

Privacy for Digital Rights Management Products and their Business Cases *

Rüdiger Grimm

*University Koblenz and Fraunhofer Institute for Digital Media Technology
grimm (at) idmt.fraunhofer.de*

Abstract

DRM systems and their shops are analyzed with respect to the privacy of their customers. The analysis follows a structure of privacy principles in accordance with the European Directive for Privacy Protection. As an example, Apple's iTunes is analyzed in detail. From the analysis, recommendations for a better practice are derived.

1. Introduction: the problem with DRM

Before the efficient digitization of intellectual property such as music, films, games, images, and text, their content was strictly bound to physical media, from which they could not easily be copied and sent without loss of quality. The traditional business model of intellectual goods was the business model of their physical media. With the digitization, latest with the highly efficient compression method of MP3 (Brandenburg/Stoll 1992, and MP3 1992/94), this business model was corrupted. Customers don't pay for CDs, DVDs, or books, if they get the related content in (almost) original quality for (almost) free in the Internet.

Digital rights management compensates for the loss of physical binding, in that their mechanisms block the duplication of digital products on the end-user devices. More sophisticated, digital rights enforcement mechanisms execute specified user rights, coming as part of the content code, within the end-user devices. Digital rights management and enforcement aim to restore the classical business models in that they make sure that users can consume only those products they have paid for. Moreover, users can consume the products only in the very specific way, they have paid for. The history and functioning of DRM systems are discussed by many authors, see for example the excellent books of Rose et al. 2002, and Becker et al. 2003.

There are many DRM products out in the market, for example the Fairplay DRM kernel in Apple's iTunes, Microsofts Windows Media Rights Manager (WMM) in many shop systems, for example in Musicload of T-Online, the OpenMG in Sony Connect-Europe, to mention only the music market. For electronic documents, Adobe PDF has integrated DRM functions to support the E-Books format. There are many more systems in the market, they are all mutually incompatible, difficult to use, and – this is the topic of this paper – they are intransparent with respect to their handling of personal data. After all, they are badly accepted by the market, compared to the market potential. The Musicload Factsheet of 05.04.2005 (www.musicload.de) with reference to the 2004 Digital Music Report of the International Federation of the Phonographic Industry (IFPI), estimates 200 million legal music-downloads world-wide, while they estimate that this covers only 20% of all downloads: that is, 80% or 800 million downloads were illegal. Moreover, the numbers of legal downloads have remained constant from 2003 to 2004. The download market of digital music is both far beyond its potential, and it is in a stagnation phase.

In our study (Bizer et al. 2005) we claim, that this is due to the fact, that DRM systems are uncomfortable to use, they don't meet the real needs of the users, and they undermine the trust of users, in that they either misuse personal data of their customers or, at least, handle personal data with little care. In this paper I will explicate especially the privacy problem of existing DRM systems.

For this purpose, this paper explains how existing DRM systems and their shops can be analyzed with respect to the privacy of their customers. The analysis follows a structure of privacy principles in accordance with the European Directive 95/46/EC on Privacy. (www.cdt.org/privacy/eudirective/). As an example, Apple's iTunes is analyzed in detail. From the

* This study was sponsored by the German Federal Ministry of Education and Research under the program "Innovationspotenziale der Informationstechnik" 2005

analysis, recommendations for a better practice are derived.

An economic analysis of DRM systems is discussed by Will (2005) in his contribution to the Axmedis conference 2005.

2. The privacy problem in existing DRM systems

There are three fundamentally different approaches to the protection of rights on digital items. Approach number 1 is “strong DRM” which enforces rights: users *cannot* act illegally. Copyright mechanisms are the most simple form of “strong DRM”. Approach number 2 would not prohibit users to act illegally by technical means, but would personalize products in order to identify the origin of products in illegal environments. An example for this “trace model” is the LWDRM technology (Grimm/Aichroth 2004) by which users earn fair usage of digital items if they sign them: users do *not dare to* act illegally. Approach number 3 does not use technical rules enforcement mechanisms, but would encourage users to act legally by incentives: users do *not want to* act illegally. An example for this “incentive model” is the “Potato system” which encourages users into a provision model (www.potatosystem.de).

Obviously, the trace model and the incentive model must somehow use personal data in order to identify users either for prosecution or in order to realize the incentives. Both approaches must work carefully on a personal data protection model. The third approach, however, the “strong DRM”, which prohibits users to act illegally by technical means, should be strictly product oriented, with no reference to the person who legally owns the product. Because the technical mechanisms enforce well behavior, there is no need to either prosecute or reward the user.

However, most shop systems which use DRM, do not trust the built-in mechanisms of DRM to enforce the usage rules in the end-user devices. Therefore they use the trace method as a second line of defense. They collect data to identify users, not only for business purposes, but also to link products to their buyers in order to identify the origin of products in illegal environments. Most often, the usage of personal data within DRM protected products is intransparent to the customers.

3. The analysis structure of the privacy principles

It is helpful to use a privacy model in order to analyze DRM systems with respect to their usage of

personal data. The European Directive 95/46/EC on privacy provides such a model, which is widely accepted, not only within Europe, but also, for example, in the USA by means of the Safe Harbor Principles (US DoC 2005). The directive defines “personal data” as “any information relating to an identified or identifiable natural person”. The identified or identifiable person is called the “data subject” (Art. 2). The directive regulates the storage, processing, and usage of personal data by explicating the following principles

- Quality (Art. 6): the data must be lawful, fair, adequate, relevant;
- Legitimation (Art. 7): personal data must be bound to the purpose of the service, they may be used only by consent of the data subject or by a legal obligation;
- Purpose binding (Art. 7): the personal data must be necessary for the purpose, e.g. a contractual cooperation or the administration of a service, etc.;
- Transparency (Art. 10-12): the right of access by the data subject;
- User control: beyond transparency, the right of access, esp. the right of rectification (Art. 10-12), the right to object (Art. 14);
- Confidentiality (Art. 16): the organization must ensure the confidentiality of the personal data;
- Correctness and security (Art 17 on security, and the right to rectify the data, in Art. 10-12): the organization must protect the personal data against loss, distortion, and correctness with respect to the content;
- Supervision (Art. 18-19, and 28-30): regulations on a supervisory authority;
- Remedies, liability, and sanctions (Art. 22-24); regulations on the sanctions in case the service provider does not comply with the principles.

There is also an obligation to ensure that personal data may be transferred to a third country outside of the EU only if the “third country in question ensures an adequate level of protection” (Art. 25). The US Safe Harbor is an agreement which service providers may join freely in the USA in order to guarantee such an “adequate level of protection”. The principles of the European Directive 95/46/EC on privacy are not exactly, but somewhat closely mapped on the seven Safe Harbor principles: notice, choice, onward transfer, access, security, integrity, and enforcement. (US DoC 2005)

Of course, these principles may be checked against any shop that offers digital music. Web shops offer privacy statements on their Web pages, for example <http://www.apple.com/de/legal/privacy/statements>.

However, it is hard to know, where exactly, at which

point of communication between buyer and seller, and – most difficult – for which purpose, data are collected, stored and used by a shop provider, either directly, or by means of its DRM system.

In order to find out, which data are collected and stored at which point of the communication and at which place, each Web shop and its incorporated DRM system can be analyzed according to the following categories (Bizer et al. 2005, 2.4-2.5):

- Data flow *before* concluding a deal: while preparing a purchase, during user registration, and when selecting a product for purchase (placing a product into the shopping cart);
- Data flow *at* conclusion of a deal: at end of selection (closing the shopping cart), for payment of the products, and for delivery of the products;
- Data flow by checking the right to use a product: at first initialization of a player; at repeated usage; for rights update;
- Data flow through service functions: for example improvement of service, direct marketing, and security functions such as encryption;
- Data flow through hidden interfaces and by linkage of different functions: cookies, pixel tags (web bugs), combining customer data with clickstream data such as IP addresses or encryption keys.

There is an additional category of personal data, which is orthogonal to the other five categories, which we call “general data traces”:

- General data generated by the user consciously by filling forms;
- General data collected by the shop from the communication data;
- General data encoded within the product itself.

4. For example, Apple’s iTunes

4.1 Overview

Apple’s iTunes is a music portal with online stores in 19 countries worldwide. It offers 800 000 titles in 28 genres (by January 2005). Five major and numerous independent labels offer their titles in iTunes stores. Until March 2005 iTunes has counted over 300 Million downloads over all stores, that is ca. 40 Million downloads per month. In addition to music the iTunes stores offer videos, reading books, film trailers and radio streams.

The number of customers is not published. The target customer group is young people who love music, who are used to online surfing, and who are ready to pay for what they get.

The iTunes servers are located in the USA, they support the formats of M4P (MPEG-4 protected) and Codec AAC. The DRM system incorporated is called “Fairplay”.

See (iTunes Musicstore 2005).

4.2 Process model

Apple’s iTunes works like this, see figure 1. The user interacts with a shop server (steps 3 and 5), which stores registered content from the content provider (step 1). After installing the iTunes client software (2), the user browses the shop’s Web site and selects a piece for purchase (3). The shop first organizes the payment of the chosen product (4) and then delivers the content to the user (5). The user may consume the purchased products or share it with a specified number of devices according to the iTunes usage rules (6).

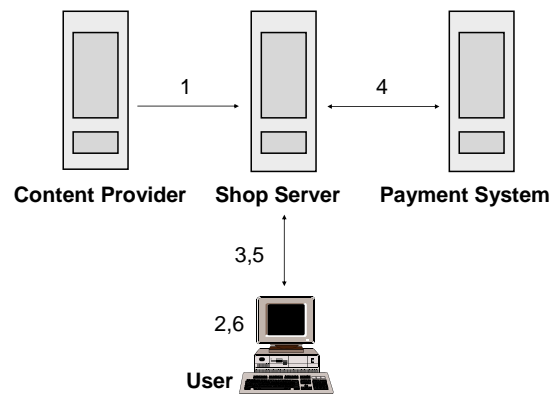


Figure 1: The process model of iTunes

The user may burn the purchased iTunes products on up to 10 CDs, and he may transfer the products to Apple iPods. He may also share the products with not more than six other computer devices which must be registered to the iTunes shop server in advance. Before a product can be transferred to a seventh computer, one of the first six must be de-registered before. iTunes products are encrypted with a symmetric key, which is encrypted by an asymmetric public key of the user device, before the product is delivered. The key pair of the user device is a function of some hardware parameters of the device, such that only a registered device can decrypt the symmetric content key (Bizer et al. 2005, section 3.1).

4.3 Data traces

Following the structure of section 3 above, we found the following personal data collected and stored by iTunes:

General data provided consciously by the customer at registration:

During registration, the customer fills a form which contains these mandatory fields:

- name, address, and telephone number;
- the user's e-mail address, which is used as his "Apple-id", and a related password on the user's choice;
- a secret question plus an answer, as well as the birth date in order to reveal a lost password;
- payment information such as credit card number and validity date.

Optionally, a client can provide his

- postal address for delivery of hard goods;
- fax number;
- mobile telephone number;
- tax number.

General data collected by the shop during communication:

- The operation system used by the client (version, language), the version of the iTunes software used, and the IP address of the client by means of the HTTP protocol;
- Cookies and session-ids. Cookies are expressively used to check how often which special sites of iTunes are visited by clients in order to improve the service; session-ids are used to organize the online session effectively;
- Device-id: for registration of a client device, the iTunes client-software derives a so-called Device-id from the hash of some hardware parameters and sends it to the iTunes shop; the Device-id will be used by the shop to encrypt a symmetric user key with which the delivered products are encrypted in order to bind it to a specific registered client device: this is the kernel of the Fairplay DRM system.

General data encoded within every iTunes product:

This part is rather intransparent to the customers of iTunes. However, it is simple to find these data in a hexdump of an iTunes product:

- Product-id, provided and maintained by the store;
- Apple-id, which is identical with the e-mail address of the customer;
- Meta data about the product such as title, name of composer, genre, year of production, etc.

It may be surprising to many customers, that their e-mail address is part of the product code. So, if they copy and send a product to other devices, their email address will reveal the origin of this file.

Data flow before concluding a deal:

Before completing a purchase, a user must be a registered iTunes customer. For this purpose, a user fills an online registration form. See above the "general data provided consciously by the customer at registration". Moreover, the registration is done over the HTTP protocol, by which more personal data are sent to the shop, like the operation system used by the client (version, language), the version of the iTunes software, and the IP address of the client, see above the first bullet point of the "general data collected by the shop during communication".

Other data are sent when using coupons from other users to fill up an internal iTunes account.

Data flow at conclusion of a deal:

A registered iTunes customer concludes a purchase in that he logs into the system with his Apple-id (which his e-mail address) and his password. The products in his shopping cart are associated with Product-ids by the shop. The shop stores the association of the Apple-id (the user's e-mail address), the list of Product-ids and the download-status. Even if the download status is "completed", this line remains in the store's database.

The shop encodes Product-id, Apple-id, and some meta data about the content within the code of every product before the content is encrypted with a user key and then downloaded by the customer.

Data flow by checking the right to use a product:

A PC must be initialized before it can play an iTunes product. Remember, that the content of iTunes products does not come in clear text, but it is encrypted with a symmetric key. The symmetric key is encrypted itself individually for any registered end-user device. For this purpose, a PC establishes a user account which contains private information to decrypt the symmetric product keys. The private decryption information of the user device is derived from some hardware parameters of the device (Device-id) and from the e-mail address of the user (Apple-id). For registration, the client PC sends its Device-id related Apple-id to the iTunes shop and, in turn, receives the necessary information to decrypt the symmetric keys that decrypt the content of iTunes products.

Another device, which is not registered, would not be able to decrypt the symmetric product keys, because the private information necessary for this step includes the local hardware parameters. Therefore, only registered end-user devices can play iTunes content (unless the DRM system is broken).

Registered devices play iTunes products offline. Therefore, on consumption, there is no more data flow between customer and shop.

Data flow through service functions

When a registered iTunes customer browses through an iTunes shop, key-words for search and visited Product-ids may be associated with IP addresses or Apple-ids.

A special feature is the so-called iMix. An iMix is a personal hit list of products which are assessed as best by the user. An iMix is associated with a self-chosen pseudonym (persona). An iMix together with its pseudonym (but not more) can be published. An iTunes customer can recommend his personal iMix to another user, in that he fills a special form which includes his own e-mail address and the e-mail address of the recipient. The recipient needs not to be an iTunes customer (but might now become one, because he likes the iMix). iMix, pseudonyms and e-mail addresses are associated by the iTunes system. This personal data reveals a certain user behavior.

Another special feature is the “pocket-money account”. There is a personal user-account internal to the iTunes system. It contains a kind of “pocket-money”, which can be used to purchase iTunes products. The smart idea behind the “pocket-money account” is that a registered iTunes customer can send money to another user who is not necessarily an iTunes customer. But he will certainly become one in order to enjoy the money presented to him. All related personal data (e-mail addresses as Apple-ids, money flow between the accounts, and subsequent purchases) can be linked by the iTunes system in order to learn more about its customers.

Coupons are another way by which customers can send money to other users. Again, the e-mail addresses which are or may become Apple-ids, are used for sending coupons.

Metadata of content can also be communicated in the iTunes system. There is a central metadata service, CDDB run by Gracenote (www.gracenote.com/music). An iTunes customer can access the metadata server of CDDB during his iTunes session. He can send the metadata of his content to the meta data server. And in case he has no metadata or he wishes to update his existing metadata, an iTunes customer can send

content to CDDB which would recognize it send metadata back to the user. During metadata communication all associated data like Apple-id and Product-ids are accessible by iTunes.

Data flow through hidden interfaces and by linkage of different functions

At every communication step with iTunes, all clickstream data out of the HTTP protocol, including IP address, language, HTTP-referer (“from which site am I coming”) and search key-words are accessible by the iTunes system and may be associated with other personal data stored in the internal data bases of iTunes, especially the Apple-id which identifies the customer by his e-mail address.

The following data set is an example of an HTTP header of a client request to an iTunes store:

```
GET
/WebObjects/MZStore.woa/wa/com.apple.jingle.app.store.DirectAction/viewNewReleases?fcId=14094475&pageType=newReleases&id=100
HTTP/1.1
Referer:
http://ax.phobos.apple.com.edgesuite.net/WebObjects/MZStore.woa/wa/storeFront
Accept-Language: de-de, de;q=0.75, en-us;q=0.50, en;q=0.25
X-Apple-Tz: 3600
User-Agent: iTunes/4.6 (Windows; U; Microsoft Windows 2000 Professional Service Pack 4 (Build 2195)) DPI/96
Cookie: countryVerified=1
X-Apple-Validation: 0F56EB06-7A5FFCA9109C3FC4E2B0CCA304ADC981
Accept-Encoding: gzip, x-aes-cbc
X-Apple-Store-Front: 143443
Host: ax.phobos.apple.com.edgesuite.net
```

More information about personal behavior is uncovered by the association of the Apple-id, iMix (favorites hit list), sent and received coupons and pocket-money. From the frequency and type of communication iTunes may learn a lot about its customers, not only as an anonymous customer, but as a real person in that iTunes knows name, address etc.

E-mail addresses are certainly personal data. Even if they may provide a certain degree of pseudonymity, the iTunes system can map it to the real person by means of its customer record exactly.

A clearly hidden interface is the e-mail address of the customer encoded in every product he has purchased. With this information every reader of the product knows its origin.

5. Conclusion: recommendations for a better practice

To complete the picture, other DRM system must be analyzed by the same structure. This would allow to compare the systems in a fair and transparent way. This has been done in the study (Bizer et al. 2005), which compares iTunes with Microsoft's WMRM in Musicload (T-Online), the OpenMG in Sony Connect-Europe, and the alternative PotatoSystem. As a result, the state-of-the-art of DRM systems can be described as follows:

They all collect more personal data from their customers than necessary to fulfill the purchase service. There are many hidden interfaces, both by encoding personal data within the products, and by linking clickstream data with contractual data.

It must be noted, that the online shops utilize a good part of their knowledge about their customers for service improvement or extra features to the benefit of the users, like the assessment of more or less purchased or visited products. But there is no need to link this information to the real users and their personal behavior like favorite lists or personal relationships to other users.

There are two important parameters which govern business, not only, but also in the electronic world: Trust and reputation. Hidden interfaces and encoded personal data demonstrate that the shops do not trust their customers. The second line of defense is prosecution of customers who use the products illegally by copying them to other users. But by encoding personal data within the products, *all* customers are put under suspicion. If this is done secretly (intransparent to the users), customers lose trust to their shops, and the shop will lose reputation. And this will reduce the business of a shop beyond its market potential. This is the situation as is.

It is recommended to do all personal data processing in a clear transparent mode. Customers, who are ready to pay for what they get, are ready to provide personal data if they know what it is good for. If it is really for the benefit of the customers, they would accept it. If not, these data shouldn't be used.

6. References

[1] Becker, Eberhard; Buhse, Wilms; Günnewig, Dirk; Rump, Nils (2003): Digital Rights Management. Technological, Economic, Legal and Political Aspects. Springer Lecture Notes 2770, Springer, Berlin etc. 2003.

[2] Bizer, J.; Grimm, R. Will, Andreas; Möller, J.; Puchta, S.; Müller, A.; Müller, M. (2005): Privacy4DRM – Datenschutz-

verträgliches und nutzungsfreundliches Digital Rights Management. Study for the Ministry of Education and Research (BMBF) of the Federal Republic of Germany. July 2005.

[3] Brandenburg, K.; Stoll, G. (1992): The ISO/MPEG-1 Audio Codec: A Generic Standard for Coding of High Quality Digital Audio. Journal of the AES, Oktober 1994, 780-792. First publication at AES-Convention, Vienna 1992.

[4] EU (1995): Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (www.cdt.org/privacy/eudirective/EU_Directive_.html).

[5] Grimm, R., Aichroth, P.: Privacy Protection for Signed Media Files: A Separation-of-Duty Approach to the Lightweight DRM (LWDRM) System. Proc. ACM Multimedia and Security Workshop 2004, Magdeburg, Germany, September 2004.

[6] iTunes Music Store (last download at 8.7.2005): www.apple.com/de/itunes, www.apple.com/de/itunes/store, www.apple.com/de/legal/privacy/statements.

[7] MP3 (1992/94): MPEG-1: ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 11172-3, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 MBit/s. Part-3: Audio. 1992. And MPEG-2: ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 13818-3, Generic coding of moving pictures and associated audio. Part 3: Audio. 1994. („MP3“ is „MPEG-1 Audio“ + „MPEG-2 Audio low sampling rates“).

[8] Musicload (2005): The Musicload Factsheet of 05.04.2005 (www.musicload.de) with reference to the 2004 Digital Music Report of the International Federation of the Phonographic Industry (IFPI).

[9] Rosenblatt, B.; Trippe, B.; Mooney, S. (2002): Digital Rights Management – Business and Technology. M&T Books, Hungry Minds Inc., New York, 2002, 288 pages.

[10] US Department of Commerce (2005): Safe Harbor. <http://www.export.gov/safeharbor/> (last update 3/2/05).

[11] Will, A. (2005): An economic analysis of music download platforms. Contribution to Axmedis 2005, Virtual Goods Workshop.