Trading digital intangible goods: the rules of the game

Konstantas, Dimitri; Morin, Jean-Henry

Trading digital intangible goods : the rules of the game

Dimitri Konstantas
Jean-Henry Morin

Abstract

In this paper we first specify the terms and conditions for the commercialization of intangible goods, and namely digital documents, defining a mapping between the commercialization terms of the tangible good (printed information) to the commercialization terms of the intangible good (digital document). The meaning and limits of purchasing, accessing, redistributing, copying, preserving user anonymity, etc. are defined in accordance to the well established rules governing the trade of printed information. We next specify the requirements that a digital commercialization system of intangible goods should fulfill in order to guarantee the rights of its users (providers and consumers) like user anonymity, superdistribution, marketing policies, etc. Finally we give a brief presentation of a framework and a pilot application that we designed and implemented for the commercialization of digital documents.

1 Trading of Goods

The commercialization of tangible goods, ranging from cars to books, is a long established trade with well defined and understood terms and conditions. The rights of both the seller and the buyer are clearly defined either formally by laws and contracts or informally by what is called "common practice". Also a certain number of intangible goods, like copyright of an intellectual work or the commercialization rights of a certain product for a certain region, are traded as if they were tangible goods.

With the wide use of computers and networks a new class of intangible goods appeared in the market: the digital intangible goods, which in most cases are the digital representations of tangible goods, like books and documents but also work tools like document editors and business applications. This new class of intangible goods are considered by both buyers and consumers as just different representation of tangible goods. Thus, it is expected that their commercialization will be done under similar terms as tangible goods. However due to their different nature (i.e. atoms versus bits) the *similarity* of the commercialization terms and conditions is not always an obvious one.

Another aspect in the commercialization of tangible goods is the distribution channel. That is, the underlying infrastructure and commercialization network used for the dissemination of the goods. The rules regulating the distribution channel are also well defined and each intermediary in the distribution chain is bound by law or custom to a certain behavior. With digital intangible goods however the distribution infrastructure is the commercialization computer application and the network. This infrastructure should satisfy certain requirements for the protection of the rights of both the provider and the consumer (among the major, copyright and privacy) acting as an intermediary in the delivery of the (intangible) goods. For example it should protect

the user's anonymity, allow for marketing policies, the free choice of provider etc. Although these requirements stem directly from the tangible goods commercialization model, in the case of digital intangible goods' trading they need to be explicitly stated, since they serve as the basis for the development of digital intangible goods commercialization systems.

In this paper we take the example of digital documents as a target intangible good and define the commercialization terms and conditions for digital information trading. We then specify the requirements that a digital commercialization system of intangible goods should fulfill in order to guarantee the rights of its users (providers and consumers). Finally we give a brief presentation of a framework and a pilot application that was designed and implemented based on the above identified rules, for the commercialization of digital information.

Note that although in this paper we talk about "digital documents", the ideas developed are equally valid for a large class of digital intagible goods, ranging from video and audio streams to books and news articles. In this sense the term document should be seen as a generic term encapsulating almost any digitally representable information or tool.

2 Commercialization of Digital Documents

The commercialization of printed documents, like books, newspapers and reports, is a long established trade with well defined and understood terms and conditions. People are used of trading material documents in different forms. For example, a book can be sold as a pocket book or in parts, when published within a journal, or even as an audio tape. With the introduction of digital technologies digital documents were initially considered as just one more representation of a material document, and were traded in the form of diskettes and CD-ROMs. However the evolution of technology and the introduction of the Internet allowed people to trade digital documents without the need of a physical support. As a result the long established rules and conditions of document trading were no longer applicable. Practical restrictions and problems that in the past were acting as a regulator in the protection of intellectual rights were no longer existing. For example, a reproduction of a printed book costs as much and in most cases more than the original and people prefer to buy a new original rather than photo-copying it. A digital book however can be copied and distributed over the Internet indefinite times at virtually no cost with no quality loss (i.e., no notion of an original).

The trading of intangible goods is governed by its own terms and conditions [19]. However, if we wish to promote trading of intangible goods we must define the applicable rules and conditions in a way that is understandable and acceptable by both the publishers and the readers. For that we must start from the existing model of trading tangible goods and define the relations with the trading of intangible goods. Our claim is that intangible goods can be traded under *similar terms* as tangible goods. However we need to define what we consider as "similar terms".

Document content advertising. The first action of a reader is to identify if the document he is about to purchase interests him. With printed document, like journals and newspapers, the reader finds in the front page the titles and possibly a few lines of abstract describing the contents of the document. This information is provided free of charge (newspapers and journals are posted outside kiosks in order to raise the interest of the readers). With digital information the reader

should be able to obtain a brief summary of the content of the document without having to pay for it. The summary depends, known also as *value stripped content*, on the policy of the provider and can range from a simple title to a full abstract.

Document purchasing. The purchasing of a printed document and the payment of the corresponding fees is done at the moment the reader requests the document. It is at this moment that the reader expresses his will to read the printed document and consequently pays the corresponding fees. Thus, with a digital document the payment of the corresponding fees should be done at the moment the reader expresses his interest to read it. That is, when the reader attempts to open the document for reading. It is at this moment that payment of the corresponding fees should be made according to the policies attached to the document.

Document reading. A person who purchased a magazine or book expects to be able to read it as many times as he wishes without having to pay again every time he wishes to re-read it. With a digital document, where document purchasing is done at the moment that reader attempts to read the document, the reader should also pay once and be able to read it as many times as he wishes without having to pay for it again (provided that the attached policy allows it). Furthermore even if he possesses the digital document (that is, the digital data) he should not be able to read it without first paying for it.

Document re-distribution. It is quite common that a person passes a document he purchased to a friend. However this action results in the original purchaser loosing ownership of the document or at least of the right of usage (i.e., reading). If he wishes to read it again he has to buy a new one or retrieve the borrowed copy. Alternatively the owner of the document can give an indication of where the document can be acquired and paid for, keeping his own copy. What is important to note in this transaction is that we always have a single copy of the document which can be read at any given instant by one and only person. With digital documents on the other hand the case is different. When one passes a digital document to a friend he actually makes a (indistinguishable) copy of the original. As a result both persons have now a copy of the document. However considering the previous term (document purchasing) the second person should not be able to read, without paying, the copy of the digital document he received, unless the original owner looses his right to read his copy of the document.

Document life time. A reader buying a printed document today and preserving it in good condition is expecting to be able to (re-) read it after long time periods (in the order of decades or even centuries) without having to pay again for it. This should also be true for digital documents. A digital document which the reader bought today (that is, for which he paid the fees for reading it) should be readable free of charge after long time periods.

Document copying. A common and (up to a certain level) tolerated practice with printed documents is photocopying. Photocopying is tolerated by the publishers for a number of reasons: first of all the quality of the copy is (in general) lower than this of the original; second in many cases, like for example for books, photocopying the complete document is more expensive than buying a new copy; third, photocipies can be made within the established context of "fair use"; finally photocopies are easily identifiable and, if needed, legal action can be taken against the malefactor. Another way to copy a printed document is through *Optical Character Recognition (OCR)*

systems. However the reproduction of a printed document using OCR is in most cases time consuming and costly. With digital documents photocopying can be compared with the printing of the computer screen, something which is, as with photocopies, difficult to prevent[1]. In addition, as with printed documents, one can consider reproducing the digital document from the captured screen dump using OCR techniques. However in both cases the quality of the document is lost and in addition any special features of the digital document, like for example hypertext links, active parts (code, animation, sounds etc) will disappear.

Document purchaser identification. A major issue in the commercialization of printed documents is the ability of the reader to buy the document without revealing his identity to anyone. One can buy, for example, any magazine, newspaper or book from a kiosk or book-store keeping his full anonymity from both the sales person and the publisher. On the other hand a reader can decide to reveal his identity to a publisher or reseller agent via, for example, a subscription and benefit from any possible special offers, like discounts, extra editions, advance copies etc. Note that the fact that a person is reading a specific document is in itself information. Thus it should be up to the reader to decide if he wishes to reveal this information or not. With digital documents the reader should also be able to read a document without having to reveal his identity. The document provider should not be able to relate the collected document fees to a specific reader. Of course if the digital document provider offers nominative subscriptions with possible side benefits, it should be up to the reader to decide if he wishes to subscribe and thus reveal his identity, or if he prefers reading the digital document anonymously.

Document authoritativeness. The cornerstone of a printed document is the indisputable identification of the source of the information. The reader of a printed document knows with certainty who created the specific document and can easily identify modifications done on it, like corrections or additions. It is in general very difficult for one to modify or falsify a printed document in an untraceable way. Nevertheless given enough money, time, effort or power any printed document can be untraceably modified or falsified. For example someone with enough money can very easily print a false edition of a newspaper which is indistinguishable from the original. With digital documents the reader should thus be able to indisputably identify the source of the document and verify its integrity. However, as with printed documents, a person or organization with enough money, time or power will always be able to falsify any digital document.

3 Requirements for a Commercialization System of Digital Documents

Once the terms and conditions of the commercialization of digital documents have been defined, we need to translate them to requirements that will serve in the design of the commercialization application. These requirements will reflect the fundamental interests of the digital information publisher and consumer, namely the fact that the publisher is interested in providing a profitable service fulfilling the needs of the consumer, while the consumer is interested in obtaining a re-

1. Of course there are techniques that prevent one from making photocopies, like for example the use of special ink, but these are not so often used due to their high cost.

liable service for the right price. Here we give an overview of the basic requirements; an extended description can be found in [7][11].

Anonymity. As with traditional commerce an information consumer does not need to reveal his identity to the publisher of a magazine or newspaper in order to buy it, so with digital information the consumer should be able to buy information without having to reveal his identity in any direct or indirect way.

Information granularity. In the traditional publishing industry, the smallest information unit that can be put on the market is the issue, which bundles a substantial number of information pieces and its price is fixed accordingly. In a digital environment however, the granularity of the marketable information unit can be brought down to the level, for example, of a single newspaper article. In fact what the digital information consumer will be interested in is buying independent pieces of information and not complete editions.

Superdistribution. It is quite common for a person to read something and to wish to show it to somebody else. With printed material this is easily done by simply cutting the item or giving the complete edition to another person. In a digital system however this process although feasible, has important copyright violation side effects. The reason is that instead of passing to the other person the specific item, and in consequence losing possession of it, we actually make a copy of the item. This copying can violate copyright law. Thus the digital system should allow the reader to freely pass information items to other persons without violating copyright law.

Subsequent access. Once the consumer has paid for an information item he should be able to read it again at a later time without having to pay for it a second time. Although this might look like an obvious possibility it is an important requirement since payment of author rights is done at the moment of reading the document. Subsequent reading of the information by the same reader should not result in a second payment of the author rights (unless explicitly expressed by the policy attached to the document).

Free Choice of Providers. In a digital commerce environment the consumer should be able to choose freely from where he buys services and goods. This means that any system installed should not be bound to a specific provider, allowing the consumer to freely choose the provider from whom he will buy information.

Off-line Activity. It is a common practice for a person to buy a magazine or newspaper and read it at different locations, like when traveling or even at the beach. A digital information system should allow the reader to read, and consequently pay for, information items that are stored locally, even in the absence of a network.

Notification of Update Availability. Being informed "on-time" is a major issue for the information consumer. In a digital information dissemination system means are needed to offer the information consumer the possibility of being notified immediately upon availability of information updates on desired issues.

Information selection. Different information providers have different specializations and present information in different ways (classification). In a digital information world we will have a large number of digital information providers available. The information consumer

should thus be able to define which kind of information he wishes to obtain from each digital information provider.

User Interface. Since it is neither feasible nor desirable to create a new hypertext browser, the digital information system should be able to run within any widely available browser, like a Java enabled Web browsers.

Marketing policies. Of major importance in the success of a service is the choice of the right marketing policy. The information providers should be able to implement flexible and adaptable payment policies. For example the price of an article can change depending on its publication date (last week news have in general no value). This should be feasible without any need for the reader to interact with the publisher: the article itself should be able to figure out its current price.

Information Access and Information Evolution. The information consumer should be able to easily access older information and trace the evolution of events. This means that published information should be immutable and identifiable.

The Information Consumer as an Information Provider. To illustrate this requirement we consider the following two examples where a user receives an information item: (i) the user decides to forward it to a friend together with some comments. (ii) the user decides to forward it to a client together with some comments for which a fee must be paid. In both cases a new information item is created, which contains the original information and the comments with a possible corresponding price. The idea here is that an information consumer can become an information provider of his own added value and a reseller of other information provider's material without infringing any copyright or intellectual property law. Thus information consumers should be able to publish new information items embedding information of other providers together with their own added value information. The final reader will have to pay both providers in order to read the information and the attached comments.

4 The MEDIA Approach for the Commercialization of Digital Documents

The aim of the *MEDIA (Mobile Electronic Documents with Interacting Agents)* [6] project is to develop the means that will allow protection, commercialization and dissemination of digital documents under similar conditions as those for printed documents, as defined above. The MEDIA approach is based on the encapsulation [1][3][4][5] of the documents in agents. The document is no longer a simple collection of data but a program which the reader must execute in order to be able to read it. The document agent can thus enforce the copyright control and payment at the time the reader requests to read the document.

In the context of the MEDIA project we designed and developed the *HEP (Hypermedia Electronic Publishing)*[7][8][9][10][11] framework that implements the MEDIA digital document commercialization model. The Hep framework enforces a *pay per use* scheme for the digital documents. The reader pays only for what he requests explicitly to read and he cannot read a document which he has not payed for. The document distribution model of the Hep framework is based on public key encryption with a security schema that discourages infringements [12].

Information consumer anonymity and privacy are protected so that the reader need not reveal his identity to the document provider. Furthermore the Hep framework supports off-line operations for the payment of the document fees and the release of its contents to the reader, with the simultaneous delivery of receipts (proof of purchase) for subsequent accesses to the document. An overview of the MEDIA approach and system can be found in [13].

5 Security issues

A very important issue in the conception and design of the Hep system, as well as in any system with similar goals, is the definition of the security concept. It is a well known fact that today absolute security via software encryption can not be achieved. The best we can do is to asymptotically approach the ideal absolute security accepting a certain risk level. How close we reach and what is an acceptable risk depends on the specific application.

In the Hep system the core issue is the protection of the content encapsulated in the document agent. The target is to protect the owner of the content from malicious users that will try to extract the content of the documents by breaking the security schema of the agent [2].

5.1 Content Encryption

The core of the Hep system security relies on the encryption of the document contents. The idea is that the user should not be able to break the encryption using reasonable processing power and within a reasonable time frame. What is reasonable, of course, depends on the importance of the information contained in the document. If the information contained has a commercial value for a short time period, like for example a quotation of the stock market, then the security level employed should be sufficient for preventing the user from breaking it within the commercial life time of the information. The major problem however comes from contents with very long life time, like for example books which have a life time of more than 30 years. In this case it is difficult to anticipate an encryption schema that will retain its properties for this long period. For example, ten years ago the technological state of the art was defining that breaking a specific encryption schema would have required more than 50 years of processing time. Today the same encryption schema can be broken in less than 0.5 years of processing time (or even less). In a few years, with the projected evolution of technology the same encryption schema might be breakable with a few hours of processing on a home computer. Therefore we cannot expect that an encryption schema used today will retain its security level after a few years.

Systems like Hep, based on the principle of content encapsulation and encryption, can provide adequate protection of intellectual rights only for the time periods for which the technology evolution can be anticipated. It would be unrealistic to try to provide long lasting protection, i.e. 20 years or more, employing only content encapsulation and encryption.

Nevertheless the Hep model offers a customized encryption model. The provider can use any encryption algorithm he wishes, create new ones and choose the encryption key length to any size he believes will be suitable for a specific document. This way the encryption security of the Hep platform can be adapted as needed according to the technology evolution and needs of the providers.

5.2 Security Target of Hep

If we consider for a moment that the offered protection of the used encryption schema is sufficient, then the question to be asked is how secure is the overall Hep system. That is, how difficult can it be for a user to maneuver the agent document in releasing its contents without actually having the proper authorization. This can be done either by tampering with the agent or by compromising the agent platform.

If we assume that the agent platform is secure (the user has/can not tamper with it) then the attack on the document agent can be easily prevented. The agent can include a signature which will allow the verification of its integrity by the platform. If the agent has been tampered the platform will simply reject it and will not handle it. Certificates can be used, for example, to compose this signature so that the user will not be able to fake it. In addition the document-agent provider can use code obfuscation [20] techniques in order to make the task of tampering with the agent code more difficult. This way the average user will not be able, without considerable effort, to modify the agent behavior and in addition it will be difficult to automate the task since the code implementing the agents will be different from one agent to another.

The second type of attack will be to tamper or fake the agent platform itself. In this case the arriving document agent will not be able to verify whether or not the platform has been tampered. The fake platform can easily provide all required replies to hide itself. One might consider using secure external devices, like a smart-card or a crypto-card [17], that will allow the agent to communicate with and verify the platform. However in order for the agent to establish a secure channel with the secure device it will have to execute (run) on top of the agent platform which, by definition, is compromised! In other words the agent can never be sure that the platform has not been compromised. On the other hand the secure external device might be able to detect a compromised platform, in which case it can refuse to collaborate, or if needed self-erase its contents. Nevertheless, the secure external device will be able to verify only that the software platform has not be compromised. The complete system can very well run on top of a CPU emulator system allowing capture of all relevant data at the CPU level.

5.3 Calculated Risks

It is clear that absolute security can not be achieved; there will always be a certain risk involved. In a commercial world however it is quite common to operate and make business accepting different well known security breach risks. This is what it is commonly known as *calculated risks*. That is, risks we are well aware of, but which we accept in order to achieve our goal.

In the design and implementation of the Hep platform we have accepted a certain level of security risk. That is, we assume that the Hep platform is *not* secure, but a substantial effort is needed in order to break the security. What we target is to eliminate generic security failures that will allow the malicious user to access the content of documents in a simple, fast and automated way.

The most sensitive part in the Hep platform is the financial institution's private key. A user gaining access to this key can access all documents from all providers. Thus, in our design the financial institute's private key is never communicated. A special case however is the off-line

operation where a smart card is used for the payment and document key release. In this case the smart card plays the role of the financial institution, holding the financial institution's private key and decrypting the document keys. All the key decryption operation is performed internally in the smart card and the user never gains access to the financial institution's private key. This way the user has only the possibility to extract independent document keys. However extracting the document key gives him access to only one document. By designing the Hep security system in a way making the extraction of the document key very difficult, requiring long time periods of manual work, we can say that we have an acceptable solution with a well known calculated risk.

What we must ensure is that the user does not have a way to come up with a generic way to extract document keys. For example, if the user knows (after analyzing the Hep system) that the key is always stored at a certain position in the memory, then he can easily write a program to dump this memory position. Possible solutions to this problem might include the allocation of a large memory area where the keys are stored at a random position or even fragmenting them and storing them in pieces in different positions.

Another calculated risk we take in the Hep approach is the fact that a user might be able to program a fake platform imitating the functionality of Hep. However, we can assume that the effort to write this program will be great enough to discourage such approach.

6 Conclusion

The rapid expansion of the Internet has boosted the trading of digital intangible goods. However the rules and conditions of this trade are not yet well understood. The existing models used for the trading of tangible goods need to be adapted to the new digital trading world. In this paper we presented how the trading of intangible goods, and specifically digital documents, can be viewed as similar to the trading of tangible goods, by defining the notions of similarity. Furthermore we presented how the requirements of this view are translated to functional requirements for the design of a commercialization platform. Based on these requirements we designed and implemented Hep, a platform aiming at the commercialization of digital documents under similar conditions as printed document.

One of the major issues in the commercialization of digital documents is the protection of the intellectual rights of the document owner. We do not claim that the Hep platform provides 100% security of the owner's intellectual rights, but that the effort to break it exceeds the available resources of the user. Someone with enough power will always be able to break the security. The real question is not how to achieve absolute security, but where do we stop; that is, at which level of difficulty for breaking the security schema. If we accept that an effort of 12 months, for example, is acceptable for breaking the security of Hep then we can design the document agents and platform accordingly. If we also forsee that future PCs will have crypto chips and possibly copyright chips incorporated at the hardware level, then we can improve the security protection by incorporating their functionality in the Hep platform.

Nevertheless as the protection of intellectual rights cannot be achieved only with technological means, in the same way it cannot be also achieved with strict regulatory means. Technol-

ogy and legislation (law and public policy) should be combined to provide a consistent environment protecting the efficiently the trade of intangible goods.

References

[1] Kaplan M.A., "IBM Cryptolopes™, SuperDistribution and Digital Rights Management", IBM Corporation, December 1996, http://www.research.ibm.com/people/k/kaplan

[2] Kohl U., Lotspiech J. and Kaplan A., 1997, "Safeguarding Digital Library Contents and Users Protecting Documents Rather Than Channels", IBM Research Division San Jose, California, and Hawthorne, New York, *D-Lib Magazine*, September 1997, http://www.dlib.org/dlib/september97/ibm/09lotspiech.html

[3] O. Sibert, D. Bernstein and D. Van Wie, "The DigiBox: A Self-Protecting Container for Information Commerce", proceedings of *First USENIX Workshop on Electronic Commerce*, New-York, July 11-12, 1995.

[4] Softlock Services Inc., *SoftLock*, http://www.softlock.com/

[5] Open Market Inc., *Folio 4*, http://www.folio.com/

[6] Dimitri Konstantas, Jean-Henry Morin and Jan Vitek, "MEDIA : A Platform for the Commercialization of Electronic Documents", in *Object Applications*, Ed. D. Tsichritzis, CUI, University of Geneva, August 1996.

[7] Jean-Henry Morin and Dimitri Konstantas, "*Towards Hypermedia Electronic Publishing*", Proceedings of second IASTED/ISMM International Conference on Distributed Multimedia Systems and Applications, Stanford, California, August 7-9 1995.

[8] Jean-Henry Morin, "Requirements for a Hypermedia Electronic-Newspaper Environment Based on Agents", in *Objects at Large*, D. Tsichritzis (Ed.), Centre Universitaire d'Informatique, University of Geneva, July 1997, pp. 177-193.

[9] Jean-Henry Morin, "HyperNews: a Hypermedia Electronic-Newspaper Environment Based on Agents", in *Proceedings of HICSS-31, Hawaii International Conference On System Sciences*, IEEE 1998, January 6-9, 1998, Kona, Hawaii, Volume II, pp 58-67.

[10] Jean-Henry Morin and Dimitri Konstantas, "HyperNews: A MEDIA Application for the Commercialization of an Electronic Newspaper", in *Proceedings of SAC'98, 1998 ACM Symposium on Applied Computing*, Atlanta, Georgia, February 27 - March 1, 1998, pp. 696-705.

[11] Jean-Henry Morin, "Commercial Electronic Publishing over Open Networks: A Global Approach Based on Mobile Objects (Agents)", PhD Thesis, University of Geneva, Faculty of Social and Economic Science, Departmant of Information Systems, March 1999.

[12] Vassilis Prevelakis, Dimitri Konstantas and Jean-Henry Morin, "Issues for the Commercial Distribution of Electronic Documents", in *Communications and Multimedia Security*, Sokratis Katsikas (Ed.), Vol 3, Chapman & Hall, 1996.

[13] Dimitri Konstantas and Jean-Henry Morin, "*Agent Based Dissemination of Commercial Electronic Information*", in this collection.

[14] J. Baumann, F. Hohl, K. Rothermel, M. Schwehm, M. Straßer, "Mole 3.0: A Middleware for Java-Based Mobile Software Agents", in Proceedings of Middleware'98, Chapman & Hall, 1998.

[15] Markus Strasser, Joachim Baumann and Fritz Hohl, "*Mole - A Java Based Mobile Agent System*", Second ECOOP Workshop on Mobile Object Systems, University of Linz, July 8-9, 1996.

[16] Jan Vitek, Ciaran Bryce and Walter Binder, "Designing JavaSeal or How to Make Java Safe for Agents", in *Electronic Commerce Objects*, D. Tsichritzis (Ed.), Centre Universitaire d'Informatique, University of Geneva, July 1998, pp. 105-126.

[17] IBM 4758 PCI Cryptographic Coprocessor, http://www.ibm.com/security/cryptocards/index.html

[18] Jan Vitek and Giuseppe Castagna, *"Towards a Calculus of Secure Mobile Computations"*, IEEE Workshop on Internet Programming Languages, Chicago, Illinois, May 1998.

[19] Cox B. (1996), "Superdistribution Objects as Property on the Electronic Frontier", Addison-Wesley.

[20] Christian Collberg and Clark Thomborson, "Software Watermarking: Models and Dynamic Embeddings", proceedings of the 26th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Langauges, San Antonio, Texas, January 20-22, 1999.