6.857 Computer and Network Security November 1, 2001

Lecture Notes 16 : Digital Rights Management, DMCA

Lecturer: Michael Freedman/Ron Rivest *Scribe: Mineweaser/Pick/Stein/Traub*

Outline

- Digital Rights Management (DRM): "Presented by Michael Freedman, an alum of 6.857."
- Digital Millennium Copyright Act (DMCA): "To say the least, a bit controversial"

1 Digital Rights Management (DRM)

Michael Freedman presented slides on Privacy Engineering Techniques in Digital Rights Management" which will be part of his presentation at the Workshop on Security and Privacy in Digital Rights Management, held in conjunction with the Eighth ACM Conference on Computer and Communications Security in Philadelphia, Pennsylvania. These lecture notes accompany a slide presentation available at `http://pdos.lcs.mit.edu/~mfreed/publications`

1.1 Privacy Engineering Techniques

Online services such as travel agencies and newspapers[1] often require users to provide personally identifying information (PII) in order to access even the most basic features.

Further, browser cookies are often employed without the user's knowledge or consent in order to trace activities and develop a profile of the user. Software which blocks these *cookies* is available, but its use often prevents access to common services. Internet users are left to choose between two alternatives: no privacy or no service.

Businesses collect PII because of its value in marketing to current and potential customers. For competitive and other reasons, PII is normally collected separately by each service. Two common approaches to the collection of this information are described below.

1. Based on evaluation of the requirements for the service, collect only the PII that is essential to the level of service required or requested. The evaluation process can be labor-intensive for the business, and it may be necessary later to obtain more PII from existing users in order to deploy new features.
2. Regardless of current requirements for the service, obtain as much PII as possible at the time of initial user registration. Relevant details can be extracted later, and new features deployed, without requiring further inquiries to the customer.

[1]The New York Times inquires about the "household income" of its users.

We begin with an overview of enabling technologies for digital rights management, including oblivious transfer (OT), private information retrieval (PIR), and secure function evaluation (SFE). We then address some practical privacy engineering and policy issues for DRM systems.

Digital rights management is of particular interest to copyright owners concerned about violations facilitated by Napster, Gnutella, and other peer-to-peer file sharing systems. Such unauthorized use of materials is increasingly problematic for copyright owners.

Ethical questions are best left to lawyers and civil activists. Instead, DRM schemes will focus primarily on the technical issues.

An access control list (ACL) is one common method for expression of digital rights. For each activity, an ACL specifies which users are permitted to perform that action. Examples include:

Example 1 *Listen to music for one week.*

Example 2 *Access an online resource for one month.*

In one DRM scheme, encrypted content is available for download by any user. However, only authorized users will be provided with the decryption key. A verification protocol in the DRM system checks the ACL when responding to each user request.

The focus of current DRM schemes is on controlling access to official data storage systems. Current schemes do not yet address subsequent unauthorized use of digital content. Future efforts to meet this requirement will likely involve *fingerprinting* of digital content with a user-specific watermark. Such watermarks will enable copyright owners to identify the source(s) of illegally distributed content. These methods are known as *traitor tracing.*

Question: Do you believe that fingerprinting is feasible?

Answer: Freedman worked at a company that develops software for fingerprinting of digital content. Most watermarking technologies are intrusive due to conflicts with end-to-end security mechanisms. Current DRM schemes are intended to protect the viability of digital content distribution for mass market applications. Numerous attacks, such as digital-analog-digital conversion, have been shown to degrade or defeat digital watermarks.

There are several major reasons not to define privacy as part of a DRM scheme:

- The definition of privacy is best left to members of the legal establishment that are well-versed in such matters.
- Personal preferences vary. Some users are willing to exchange personal privacy for online benefits, while others value privacy protection over convenience.

Example 3 *The CEO of a large company makes an online purchase of a book regarding bankruptcy. Public knowledge of this purchase may have an adverse impact on the company's stock price. In contrast, knowledge of the CEO's purchase of a science fiction novel would have a relatively minor economic impact.*

Example 4 *The TiVo digital video recorder keeps track of the programs watched and recorded by users. Such detailed records of user behavior provide highly valuable context to marketers and content providers at the cost of personal privacy.*

In a DRM scheme, users obtain access to online resources through the use of a *token* obtained from a rights fulfillment server (RFS). Several kinds of tokens are commonly available:

1. Pay per use —time-insensitive
2. Subscription —time-limited

Content and related tokens can be stored in a centrally located online locker to provide access to purchased materials from any location.

Question: What are the differences between rights and keys?
Answer:

- A *key* in this context is the decryption key for a symmetric-key cryptosystem. With this key, digital content can be decrypted for use.
- The policies of the copyright owner are expressed in a *rightsmanagementlanguage*. For a DRM client, *rights* define what actions are permitted of each user. The DRM client may reside with the end user, with the vendor, or with the copyright owner. There are many variations on the architecture.

Electronic cash, or E-cash, is a digital abstraction of physical cash. E-cash is often implemented with blinding techniques and mix networks to provide anonymity for buyers and/or sellers.

Mix networks for e-cash are similar to those previously described for electronic voting schemes. A simple flow diagram is shown in Figure 1 below.

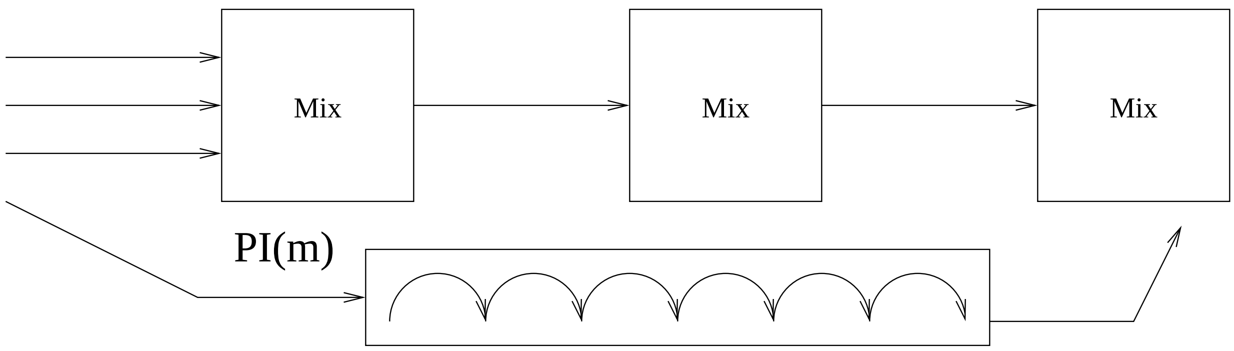

Figure 1: The mix network prevents an outsider from linking a given input to a given output. This is accomplished through the use of timing delays, padding of messages to a fixed size, introduction of false messages to foil traffic analysis, etc.

Question: Certificates vs. Credentials?
Answer: Credentials would prove some properties such as 'over 18 years old' without telling your name, and certificates would do the same thing, but provide the name as well.

1.2 Oblivious Transfer (OT) and Private Information Retrieval (PIR)

This section reviews two protocols that enable a user to obtain selected data from a server without revealing to the server which data has been received.

There are two competing goals:

1. Minimize computation overhead.
2. Minimize communication overhead.

Alice $b \in \{0, 1\}$	Bob M_0, M_1	These are inputs
M_b	Does not learn b	These are results

Figure 2: Oblivious Transfer

A discrete log problem: $C \in Z_g$ with g as its generator. C has a discrete log, but Alice doesn't know it.

$$\begin{aligned} PK_b &= g^k \\ PK_{1-b} = C/PK_b &= C/g^k \\ k &\in_R \{1 \dots q-1\} \end{aligned}$$

$$\begin{array}{ccc} \text{Alice} & & \text{Bob} \\ PK_b = g^k = C/g^k & \xrightarrow{PK_0} & PK_1 = C/PK_0 \\ H((g^{r_b})^k) = H(PK_b^{rb}) & \longleftarrow & \left\{ \begin{array}{l} r_0, r_1 \\ < g^{r_0}, H(PK_0^{r_0}) \oplus M_0 > \\ < g^{r_1}, H(PK_1^{r_1}) \oplus M_1 > \end{array} \right. \\ \text{gets } M_0 \text{ or } M_1 & & \end{array}$$

Question: What is the use of OT? How is it used?
Answer: This can be extended to be 1 of n schemes. So the sender sends n copies and does not know which one I get. This result is achieved at the cost of high communication overhead. Bob does not know what files Alice has retrieved.

Use: Online auctions. It helps hide which one of N, but it is not practical for mass market content distribution. There's also a lot of network overhead here.

Next Scheme: PIR; similar but removes the property that Alice only learns one thing. In this model, there are multiple servers, S_i, that each contain a copy of the same database $DB_i = DB$. The database contains n items $X_1, \ldots, X_n$.

Alice	Databases
—	$X_1 \ldots X_n$
Learns X_i	Does not learn i
possibly more	

Figure 3: Private Information Retrieval

The user inputs i, and learns X_i. However, no server should be able to learn i. It is possible to show that PIR and OT are essentially the same thing. Consider the following example in which there are two servers, S_1 and S_2 which contain exactly the same elements DB_1 and DB_2 respectively. We assume that the user is allowed to make a query to both databases, and that the databases are not allowed to communicate among each other [if they were able to communicate, they could tell each other what the user asked and learn something about the requested element]. The user picks two subsets of indicies E_1, E_2, such that $E_1 \subseteq [1, n]$ is a random set of n indicies and E_2 is

$$E_2 = \begin{cases} E_1 \cup \{i\} & \text{if } i \notin E_1 \\ E_1 / \{i\} & \text{otherwise} \end{cases}$$

For example, if Alice wants to learn element X_3, then Alice could form the sets $E_1 = \{1, 2, 3, 4\}$ and $E_2 = \{1, 2, 4\}$. Then Alice asks the first database for the XOR of all of the elements indexed by E_1 and asks the second database for the XOR of all of the elements indexed by E_2. In our example, the first database would return $v_1 = X_1 \oplus X_2 \oplus X_3 \oplus X_4$ and the second database would return $v_2 = X_1 \oplus X_2 \oplus X_3$. Notice that $v_1 \oplus v_2 = X_i$ always hold by this construction. In this case, however, Alice is able to learn more than just X_i; she learns the xor of a subset of values of her choosing.

1.3 Secure Function Evaluation (SFE)

Applies: Trying to compute a function among two parties. Alice knows x, while Bob knows y. Therefore, $z = f(x, y)$ which means compute z where Bob does not learn x, while Alice does not learn y.

Example 5 *Recommendation service at Amazon*

Example 6 *The Millionaire problem - how do we find out which one of us is richer without asking for the other's actual value?*

Example 7 Yao: *Garbled Circuit Evaluation*

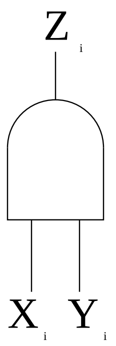

Figure 4: Garbled Circuit evaluation for SFE

1.4 Garbled Circuits Evaluation (for SFE)

First Step: Bob picks possibilities for $X_i : \{0, 1\}$. Bob then computes the table:

00	$E(G(0))$	$\longleftarrow 0$
01	$E(G(0))$	$\longleftarrow 0$
10	$E(G(0))$	$\longleftarrow 0$
11	$E(G(1))$	$\longleftarrow 1$

Figure 5: Example for AND gate

Alice learns $G[0|1]$ using OT since OT is used to exchange value between Alice and Bob.

1.5 Privacy Engineering in DRM Systems

DRM schemes which are effective in principle may be impractical to implement. For example, services offering anonymity or pseudonymity to users often control access with a passphrase that must be memorized by the user. If forgotten, often there is no way for the user to "reset" the passphrase without divulging his/her identity to system operators.

Another common problem involves the revocation of rights in a timely manner. When a user is removed from membership in a privileged group, the DRM system must promptly deny the user further access. This can be difficult to achieve in an efficient manner.

Case Study: DoubleClick, an Internet advertising and marketing firm, merged with Abacus Direct, a direct mail firm.[2] The companies' databases were integrated, bridging online and "offline" customer profiles. The public response to this action was predictably negative.

The goal of effective **fraud deterrence** is to reveal the identity of those users who misuse systems and/or digital content in violation of established policies and laws. A challenge for DRM is to deter fraud while preserving privacy for law-abiding users.

In many industries, such as music distribution, detailed purchasing records are required in order to appropriate calculate royalties for each copyright holder. One approach to preservation of privacy is to offer general subscription services to entire libraries of digital content. However, it is often difficult to provide details for royalty distribution under the subscription model.

In addition, an effective customer service operation often requires the collection of personally identifying information (PII). A customer may need to contact support personnel on multiple occasions to resolve a problem; the service team may issue a trouble ticket to the customer in order to manage open problems.

Further, effective capacity management requires sufficiently accurate and precise feedback on system utilization in order to minimize costs.

The focus of **European privacy legislation** is on management of personal information. Fair Information Practices (FIP) were developed to describe fair use properties.

Zero-Knowledge Systems tried to market "the totally anonymous Internet", but it was not commercially viable. They found that customers were either unwilling to pay for the service or simply not interested.

Network Externalities: Early e-cash systems on the Internet were less popular than existing credit card systems. The dilemma was that few users had e-cash, so few vendors were compelled to offer it as a payment option.

In an increasingly popular and effective approach to privacy engineering, a trusted third party issues a "seal of approval" to online businesses that have established privacy policies and demonstrated proper enforcement.

Another approach is to manage two separate databases: one for accounting and billing and another for capacity management. System logs can also be sanitized for long-term storage.

A business which offers only a choice between no privacy or no service does not serve the best interests of its customers. Instead, a rights management language, such as the Platform for Privacy Preferences (P3P), is needed to permit the expression of policies.

Question: Why should business care?
Answer: PriceWaterhouseCoopers (PWC) provides privacy audits with regard to legislation since users do not pay attention to math, so legal binding to practices is more important.

DRM: People have ideas that it is all bad. Not so ... It *can* be a privacy-enabler.

More information on DRM is available at `http://michaelfreedman.org`.

Question: Should an audit be enforced?

[2] `http://www.junkbusters.com/doubleclick.html`

Answer: Several stages. Business show third party that they are in compliance. They use the stamps of security approval sites on their web sites regarding policies.

Question: DRM schemes discussed today were focused on the web; what about digital systems in general?
Answer: These schemes are Internet-based and are not limited to the World Wide Web. In the future, DRM schemes must also be applied to TiVo and other consumer electronics, such as digital set-top boxes.

This concludes Michael Freedman's lecture on privacy engineering in digital rights management systems.

Rivest: DRM is the hardest thing to do in computer security, next to electronic voting.

Anonymization of voter is very hard. With DRM, you are putting control in the hands of your adversary. Therefore, deterrence and detection are both important.

In past year, several traitor training schemes have been discussed.

Example 8 *It is possible to insert messages in movies by filming in stereo, and in the editing stage build in a message by inserting one bit per scene.*

Copyright Issues:

1. Fair use
2. First sale
3. Resale —Not well-supported with DRM.
   Traditions: Do they need to be retained in the digital world?

"We are the government. We are here to help."

2 Digital Millennium Copyright Act (DMCA)

The DMCA was passed 3 years ago (in 1998).

2.1 High Bandwidth Digital Content Protection System (HDCP)

HDCP is intended to control content distribution by performing authentication between a personal computer and an attached monitor, then transmitting encrypted digital content directly to the monitor, where it is decrypted and displayed only for authorized users.

At a conference last summer, cryptographer Niels Ferguson announced his discovery of flaws in HDCP.[3] Ferguson cannot publish his research regarding the breaking of the HDCP security scheme

[3] Professor Rivest informed the class that he also discovered flaws in HDCP with only minor time and effort. After an initial review of the system, major fatal security flaws were apparent in the design.

due to the DMCA. According to professor Rivest, Ferguson explained: 'I was just reading it and it broke.'[4]

2.2 Dmitri Sklyarov

In Russia, it is legal to develop and sell software for (potentially unauthorized) decryption of protected digital content. Such protection schemes are implemented in many content distribution systems. One example is Adobe Systems' eBook system for the sale and control of electronic books. At a conference in Las Vegas, Nevada in the summer of 2001, Russian programmer Dmitri Sklyarov presented an technique for bypassing the eBook copyright protection scheme. After his presentation, Sklyarov was arrested by the Federal Bureau of Investigation (FBI) and charged with violations of the DMCA.

Skylarov has written a letter explaining how science progresses through the study, identification of weaknesses, and improvement of technology. He feels that the DMCA inhibits such progress by making it illegal to analyze copy protection schemes.

More information on the Sklyarov case can be found at `http://freesklyarov.org` and at `http://eff.org`.

2.3 Q and A with Professor Rivest

Question: A security scheme is not effective if it is breakable?
Answer: The word 'effective' is a term used in legislation, and it is a contradiction of facts with this statement.

Comment: One problem is that the DMCA is a very large piece of legislation. There are many other clauses beyond those we've discussed today.

The specifications for HDCP are available from Digital Content Protection, LLC.[5]

Rivest is divided on the issue of whether object code which exposes vulnerabilities in copyright protection schemes should be released. He said that the purpose of the release is the key issue. How do we design our society? Movies, software, textbooks (Rivest is a co-author of the popular *Introduction to Algorithms* textbook). Intellectual property is important to businesses and must be protected for economic reasons.

Question: What's up with the new law, the "Security Systems Standards and Certification Act" (SSSCA)?
Answer: Rivest thinks it is dead now. "A Trial Balloon"

Hal Abelson's course on Ethics and the Law on the Electronic Frontier is a good class to discuss this kind of thing in more depth.

The asset-rich film, television, and music industries maintain strong congressional lobbies to develop and promote favorable legislation.

---

[4] More information about this matter can be found at `http://www.macfergus.com/niels/dmca/`

[5] `http://www.digital-cp.com/`

Question: Laws are driven by industry. Reportedly, some are even *written* by industry?!
Answer: That's the way it works.

2.4 Recent Developments in the DeCSS Case

Victory for DVD Code Cracking `http://www.wired.com/news/print/0,1294,48075,00.html`

A California State Appeals Court ruled on Thursday that computer code used to "descramble" DVDs is "pure speech," and citing the First Amendment, the court reversed a trial court's order to block the code from appearing on the Web.

But since a federal appeal is still pending in a similar New York case involving the right to publish the DVD descrambler – called DeCSS – it's still unclear whether the "code is speech" defense sung by many techno-libertarians will provide them any shelter from the law.