/m-government, and e/m-generic-services, and the support to solve complex problems in science, society, industry and business objectives. It is our considered opinion and firm conviction that such an integrated research framework will be a much-needed shot in the arm for the understanding and uptake of knowledge-based digital economy.

Acknowledgements

The authors would like to express their gratitude to the DigiRight partners [2] for their contributions and efforts during the proposal writing, and Martti Mantyla

for his review of the DigiRight proposal and for his useful comments.

References

[1] Research DG, European Commission, Provisions for Implementing Network of Excellence, Background document, Draft 2002 edition: 11 November 2002

[2] Network of Excellence for a Research Framework for Privacy, Policy, Security, Trust and Risk Management for Digital Rights Management, a proposal for network of excellence under FP6 for the IST Call 1, submitted to EC, 24/04-2003, http://digiright.nr.no/nuke/html/

[3] H. Abie, B. Blobel, J. Delgado, S. Karnouskos, R. Marti, P. Pharow, O. Pitkänen, and D. Tzovaras, DigiRight: Relevance to and Potential Impact on Europe's Need to Strengthen the Science and Technology Excellence on DRM, Mobile IPR, HIIT, Finland, August 27-28, 2003.

[4] M. Fetscherin, Present State and Emerging Scenarios of Digital Rights Management Systems, JMM – The International Journal on Media Management Vol. 4 – No.3, pp 164-171, 2002

[5] R. Iannella, Digital Rights Management (DRM) Architectures, D-Lib magazine, June 2001, Vol. 7, No 6, http://www.dlib.org/dlib/june01/iannella/06iannella.html

[6] J. Feigenbaum, M. J. Freedman, T. Sander, and A. Shostack, Privacy Engineering for Digital Rights Management Systems, November, 2001, http://www.pdos.lcs.mit.edu/~mfreed/docs/privacy-engineering.pdf

[7] J. E. Cohen, DRM and Privacy: http://www.law.berkeley.edu/institutes/bclt/drm/papers/cohen-drmandprivacy-btlj2003.html, 2003

[8] Electronic Privacy Information Centre. Digital Rights Management and Privacy: http://www.epic.org/privacy/drm/

[9] L. J. Hoffman and K. A. M. Carreiro, Computer Technology to Balance Accountability and Anonymity in Self-regulatory Privacy Regimes, Cyberspace Policy Institute, School of Engineering and Applied Science, The George Washington University

[10] G. Brose, M. Koch, and K.P. Lohr, Integrating Security Policy Design into the Software Development Process, Technical Report B-01-06, Institut fur Informatik Freie Universitat Berlin, Germany, November 13, 2001

[11] J. Delgado, I. Gallego, and X. Perramon, Broker-Based secure Negotiation of Intellectual Property Rights, ISC'01, LNCS 2200, Springer-Verlag, 2001

[12] H. Abie, A Rights Management Model for Distributed Object-Oriented Information Distribution Systems, Proceedings of the IFIP WG6.7 Workshop and EUNICE on Adaptable Networks and Teleservices, September 2-4, pp. 185-194, 2002

[13] J. Bing, The contribution of technology to the identification of rights, especially in sound and audio-visual works: An overview, Norwegian Research Centre for Computers and Law, University of Oslo, Norway

[14] J. Bing, Intellectual property exclusive rights and some policy implications, Norwegian Research Centre for Computers and Law, University of Oslo, Norway

[15] Risk Analysis Resource Page, Norwegian Computing Center, http://www.nr.no/~abie/RiskAnalysis.htm

[16] AS/NZS 4360:1999, Risk Management, Australian Standard, 12 April 1999-09-17

[17] Norwegian Standard, NS 5814, Requirements for Risk Analysis, August 1991

[18] T. Dimitrakos and J. Bicarregui, Towards Modelling e-Trust, 3rd Panhellenic Symposium on Logic Anogia academic village, Crete, Greece, July 2001

[19] O. Pitkänen, and M. Välimäki, Towards a Digital Rights Management Framework, IEC2000, Manchester , UK , 2000

[20] H. Chesbrough, and R. S. Rosenbloom: The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-off Companies, Harvard Business School, To be submitted to Industrial and Corporate Change.

DigiRight: Relevance to and Potential Impact on Europe's Need to Strengthen the Science and Technology Excellence on DRM

H. Abie[1], B. Blobel[2], J. Delgado[3], S. Karnouskos[4], R. Marti[3], P. Pharow[2], O. Pitkänen[5], and D. Tzovaras[6]

[1]*Norwegian Computing Center, Norway,* habtamu.abie@nr.no

[2]*University of Magdeburg, Germany, {Bern.Blobel, Peter.Pharow}@medizin.uni-magdeburg.de*

[3]*Universitat Pompeu Fabra, Spain, {jaime.delgado, ramon.marti}@upf.edu*

[4]*Fraunhofer Institute FOKUS, Germany, Stamatis.Karnouskos@fokus.fraunhofer.de*

[5]*Helsinki Institute for Information Technology (HIIT), Finland, olli.pitkanen@hiit.fi*

[6]*Center for research and Technology Hellas, Greece, Dimitrios.Tzovaras@iti.gr*

Abstract

In today's fast changing digital environment where global communications transcend national boundaries, and where digital products are sold on international markets, the development of digital rights management (DRM) in order to protect digital assets is becoming an increasingly important global issue, and is emerging as a formidable new challenge since the different national laws, policies and practices must interoperate and be reconciled. To address these challenges, we have therefore proposed to establish a Network of Excellence for a Framework for Policy, Privacy, Security, Trust and Risk Management for DRM, DigiRight, that will consist of teams of experts on technology and development, business, law, ethics, and societal questions who will organize on-going and future high quality research on those areas of their respective disciplines that apply to DRM. This paper describes DigiRight's relevance to and potential impact on Europe's need to strengthen and reinforce science and technology excellence on DRM.

1. Introduction

In today's digital world there is an enormous and increasing amount of digital content. In the future world of ambient intelligence, digital content will be ubiquitous and people will interact with it one way or another in all areas of their lives, a situation that presents new challenges in the area of DRM. While valuable information products need protection from theft and prying eyes, access to information and the ability to contribute to information products as well as to share information within communities are also essential to all citizens of the information society. The need for security and privacy but also lawful usage of the content sets the basis for an ambient intelligence dominated future. Therefore, it is imperative to establish a Network of Excellence (NoE) for a Policy, Privacy, Security, Trust and Risk Management for DRM. DigiRight [1] is such an NoE proposal, which will consist of experts from various disciplines who will direct and conduct on-going visionary high quality research.

This cross-disciplinary interaction of experts on complex technical, legal, societal and business issues will determine how consumers will access content in the networked digital age and how rights holders will be compensated for and protected from unauthorized use of content, and how information will be shared between communities. In a perfect world, technology companies would be able to design their products as they thought best, copyright owners would be able to market their products as they think best, and users would be able to consume the content in their preferred way and in their preferred medium with minimal interference and cost. The economy would benefit from vitality of the content-based businesses, and indirectly the entire private sector at large. Low distribution and transaction costs would facilitate bridging the digital divide.

At the present, digital copying and redistribution have made these objectives incompatible, at least in part. The quest is for new generation of DRM technologies and related business models that best accommodate all these conflicting objectives. The first generation of DRM focused on security and encryption as a means for preventing unauthorized copying of content and limiting distribution to those who pay. The next generation of DRM should cover the description, identification, trading, protection, monitoring and tracking of all forms of rights of usage over both tangible and intangible assets, and would manage rights holders relationships [2]. In all this trust is essential. The ability of this next-generation DRM System (DRMS) to track and monitor will lead to a need for more efficient mechanisms for the protection of personal privacy, protection that the DRMS itself must ensure. Although there are those who claim that this is a red herring on the basis that such privacy is protected and guaranteed by law, it should be pointed out that unscrupulous manufacturers and individuals may be technically capable of violating privacy undetected and therefore unpunished.

DigiRight aims to create the combined understanding of these issues that will be a basis of the next-generation DRMS. This article will expand on DigiRight's relevance to and impact on the Sixth Framework Programme (FP6) [3] and Europe's need to strengthen and reinforce the Science and Technology (S&T) excellence on DRM. For the interested readers, a detailed description of DigiRight may be found in [1].

2. DigiRight: The sixth framework programme

"The current EU strategy adopted in Lisbon 2000 is focused on an accelerated transition to a competitive and dynamic knowledge economy capable of sustainable growth, with more and better jobs and a great social cohesion" [4]. The established European Information Society Technology (IST) Programme thus plays an important role towards the strengthening of European competitiveness. DigiRight is an integrated approach to address this IST vision and particularly the **trust** and **confidence** in building the future e/m-Europe, as well as to support innovative solution to complex problems in science, society, industry and business objectives. The social cohesion objective will also be one of the main focus areas of DigiRight since it will deal with the development of enhanced and less obtrusive communication tools. By developing innovative tools for protecting digital assets, DigiRight will contribute to a better social cohesion in Europe and beyond, by providing efficient, secure and private systems for communication, business, health, transport, risk management, environment, learning and cultural heritage. DigiRight will also contribute to the development of "codes of ethics" (guidelines) for best practice and legislative proposals which could support IST priorities, e.g., to ensure the co-evolution of technology and application, which are expected to become realizable through the research collaboration of an excellent research community like the one assembled in DigiRight.

Another issue that is considered essential for the DigiRight NoE is the participation of so-called "Small and Medium-sized Enterprises (SME)". The number of SMEs associated with the DigiRight network is expected to increase by time but at the beginning the SMEs represent 40% of the consortium's industrial partners. Their reaction time is incredibly faster when compared with large entities and provide an invaluable test bed for emerging applications.

DigiRight is also expected to address issues relevant to the eEurope 2005 Initiative: an information society for all plan and especially the development of a 'Virtual Campus for all students' that every EU member state should be able to offer to its students by the end of 2005. As a network connecting more than 16 universities, DigiRight is aware of the needs to connect students through an efficient network in order to maximize the quality and efficiency of the learning processes and activities. That is one of the main reasons why DigiRight is articulated around a strong virtual organizational network, a new paradigm for "Internet plus security and privacy". The idea is to use the results of research and development in DRM issues within DigiRight and incorporate them as soon as possible in the tools used for supporting communication within the NoE.

DRM aims to protect Intellectual Property Rights (IPR) over digital assets, and increases security, trust and privacy when information is exchanged over open networks throughout the entire value chain from producer to distributor to consumer, and potentially to consumer to consumer. Society in general and the IT in particular would thus greatly benefit from DRM.

DigiRight will be organized as a collaborative organizational network with various access levels (ranging from highly restricted access to unrestricted public access) in order to harness the management of research resources across Europe and bring the generated innovative results within the NoE to interested researchers in the DRM area. This is completely in line with IST priorities by supporting complex problem solving in science, society, industry and business. All work in DigiRight produces scientific results and publications. Various levels of IPR protections will be involved but in many cases, software will be produced with access for every researcher or citizen. The DigiRight NoE encourages the use and development of open standards and open source software in order to ensure interoperability of solutions and to foster integration and innovation. This will be mostly visible in the DigiRight software platforms to be accessed through the public side of the DigiRight web portal. It is expected that DigiRight will focus also on the exploitation of the acquired know-how within the NoE examining and evaluating various business models (e.g. creation of spin-offs, and co-operation with associated SMEs).

3. DigiRight: Relevance

DigiRight's integration goals address the scientific, technical, socio-economic and policy objectives of the EU IST Thematic Priority, "Towards a global dependability and security framework" [3].

In the near future, DRM will be an integral part of the end-user experience of consuming digital services. Thus, DRM will have a great influence not only on the **trust** and **confidence** relation among, and between, users and service providers, but also on any other stakeholder in content creation, distribution and management. The basic security and dependability challenges related to DRM are a major topic of research and integration between the different partners in the NoE. Therefore understanding user's concerns has a key role in the work of DigiRight. In addition, DigiRight will engage in standard setting operations,

which help to define a DRM architecture that meets these requirements.

Different security and privacy approaches will be taken into account in the NoE, integrating and sharing the experience of the different partners from the different application domains aiming at achieving a holistic framework for DRM. All domains relevant for the information society will be represented by the domain experts among the partners reflecting specific challenges, needs and solutions. Domains represented include e/m-Health, e/m-Commerce, e/m-Education, e/m-Entertainment and e/m-Government with different security objects and subjects. As said, DigiRight will attempt to address any stakeholder in any business chain including healthcare and governmental business with their specific stakeholders such as patients and citizens, but especially the domain professionals.

The different approaches of the research and experience of the partners in the NoE will be integrated, comprising architecture and technologies for security, virtual identity management, and privacy both at application and at infrastructure levels, always centered on the DRM framework.

Legal and regulatory, private and public policies, centered on the content policies will be investigated and collaboration between the partners will be done from the different DRM viewpoint aspects. Additionally, socio-economic issues will also be targeted in the research for innovative business models as well as in the content-related questions. Security and mobility should be achieved by way of integrating technical and regulatory instruments. Finally studying and integrating the end-user's experience of DRM enabled services is also vital as the user demand is usually driving the development.

Different interdisciplinary concepts need to be integrated under a common umbrella that will host all affected areas e.g., legal and regulatory, private and public policies, social, ethical and societal questions, business processes and models, and technology aspects.

Research and modeling will be done in all fields related to Digital Policy Management (DPM) as well to ensure the protection of user privacy, and enable users to control at a fine-grained level how and when personally identifying information is given to third parties and is exploited by them. The research and collaboration between partners will also be extended in the field of a specialized and improved risk management containing a risk management process, and risk analysis and assessment methodologies.

DigiRight aims at the development of some novel approaches that fully take advantage of existing standardized cryptographic approaches with regard to embedded security features and access to them, e.g. the ISO work on Privilege Management and Access Control [5]. Multimedia digital content needs embedded security mechanisms such as signatures, seals, timestamps, fingerprinting watermarks, etc., that will help enhancing the level of security for a wider spectrum of both existing and next generation applications.

DigiRight aims at concentrating the research and collaboration between the partners in all research fields related to DRM. Therefore, the development, testing and verification of technologies related to the protection, security and trust in the distribution of digital assets, combining the expertise and research of all the partners, is also one of the main objectives of the NoE.

Contribution to standardization activities is foreseen as a clear objective of the NoE, which is backed up by the great number of the partners that have long experience in the standardization process. The plan is to contribute to formal and informal cross industry standardization bodies at international level, as well as the European ones.

Biometrics is one of the major methods towards strong authentication systems. In DRM context, biometrics should play a key role with the consideration of social and operational issues that arise from such usage. Along with smart cards and other personal trusted devices, the NoE intends to investigate and develop mechanisms for the storage and processing of user's profile in a variety of heterogeneous devices (ranging from PCs and PDAs to mobile phones, smart cards and other tokens) that form a context of user's capabilities.

"Cyber-crime" could also be an issue from the risk management point of view. DRM focuses on the aspect of reducing the risk of illegal copying, viewing and processing of multimedia content. The use of digital seals and watermarks is one first step towards this direction. However, what is needed is a DRM framework that will integrate these technologies in a simple and efficient way and that will adapt to the requirements of each business model and use-case, and that will contribute to deterring cyber-crime.

As already mentioned, DRM is a key part of the future platform development that will ease the deployment of next generation applications and services. A DRM architecture that balances the interests of the various stakeholders will be a key enabler while an unbalanced architecture will only add to the existing hype and confusion. DigiRight intends to address the IPR of all digital assets, and especially the new opportunities and challenges that mobile technology offers to content providers, businesses and network operators for the development of value-added services and generation of new revenue streams.

Europe has a rich content base, technical strengths, long publishing tradition, and world ranking technology player but it is lagging in e-publishing and content-bound commerce. Therefore richer, multilingual digital assets that will kick-start a new mCommerce wave are needed. A DRM infrastructure that facilitates the entry of such assets to the market by balancing the needs of the various stakeholders, and that can fuel the

development of mobile applications, is clearly in a vital interest of the mobile Europe vision.

4. Europe's need to strengthen and reinforce S&T excellence on DRM

At present DRM in Europe has to operate and be administered across national borders, in a dynamic unpredictable heterogeneous environment, characterized by a lack of common standards at technical level, with several competing emerging technologies, uncertain business usage and cases, a plethora of different and possibly conflicting legal practices and regulatory frameworks, and an immense pressure from American content companies aiming at structuring the content business landscape according to their interests. The laws pertaining to the protection of intellectual property vary widely between countries, and are likely to remain different, despite both the European Commission's efforts to harmonize them [6] and other coordinated efforts. A general problem of European S&T is caused by the fact that, contrary to the U.S., there are national research policies operating only at national level. Currently, the EU does provide a very generic common umbrella, but most of the work is still done from a national perspective while a homogeneous approach is still a vision. This is especially true for high-level research activities. The former European Framework Programmes have tried to overcome this weakness by promoting pan-European collaboration via research and demonstration projects. Most of these projects, however, still suffered from strong national parts and rather weak "political" and "legal" interoperability in terms of a common (harmonized) European approach right from the beginning. This situation affects directly the management of rights in a digital environment, and poses challenging obstacles to overcome.

Today technical DRM solutions, models of human interaction, legislation and business models are produced by technology providers, social scientists, legislators and economists who co-exist without fully considering the side-effects on other domains. There is lack of communication between the domains, and practitioners are frequently totally ignorant of the S&T activities in another domain. We understand that maintaining the overview over different domains is an extremely challenging task that can only be done in a long time process with experts willing to accept new ideas and approaches. The non-existence of multiple domain overview by most of the researchers, often leads to production of exciting concepts and results that do not make the milestone to overcome the area that they were created for, and see if the same principle and expertise can be applied in a totally different context in another domain. The last leads not rarely into reinvention of the same things again and again, while the integration and applicability of knowledge of one domain to another has often led to true innovation, something valuable for the future. An everyday example is that of PKI. Banks established certain PKI-like services rather early. When eGovernment started to become effective, the involved partners – among them are both companies, public organizations, administration, and government ministries, etc. – tried to identify the needs and requirements of eGovernment to design and develop a stand-alone solution. Did they take the solutions of other sectors, like that of the banking industry, into consideration? No, they did not.

DigiRight aims at making easier the process of multi-domain communication, sharing and understanding of results and concepts, keeping researchers informed on cutting edge research and finally easing the task of keeping the overview that leads into production of innovative research.

What Europe really needs in terms of the existing pre-requisite to strengthen and to reinforce S&T excellence is an inter-domain, multidisciplinary approach that can be achieved by a specific restructuring of all existing research capacities and the way research is carried out. Technical problems and challenges in all major domains are often of nearly the same nature. Technical solutions from one can easily be adopted by, and adapted to the specific requirements of, other domains. It is a question of interacting and exchanging knowledge. Several companies in Europe are able to provide high-level solutions to their own customers, in their own application domain. This expertise and the results of research must be shared across the borders of sectors and domains if S&T including DRM is to be developed, enhanced, strengthened, and reinforced.

Although better co-operation between organizations involved in different aspects of systems for the management of rights in a digital environment will be very important, and the deployment of such systems seems, at the moment, important to create an international market especially for services based on material protected by copyright and related rights (especially videograms and phonograms, but also text documents, computer programs, interactive games etc), this is not only a question of technical development and the development of adequate legal instruments (legislative and contractual), but will for a large part have to be based on consensus. Such consensus presupposes an international and interdisciplinary dialogue, involving the research community, policy makers and – most importantly – the industries themselves. A Network of Excellence [7] would be a platform on which to stage such a dialogue, and would in itself at least contribute towards consensus.

DRM will boost content delivery networks (CDN) and with the infrastructure in place several business actors will be affected:

- Network operators can enhance their services and thus quickly convert the content service provisioning into profitable revenue streams. Time to market is reduced and interoperability is

promoted, thus lowering the threshold of market entry of new services.

- Content providers gain additional channels to sell their content and generate more revenue.
- Content distributors are able to offer new digital content services to their customers and implicitly make revenues by forcing the rights transaction.
- End-users gain richer content and transparent access to individual usage rights on high value content previously unavailable, without the need for proprietary flow specific plug-ins on the client-side.
- Enterprises have access to a variety of tools to create, monitor and control their assets.

Dormant content now residing unused in archives and closed, proprietary systems can be vitalized to create new markets and economic activity, thus promoting the dynamics of the economy.

5. DigiRight: Potential impact on restructuring and spreading excellence

DRM aspects play an important role in virtually any application domain. On the other hand, the related expertise and experience are closely related to a few domain experts. To a large extent, this specific knowledge never leaves the boundaries of national or application domains. What DigiRight aims at is the establishment of collaboration and information exchange between countries and domains.

5.1. Restructuring existing capacities and research methods

Drawing conclusions from what has been said earlier, in principle the S&T potential in Europe is in place already. There is no need for additional S&T work before the chances of adopting and adapting existing solutions are checked and verified. From that particular point of view - and regardless of the domain(s) concerned - the existing S&T capacities in Europe must be restructured in a way that allows enhanced information exchange, not only across the borders of countries, but also across the borders of domains and disciplines. The pre-requisite for doing so is that this expertise is available, and that the experts of the requesting domain are aware of the existence of the information. A network of experts in a certain technology must be established to solve the problems of many application domains by serving as "knowledge provider" for these application domains. This is the most important aspect of restructuring.

Similarly, there is a need for interdisciplinary research and co-operation if we are to address properly those issues relevant to the promotion of the information society. The organization of a network of excellence in DRM in different vertical and horizontal themes will contribute to the interdisciplinary understanding of the services required by the information society. Methodologies and tools used in one discipline might easily be adopted by another discipline. We thus need to integrate the methodologies and tools from technology, law, business, and social science (all of which are important operative factors in the uptake of DRM) to provide a common background and basis for combined research and an in-depth understanding of the fundamental issues and challenges of DRM systems, and to facilitate the exploitation of the synergy of the various projects, areas of expertise and stakeholders.

In sum, European S&T excellence on DRM would be much strengthened and reinforced by the integration of our existing fragmented research capacity so that research institutions can co-operate, set up liaisons, and share results.

5.2. How DigiRight will achieve this restructuring

DigiRight will guide towards restructuring of research in each organization, in order to transform from a group-based closed community approach to an open and transparent peer-to-peer approach where interaction and integration will be the driving force. The research group of an organization will be in contact with other groups involved in research in other disciplines related to DRM, exchange ideas and knowledge, and engage in joint research activities so that each group will have a holistic view of DRM. This combined expertise of the various disciplines will lead to an integration of DRM usage in all domains. Joint research activities are highly beneficial to the excellence of EU research potential. The spreading of excellence and establishment of a high-competence open group of researchers is expected to maximize the scientific outcomes and completely restructure researcher's time and organization's resource management. The multicultural constitution of the network is also expected to open new perspectives and lead to the development of best-practice approaches for the resolution of long-lasting organizational problems.

This alternative to the traditional fragmentation of our continent and research will lead to useful comparative studies which will identify the best solutions, allow for the greater exchange of views and ideas, and lead to the development of interoperable standard solutions. The network will certainly be able to achieve the goal of information exchange between different application domains by inviting experts of these domains to join the NoE. Interdisciplinary working groups with technicians, researchers, and application domain specialists (e.g. from the eHealth domain, the knowledge provider domain, the multimedia provider domain, etc.) can meet these aforementioned requirements by providing a solution that is sound, that fits and that is really applicable to the

domains in question. This solution will be based on the combined expertise of several independent domains and will thus be more generic and reliable than a solution that is solely based on domain internal knowledge of one single domain.

Because of the highly multidisciplinary nature of the DigiRight NoE, one of the tasks will be to create Special Interest Groups (SIGs), working on specific inter-related tasks. Partners involved in those SIGs will contribute to common goals and will start to establish closer links between their respective organizations. The communication tool of choice will be the DRMnet a Virtual Private Network (VPN) that will be created for this purpose. Other activities will also contribute to the fusion of the separate entities, such as meetings, workshops, summer schools, conferences, and Ph.D. 'twinning' (pairs of complementary Ph.D. on related topics from two different organizations). DRMnet will be used as a 'classical' VPN but will also connect the 'Usability Test Rooms' from the various institutions. This will allow the sharing and remote testing of various research works: shared test datasets and the common DigiRight platforms will allow remote communication, not only to help researchers in their work, but also to test new software tools either remotely or locally at the various locations in the network.

The DigiRight network is federating research centers already excellent in the domains of its interest i.e. security, privacy, trust, protection mechanisms, etc., but this excellence of laboratories is seldom used for providing holistic solutions to the DRM. The critical mass of researchers in DigiRight will make possible that kind of integration, and long term goals like the 'Secure Internet' and other grand challenges will begin to seem more attainable. The DigiRight community already has tight links with many research institutions in the security and trust fields beyond those already involved in the network, like the European Symposium on Research in Computer Security (ESORICS) (www.laas.fr/~esorics), and iTrust (www.itrust.uoc.gr), but in order to continuously reinforce links with the DRM community at large, a provision of some of the DigiRight budget would be reserved for actions like creation of a joint SIG or the joint organization of a conference. Here we consider the multimedia content providers in general and especially the music, film and art industry with their main players (among them are art directors, photographers, performers (singers, actors), composers, song writers, book authors etc.) important.

Research in DRM related issues is already funded in all partners' labs through various regional, national and international funding sources, but almost all this funding is related to short or mid-term goals and standalone project specific solutions without any integration roadmap with concurrent efforts. Within the DigiRight all activities have the long-term goal of integrating the research institutions in a large and tightly knit web. The integration of industrial partners and more specifically SMEs may provide a source of self-financing for the networking activities themselves, which will guarantee the long lasting integration of European research in the DRM field.

5.3. Continuous structuring impact on European research

DigiRight has a number of partners who are commercial enterprises, both SMEs and larger concerns. They will exploit the results of our research by implementing them in innovative products and services for several application domains. Based on the future achievements of the NoE, we intend to provide web-based technical support for users, and on a pre-agreement basis technical and scientific support (reports (e.g. state-of-the-art and case studies), design of subsystems to be integrated in their approaches, for industries and SMEs. All these activities will create revenue, and we foresee that the Network will begin to be self-sustaining after about five years from its kick-off. Being financially independent, the network will be a viable and self-perpetuating entity, that will become a permanent feature of the European research landscape, thus, by its on-going activities and the example it sets, having a durable structuring effect on European research. DigiRight intends to create a permanent virtual research organization, annual events such as conferences, workshops, and summer courses for doctoral students.

The effect of influencing the European S&T society will be a long-lasting one. The trend is towards projects of longer duration. Contrary to the Research and Technology Development (RTD) projects about 10 years ago, nowadays S&T activities focus on 10 to 15 horizon. With regard to multimedia electronic health records one can consider that today's developments will result in prototype applications in about 5 or 6 years with a general acceptance in about 10 to 12 years [8]. From that perspective, project efforts and therefore also project advisory activities and project steering efforts will last. In addition to that, some of the involved experts may decide to found their own advisory and promotion companies (spin-off companies) managed by the Steering Committee after the funding of the EC has come to an end. Several former 4th and 5th framework projects (e.g. RICHE [9]) have shown that this strategy is indeed feasible. The DigiRight NoE could easily kick-off a re-structuring of specific parts of the European S&T community from "pure" development-oriented strategies to a more business-oriented approach having a lasting effect on European research.

Nowadays the state of the art in Europe (both in DRM and related application domains) needs to be discovered by exploring the technology of DRM (done by technology experts), the current use in application domains (domain experts like physicians and health managers) and the related development and improvement strategy for the next few years to come (life time of the project) as well as for the time after.

The most important aim of DigiRight is in fact that experts who "normally" would have never met because of their "domain-and-discipline-restricted" thinking and research policy, are now brought together to discover the similarities in their problems and seek generic innovative solutions that may or may not be available in other domains. These experts can easily exchange knowledge with both their own "technical domain" colleagues - that's what they do anyway - as well as with colleagues from "application domains" such as health, media, information providers, and art providers (that's what they normally don't do). From that particular point of view, this NoE is able to restructure the current research environment in Europe and beyond. The second aspect is this enhanced knowledge that can only be compiled using this cross-domain approach. And combining both approaches EU's target of "the network will begin to be self-supporting after five years" can become a reality.

By concentrating on the tree (stand-alone solutions for isolated domains) we miss the forest (global generic concepts applicable to diverse domains). Experts on DRM can enhance existing solutions and develop new ones, but what we (the European S&T community) all need is to make this knowledge available, couple it with application knowledge, enhance both interoperability and standardization, and look for real business cases. These business cases can easily be found in virtually any application domain but need to be identified in a close co-operation and collaboration between different specialties. It is not only the regional and national boundaries that count but also the borders of domains. Ask experts from different domains about their problems to be solved, and they are going to tell you the same things. Ask them whether or not they have checked other domains' results, and they will tell you: no, not yet. We do not know where, how to look for these results, and we do not know the experts to be asked. That is the bitter reality.

Finally, DRM is a very suitable technological area for showing the important potential of these networks of experts from different technical and application domains. Any kind of multimedia content to be protected is potentially interesting for DRM experts. Regardless whether it's a video or audio file that is sold to a customer and should not be copied/used without permission, or a set of medical high resolution images that need to be protected against unauthorized changes, DRM is able to provide an applicable solution.

5.4. Spreading excellence beyond DigiRight, disseminating knowledge and exploiting results

The Network will be an important contributory factor in the creation of a common European Research Area, both in our own field, DRM, and in a number of satellite fields, thus, we expect it to lead towards a permanent and stable integration of the research community on the one hand, and, on the other, the dissemination of expertise to those who need it. The last can be achieved in conjunction with the planned international workshops and Europe-wide courses in co-operation with industry. More specifically, the dissemination mechanisms will be as follows.

The various conferences and workshops to be arranged by the NoE and their proceedings will be the main vehicle, and will be supplemented by liaisons with industry, scientific and commercial consortia, and standardization bodies, and will become a major mean of establishing contact with other application domains outside the project that have certain DRM requirements. Other aspects will be the publication of a DigiRight newsletter, journal publications, contributions to prestigious conferences, development of brochures for public awareness, kick-start of a European Forum on related technologies, and a web repository. DigiRight is going to contribute and promote its results to standardization bodies, and will effectively support the continuous exchange of ideas between universities, research centers and industry. This is achievable since the consortium is consisting of members coming from all these domains and initial contacts are already established. Last but not least, the NoE aims at the presentation of DRM technology and products at certain industrial fairs.

Experts within the network will make their experiences and collected knowledge available to the outside world. Especially the aspect of collecting knowledge is very interesting for other projects, initiatives, and activities. After having found a certain level of excellence among the experts, the network can easily become a partner providing services to other technical domains. In the context of other national or international projects, the network could act as an advisor or even supervisor securing a reasonable use of the funding by providing working solutions or services. This is a kind of re-use of existing information. The network is considered a multiplying factor for application domain dependent excellence (e.g. health, culture, and art) and application domain independent knowledge (e.g., DRM aspects).

DigiRight will push the results and findings to different consortia of regional, national, or international projects. In addition to the typical exploitation and dissemination mechanisms of the S&T community (workshops, congresses, conferences, tutorials, sessions, training courses, and fairs both technology-related and domain-related), the role of advisors, steering board members, mentors, etc. in the aforementioned activities will become important exploitation strategies that are directly linked and connected to the specific approach of the Network.

5.5. Contributions to standards

As mentioned earlier, standardization and the strict usage of standards and pre-standards is an important aim of DigiRight. Contributions aimed at the trust,

policy, privacy, security and risk management for DRM will be submitted to some of the leading standardization bodies, organizations or fora - like CEN/ISSS, ISO/IEC, OASIS, and OMA - in different fields like metadata and its interpretation, MPEG4, MPEG7, MPEG21, REL, RDD, and IPMP.

A joint CEN/ETSI Group on Network and Information security Standardization (NIS) was established in 2002, with the aim of addressing the issues raised in Communication COM (2001) 298 by the European Commission on "Network and Information Security: Proposal for a European Policy Approach". DigiRight will carefully follow the CEN/ISSS standardization on DRM issues.

With regard to the Open Mobile Alliance (OMA) we need to have a look at the new and updated MCOMM workgroups of ETSI and the network layer issues on DRM where ETSI has the lead. DRM standards within media/content industry as well as telecom industry are relevant.

6. Conclusions

In this paper we have described DigiRight's relevance to and impact on the Sixth Framework Programme and Europe's need to strengthen and reinforce the Science and Technology excellence on DRM. DigiRight is a Network of Excellence for a Framework for Policy, Privacy, Security, Trust and Risk Management for DRM that will consist of teams of experts on technology and development, business, law, ethics societal questions who will organize on-going and future high quality research on those areas of their respective disciplines that apply to DRM.

Conducting research for the purpose of extending vertically the five horizontal fields, viz security, privacy, policy, trust and risk management, will result in the provision of specialized services for the protection of IPR. These services will need to be updated, improved and enhanced as time goes on which will require further research and development. This will contribute to the durability of research beyond the termination of the project. Additionally, the established communication links within and outside the Network will accommodate research and make it much more cost-effective relative to the benefit of an enhanced service, again contributing to its durability.

DigiRight intends to create an industrial board and to increase industrial participation in the consortium during the first five years. Through this liaison, industrial partners will provide guiding input to, and receive benefits from, the research being conducted by the research partners in DigiRight. It is also expected that the Network's day-to-day working activities will be partly integrated in the day-to-day management of the participating institutions, leading to durability of the co-operation between the partners.

Acknowledgments

The authors would like to acknowledge all DigiRight partners [10] for their contributions and efforts during the proposal writing in general, and especially Marc Fetscherin, Nineta Polemi, Thomas Dreier, and Jon Bing for reviewing that part of the proposal relevant to this article.

References

[1] H. Abie, J. Bing, B. Blobel, J. Delgado, B. Foyn, S. Karnouskos, P. Pharow, O. Pitkänen, and D. Tzovaras, DigiRight: Network of Excellence for a Framework for Policy, Privacy, Trust and Risk Management for Digital Rights Management, International Mobile IPR/DRM , HIIT, Finland, August 27-28, 2003

[2] R. Iannella, Digital Rights Management (DRM) Architectures, D-Lib magazine, June 2001, Vol. 7, No 6, http://www.dlib.org/dlib/june01/iannella/06iannella.html

[3] 2003-2004 Work Programme, Information Society technologies, EC IST priority, WP 2003-2004

[4] eEurope 2002 Initiative: eEurope 2002 - An Information Society For All. http://www.eeurope-standards.org/

[5] B. Blobel and R. Nordberg, Privilege Management and Access Control, in Shared Care IS and EHR. In: The New Navigators: From Professionals to Patients, Proceedings of MIE 2003 May 4-7, 2003, Amsterdam: IOS Press. Studies in Health Technology and Informatics; 95. ISSN/ISBN: 1-58603-347-6 0926-9630.

[6] Commission of European Communities 2002, Commission Staff Working Paper: Digital Rights, Background Systems, assessment, SEC (2002) 197, Brussels 14/02-2002 http://www.politechbot.com/docs/european.commission.drm.030202.pdf

[7] Research DG, European Commission, Provisions for Implementing Network of Excellence, Background document, Draft 2002 edition: 11 November 2002

[8] CEN TC 251 TF HISA: Revision of ENV 12967: Health informatics - Service architecture; TF EHRcom: Revision of ENV 13606: Electronic Health Record Communication. http://www.centc251.org/

[9] The RICHE Consortium: A framework for open information and communication systems for health care in Europe. http://www.newcastle.research.ec.org/esp-syn/text/2221.html

[10] Network of Excellence for a Research Framework for Privacy, Policy, Security, Trust and Risk Management for Digital Rights Management, a proposal for network of excellence under FP6 for the IST Call 1, submitted to EC, 24/04-2003, http://digiright.nr.no/nuke/html/

Authors Index

Helsinki Institute for Information Technology HIIT, founded in 1999, is a joint research institute of University of Helsinki and Helsinki University of Technology.

HIIT conducts internationally high-level strategic research in information technology and related multi-disciplinary topics, especially in areas where Finnish IT industry has or may reach a significant global role. HIIT works in close co-operation with Finnish universities, research institutes, and industry, aiming to improve the contents, visibility, and impact of Finnish IT research to benefit the competitiveness and progress of the Finnish information society. HIIT also aims at creating a strong network of international partnerships with leading foreign research universities and institutions.

CONTACTS
WWW: http://www.hiit.fi/
Tel: +358-9-85012313; Fax: +358-9-6949768
Postal address: P.O.Box 9800, 02015 HUT, Finland
Visiting: High Tech Center, Tammasaarenkatu 3, Helsinki, Finland

ISBN 951-22-6675-X (printed)
ISBN 951-22-6676-8 (electronic)
ISSN 1458-9451 (printed)
ISSN 1458-946X (electronic)

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY