Courseware needs Security

Frank Graf, Christoph Busch, Stephen Wolthusen

Fraunhofer Institute for Computer Graphics
Department Security Technology for Graphics and Communication Systems
Rundeturmstraße 6, 64283 Darmstadt, Germany
graf@igd.fhg.de, busch@igd.fhg.de, wolt@igd.fhg.de

To be able to fulfill future requirements for education and training, new learning scenarios with distributed, user-adaptive, on-demand, co-operative training environments to support time and space independent learning are needed. Imparting knowledge will become a valuable service and field of business since knowledge and continuous education are becoming a major contributing factor to economic success of any company. As a consequence, courseware and the right to access it will be a major object of trade. Since its stock of courseware is the capital of each training provider there is a considerable need to protect it from any misuse. This paper will discuss the security requirements special to the area of Computer Aided Learning (CAL). It will show how courseware can be protected from illicit use and distribution when using a security system, which allows throughout the existence of the data use control instead of access control restricted to the time of delivery

Keywords: **Security, copyright, e-commerce, learning environments**

1 Introduction

The computer is considered a promising platform for learning. Particularly its multimedia capabilities and its interactivity can provide a new level of quality in learning. The computer can be seen as a universal media device which can cope with all available media formats. Besides significantly cutting down the costs for training equipment it additionally can integrate all the different media formats into coherent multi-media presentations thus addressing the student on a higher and wider receptive level. Another advantage of computer aided learning is the capability to build interactive courses which adapt themselves to the student. Up to now real interaction between student and training provider was only possible in individual face-to-face training. Although this is a very effective form of imparting knowledge, it is expensive and not feasible for large numbers of students due to the enormous human resources needed. The computer makes interactive courses available to large numbers of students by allowing the usage of interactive material such as simulations or by dynamically selecting the material most appropriate for the student. So the student is more involved in the learning process which will increase his attention and overall motivation. Further on, to gain positive results in the simulations, he must actively apply the new knowledge and is no longer restricted to passively consuming it.
The WWW has established itself as the most important platform for CAL. Before the advent of the WWW, a working style isolated either in time or space was prevalent. The network provided only a narrow bandwidth and was therefore only used as a medium of transfer. Together with the WWW, high-bandwidth networks became available. This, together with the fact that integrated digital representation of different media became available, made the network the medium of access. Information is accessed and processed regardless of its location. This is the basic idea behind the concept of a virtual university [2][6], where students learn in a time and place independent manner.
A second development with great impact on the way people will work and learn is the fact that the speed of technical development is getting faster and faster. New knowledge emerges and must be mastered and incorporated in ever shorter periods of time. As a result of this, continuous, lifelong training will become a necessity. On the conceptual level, this implies that the concept of “learning for life” will be replaced in many areas by “learning

on demand". From a practical point of view, this has two implications. First, learning must be done more frequently and since it is becoming more and more part of the normal work it must be integrated as seamlessly as possible in the working process. The learning environment must be intuitive and easy to handle and configure. Currently training is often performed in a remote classroom environment. So, besides the cost for training there might also be considerable costs for traveling and accommodation. CAL supporting group learning and collaboration especially in globally dispersed groups can be the answer to this problem and since CAL becomes an ubiquitous resource, training on demand becomes reality: Learning sessions may be limited to the absolute minimum and delivered just in time.
Using the WWW as training platform is an obvious and promising choice, because it provides all the basic multimedia, communication, and collaboration functionality needed for education and training, it is available for virtually all platforms, and the vast majority of students is already familiar with it. However, there is the problem of protection of intellectual property when distributing the learning material on the WWW. Considering the tremendous time and amount of expertise that is invested in the preparation of course material it is essential to establish well designed protection mechanisms to prevent students from redistributing the downloaded material for their own benefit and bypassing the license regulations. Suitable countermeasures against piracy will ensure that the rights of the copyright holders are not violated and avoid loss of revenue, associated with the illicit usage of the material. Such protection mechanisms will be the key issues of this paper. The technology proposed will lower the inhibition threshold of courseware producers to provide their work in an E-Commerce environment.

2 Security Requirements in CAL

Knowledge is becoming more and more a product which can be and in increasing amount is traded. This general trend can clearly be seen in the following numbers. In 1997 six percent of the U.S. GDP was created by copyright transactions. A 1996 study by Economist's, Inc for the International Intellectual Property Alliance (IIPA) shows that the copyright industries are growing twice as fast as the remaining industries[1].

	1977-94	1987-94	1991-94
Copyright Industries	5.8 %	4.6 %	5.4 %
Others	2.3 %	2.3 %	2.3 %

Table 1: Estimated Average Annual Real Growth Rate

In the area of education and training, the courseware is the carrier of the knowledge. Since this learning material is stored and distributed in digital form it is particularly easy to copy. The following are the relevant issues concerning security.

2.1 Confidentiality

Due to fact that courseware by its nature must be distributed to the outside, industrial espionage and data theft is not the major problem in CAL. However, in certain scenarios confidentiality and restricting flow of information to well-defined groups is also a relevant issue. In case CAL is used for in-house training of employees it is quite likely that courseware might contain sensitive information, like marketing strategies or internal organizational procedures. Aside from countermeasures against physical theft of storage devices, diligent security administrators have to set up countermeasures against electronic theft such as the installation of firewalls and intrusion detection mechanisms to hamper attackers from the outside. This will cover only part of the risk that confidential data could be stolen. Confidential material can be leaked from a perfectly protected intranet – stolen by insiders, having legitimate access to the data in question. It is estimated that 80 percent of attacks originate from inside an organization.

[1] Note that these numbers are for the time before the WWW got widespread and popular.

It is also important that discussions between learning group participants must remain confidential and traceable.

2.2 Reliable Identification of Users

A training provider wants only legitimate user to access the courseware. Therefore a user must authenticate himself before he can access the courseware. Further on billing is based on the information which user accessed what material. Therefore a reliable identification of users is an important issue for CAL. When dealing with copyright violations a very beneficial feature is being able to trace back the path of an illegitimate copy and in this way to identify security leaks.
Adapting a course to a student requires the existence of a user profile. Whenever the student accesses course material this will affect his knowledge, so each access must be monitored and registered in his user profile. When collecting such detailed information on the actions of the user the training provider must make sure that this is done in a way that does not violate the student's privacy and legal regulations. Reliable user identification is especially important in case the learning will result in some kind of certification (university degree, certification program). If the training takes place in a distributed WWW environment it is almost impossible for the training providers to back up their guarantee. From the user information collected it can be proven that a certain user account has accessed the data. In case of prosecuting copyright violations this might be sufficient but for certification this is not the case.

2.3 Protection of Copyright

In order to create courseware an author must be expert on the subject, in order to decide which topics must be included and how they relate to one another. Then he must subdivide these topics into separate learning goals and organize these appropriately. For this didactic skills are needed. The learning goals must be broken down into concrete learning steps. This includes the definition of the layout and interactions for these learning steps. Here expertise is needed in the area of design as well as in pedagogics. When it comes to the creation of the basic course material (pictures, videos or simulations), there is the need for wide variety of technical skills, such as creating pictures, animations, sounds video, or programming interactive elements like simulations. Since there is presumably no single person who is at the same time an expert on the subject, a teacher, a designer, an artist, sound or video technician and a programmer, a team of experts will be needed for the creation of courseware. So the value of courseware is due to the large amount of expertise and human resources necessary for its creation. From our experiences, a good rule of thumb is that in order to create courseware of fairly good quality equivalent to one hour of course one has to invest about 100 hours of manpower. There is ongoing work [1][3][4][5] to address this problem, e.g. creating courseware in a template based approach to make use of common patterns or the reuse and multiple use of courseware. Nevertheless the creation of courseware will always need a substantial amount of human expertise and cannot be automated to a significant degree. So the copyright holders of courseware have a strong interest in protecting their courseware from illicit use and distribution.
The major drawback for copyright protection in CAL is that in order to do the job the copyrighted material must be made available to the students usually in an environment outside the control of the provider. A training provider can and will restrict access to courseware only to registered customers and the courseware can be delivered in encrypted form to the customer to avoid that eavesdropper can get hold of the courseware. But giving a user read access to digital data always means giving him the opportunity to redistribute the data, too. So the security of the courseware depends on the goodwill of the customers. Withholding the courseware until appropriate payment for the purchased goods is confirmed does not solve the problem, since it does not prevent one paying customer from redistributing illegitimate copies of the courseware.
The simplest way to pirate courseware is to redistribute it under a new name and claim the copyright for this new instance. Then the legitimate copyright holder can prove his ownership only by providing evidence that part of or all of the content in the pirated courseware is a copy of his own work. Many multimedia formats support the attachment of

copyright information. However, when knowing where this information is located in the data or simply by using the appropriate editor application it is no problem at all to alter this information. Most typically this copyright information is additional data which is independent from the content. So another possibility to get rid of the original copyright information is to capture the content when it is displayed, e.g. take a screenshot of the monitor. In the case of a digital copy the pirate might also slightly manipulate the content or use only selected parts of the content in order to make the fact that one material is a copy not so obvious. For images, examples for such manipulations that do not alter the user's perception of the content are reducing the color depth, slight rotations or alteration of the brightness or the color.
Secondary copyright is also an interesting issue. Consider an author developing courseware wishing to use material created by a different author as part of his course. Reuse would save him the effort to create the courseware and on a global scale it will avoid that the same content is re-implemented over and over again. But then the problem arises who the copyright holder of the new courseware is and who is entitled to the royalties for the new courseware.

3 State of the Art in Security

The problem how to distribute copyrighted or confidential material over networks such as the Internet is not specific to CAL. Instead it has been around as long as networks exist. In current state of the art environments one finds security mechanisms for secure storage, access control, and secure transmission of multimedia material. Although each of these techniques solves a single aspect of building up a secure infrastructure there does not yet exists an overall solution. Even with means of secure transportation of valuable data there is no protection once the data has reached its destination. At the destination point all classical systems convert documents, which might have been transmitted in encrypted format, to plain text. Thus transmitted data is not protected after subsequent storage. As a result of this

- Attackers may eavesdrop on the connection from the outside and subsequently access or steal the data.
- Consumers may copy the material and redistribute the content for their own benefit.
- The Copyright holder has no control over his data once it is delivered. Therefore he can not enforce that delivered data is used according to licensing conditions.
- There is no way to determine the legitimate copyright holder directly from the contents.

As a result, it can be stated that for an effective way to protect confidential or valuable courseware, the copyright holder must have complete control over the distribution and the means to trace back illicit use and distribution to the perpetrator. Furthermore, he must be able to provide proof of illegitimate activities. Therefore monitoring every access to courseware is essential.

4 Computer Aided Learning in a CIPRESS Environment.

The acronym CIPRESS stands for "Cryptographic Intellectual Property Rights Enforcement System" and is the internal code name for a joint development of the Mitsubishi Corporation, Tokyo, Japan and the Fraunhofer Institute for Computer Graphics based in Darmstadt, Germany to protect intellectual property from illegal use and to expedite reuse of copyrighted material for new content creations by combining patented re-encryption and watermarking technology. First an overview of the CIPRESS system will be given, then it will be discussed how CIPRESS can be applied to the area of CAL.

4.1 Overview of the CIPRESS security environment

The re-encryption technology is the key technology for preventing illegal distribution and allows exact monitoring of the usage of documents. The basic idea is all data is stored in encrypted form on the storage devices of a client computer. Whenever a document is accessed it must be decrypted. However this decrypted data is available only in the volatile memory section of the computer, then as soon as the document is saved to a storage device

it is encrypted again. The crucial point here is that CIPRESS does not reuse the encryption key but creates a new personalized key specific to the document whenever the document is written. Key creation is done by a central key storage facility, the so-called Key Center, which is located in a secured facility. Keys are exclusively generated and stored in the Key Center, therefore the Key Center is involved in every access to the document data and can monitor and document every usage of the data. Furthermore, if the Key Center does not deliver any key the user cannot access its content. For practical usage this implies an attacker from the outside might steal the data or even the entire storage device but all he gets is the data in encrypted form. To view the document's content he would have to contact the Key Center and authenticate himself to obtain the keys needed. But then the Key Center would detect this illicit usage and not deliver the key. Therefore it is not sufficient to steal the data, the attacker must also have the authentication of a user which has sufficient access rights. Since the Key Center logs all accesses, it is impossible to access the data without leaving a trace. Using these logs the entire usage trail of a document can be reconstructed. In case a pirated copy of a document is found, this allows the copyright holder to prove his ownership and to identify the security leak.

All this encryption and decryption is transparently handled by CIPRESS system components. These components exploit the cryptographic hash (a.k.a. digest) of a document in order to retrieve the corresponding key from the Key Center. The digest is a quasi-unique identifier, which is calculated in a way that with sufficient probability for two different documents (even if differing only by a single bit) the same digest will never be generated. As a consequence, once a user alters the content of the document the modified content will be treated as a new document. Having to register all system files or application programs and reregister them whenever the file has changed, would create a lot of network traffic and a tremendous overhead of data in the Key Center. This is the reason why our system also supports local documents for non-shared usage. Like CIPRESS documents, they are automatically encrypted and decrypted but instead of using the re-encryption approach and obtaining the keys from the Key Center, a machine specific key, the so called master key, is used. This master key encryption is specific for every client, so that master key encrypted documents can only be accessed on the computer they have been created on. Once a document is completed and should become generally available, the document must be registered at a Content Server. From then on the document will be encrypted using the keys from the Key Center and thus will become subject to the Key Center's usage control and permanent surveillance. Registration must be done at a Content Server. Content Servers provide centralized and persistent storage of digital documents. User can retrieve documents from a Content Server either using a CIPRESS specific application, which is also used to register documents, or alternatively they can use a WWW front-end. The encrypted material can also be made available on standard file servers.

From a technical vantage the CIPRESS system runs on standard PC hardware and is a extensions to the Microsoft Windows NT 4.0 operating system. Thus the security features of CIPRESS are added transparently to the system and apart from reduced performance in no way affect off the shelf software. Figure 1 shows the system architecture. Access control is realized on top of an own user and group management based on X.509 certificates. Therefore CIPRESS can be used globally without being restricted to NT domains. Besides the pure software based version also a version supporting smart cards will be available during 1999, which gives an additional degree of security and increased performance.

To allow collaboration with other users, documents must be distributable via network. Therefore one of the CIPRESS components monitors all the incoming and outgoing

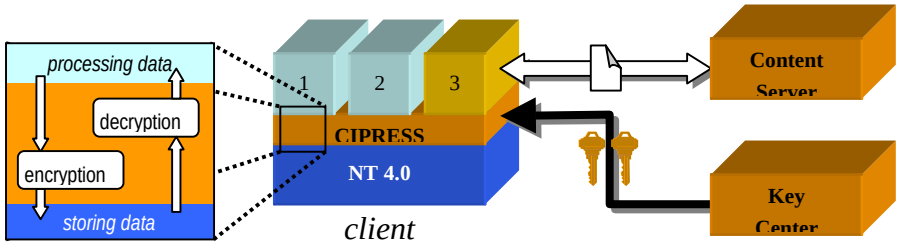

Figure 1: CIPRESS system architecture

network traffic. When receiving encrypted data this component will transparently decrypt the data. In case the communication takes place with a computer outside the secured area this component will use SSL to encrypt the communication. So, just like a Virtual Private Network (VPN), CIPRESS guarantees that information which is sent via the insecure Internet cannot be intercepted. To guarantee confidentiality CIPRESS will establish such a SSL connection only between certified hosts.
The combination of these two mechanisms secures all distribution channels which involve a computer. However, they cannot come in effect when the data leaves the computer, e.g. when a user prints out the material. By embedding an identifier of the current user into the document content itself, CIPRESS improves the changes to identify an attacker. The following types of media can be watermarked: Pictures, video, sound, and geometric data (experimental). A digital watermark must guarantee the following:

- Invisibility. The slight changes caused by the watermarks must not alter the user's perception of the document's content.
- Persistency. The watermark must not be removable by transformations which do not lead to a significant degrease of quality.
- Provability. The copyright holder must be able to prove he embedded in a watermark.

The watermarking technology used by our system for images and video is unique in that it allows the embedding of hierarchical watermarks. Since every watermark causes a slight decrease in quality and these losses accumulate, maximal five watermarks are used. The first one is a public watermark allowing automatic retrieval of document data. The subsequent secret watermark is embedded immediately when the document is registered and holds the ID of the copyright holder. By reading out the public watermark everyone can determine the copyright holder for a given document. This is of utmost importance for the copyright holder since so he can prove his claim in court. The other three watermarks are secret ones. Whenever a user accesses a document, a further secret watermark holding the ID of the accessing user is automatically added. In case a document already contains the maximal number of five watermarks a new watermarking cycle starts, i.e. the system will retrieve a new version of the document from the Content Server containing only the public and the first secret watermarks. Only a CIPRESS administrator knows the secret keys which are necessary to read out the information from the secret watermarks and is therefore capable of determine the most recent users. The entire line of document users can been retrieved from the records kept at the Key Center. The watermarks will stay in the information even when it leaves the media computer and will survive the most common transformations. CIPRESS can detect its watermarks even if they are scanned in from printouts, up to 90 percent of the original image has been discarded, or the image is converted to a grayscale image.
Although watermarking cannot prevent illicit distribution and usage of copyrighted material it is of utmost use once a copyright violation is to be proved in court. In case an insider was involved in the data theft, this malicious user can be identified and legal steps can be taken against him.
In CIPRESS a document can either be registered as a normal or a Data Linkage document. This conceptual classification allows to correctly handle the issue of secondary copyright. According to the Data Linkage concept a user, if he wants to include foreign material, must not embed this foreign material into his material. Instead he must use a external references. The user then registers the new material as a Data Linkage document with himself as copyright holder. Since composition of material is an intellectual work of its own right this is an appropriate way to handle the copyright problem. Since used material is always referrenced but never embedded, the orignal of the referrenced material must be accessed as well whenever the composition is accessed. This gurantees that the copyright protection cannot be circumvented.

4.2 Realization of Computer Aided Learning with CIPRESS

The big advantage of CIPRESS is that since it is implemented transparently at the operating system level, it provides security features in a transparent way and can therefore be combined with any learning environment. So using standalone CAL applications poses no problem at all and, due to the integrated WWW front-end, the system can also immediately be adopted for WWW based CAL systems. CIPRESS will take care of the storage, retrieval

and delivery of the courseware as well as communication with teachers, tutors, and other students. The logic for course control, maintaining user profiles as well as course and user administration must be provided by a special system, such as IDEALS MTS [3] or in the simplest case a plain WWW server.

The fact that CIPRESS, as opposed to the WWW, grants access only to students which have successfully authenticated themselves, might at first look like a limitation. However the opposite is the case. The reliable identification of user is a sine qua non for the following functionality:

- Access control. Only legitimate users can access the courseware. Due to the combination of re-encryption, data linkage and watermarking, access control within CIPRESS does not end once the courseware is delivered to the student. Instead, access control is performed over the entire life span of a piece of courseware. For the practical side this means that sensitive or confidential information can be included into courseware since CIPRESS prevents illicit usage and distribution even for courseware which is located on the students' PCs.
- User-profiling. Since CIPRESS detects and records the access to material at any time and not only when the user learns, one can reach a much higher level of accuracy of the user profiles if the learning system is given access to information gathered by CIPRESS. So a student can no longer access courseware outside a learning session without the learning system becoming aware of it.
- Accounting and billing. Accounting and billing is in the one or other form always connected to the actions of the students. Since the Key Center keeps record of every key delivery all the basic data needed for accounting and billing is readily available at the key center. Large scale accounting and billing like operating a Key Center in a secure and reliable way require resources which probably exceed the capabilities of small or medium enterprises. Therefore it would be a reasonable business model that an Internet provider or telecommunication company, which has the technical and administrative resources, operates the Key Center and, as a service to their customers, will do billing for them. This would give small and medium enterprises the opportunity to become a training provider, without having heavily to invest into hardware and accounting personnel.
- Performance tuning and quality management. The information how many students have accessed the system is important information for the training provider, since it allows him to optimize the overall performance of the system. The information which courseware is accessed and to what extent helps the training provider to cover the needs of his customer and also gives some insights for quality management.

One unique feature of our approach is the fact that, due to the re-encryption technology, the owner has a permanent control over his courseware. With CIPRESS, the student can use the courseware only if he receives the keys for it from the Key Center. Besides preventing the illicit re-distribution of courseware, this lifelong control in conjunction with the CIPRESS user administration allows some beneficial features, which are not possible in standard learning environments.

- Time restricted use of courseware. Normally a student can once the material is delivered to him access the courseware as long as he keeps the data. The owner of the courseware can, by setting the CIPRESS permissions, at any time cause the Key Center to no longer distribute any keys for a specific piece of courseware. With this feature a training provider can restrict the student's usage of the courseware to exactly the time span the student has paid for.
- Ease of maintenance. Also this feature is very beneficial for keeping the courseware up to date. Suppose the training provider decides that specific piece of courseware should no longer be used, because the content has become obsolete, is faulty, or for some legal or political reason. With traditional systems the training provider had no means to make sure that the students will not continue to work with the material delivered to them. With our system all he has to do is set the access rights accordingly.
- Delivery in advance. Although courseware will normally be delivered on demand via the WWW, there are certain scenarios where using a different distribution channel makes sense. Using CIPRESS and a WWW-based training system requires in any case the existence of an online connection for obtaining the keys from the Key Center and exchanging control information with the course control system. These are relatively

small amounts of data, so sending them causes almost no delay. Other considerations apply to multimedia material, especially high quality video, which is retrieved on demand from the server. The training provider might send the courseware in advance on a CD-ROM. Then the courseware could be loaded from the file-system and so significant delays which disturb the flow of learning and high communication bills could be avoided. When doing this, it must be guaranteed that the student cannot access the material (exercises, tests) before he is supposed to do so. When using CIPRESS, all the owner – or even better the learning system automatically – has to do is set the access rights accordingly. When the material arrives on CD-ROM the student has no rights, therefore receives no keys and cannot access the material. Immediately before the student should access a piece of courseware, the learning system grants the rights and the student can access the material.

- Personalization. Different student have different backgrounds, different learning goals or simply have paid for a different level of support. So ideally, there should be a specific set of courseware for each of the students' needs. The training provider must on the one hand ensure that a student can access all the courseware he needs or is entitled to, but on the other hand he does not want to give a student access to material he is not entitled to. When the courseware is delivered via CD-ROM it is not practical to create for each student a personalized CD-ROM. With CIPRESS the training provider can send the same CD-ROM, containing all the available courseware, to each student. Since the courseware is encrypted, the training provider controls which courseware can be accessed by a student by granting the access rights. It is also no problem to give the student later on access to more material or imply further restrictions, simply by changing the access rights. No new CD with material must be delivered to the student.

The conceptual support for secondary copyright (Data Linkage) comes in handy in case courseware authors are paid based on how often their courseware is used. Since it is CIPRESS philosophy not to merge material from different copyright holders but use Data Linkage instead, the royalties for the copyright holders can easily be derived from the logs of the Key Center.

5 Conclusion

The CIPRESS system provides an integral and transparent security approach for handling documents in secure environments that fulfills all the requirements for CAL. Our approach is unique in guaranteeing permanent control of usage for the documents it maintains due to combining the concepts of re-encryption and watermarking. This permanent control holds also when used in collaborative environments. The concept of Data Linkage is the basis for being able to handle secondary copyright in an intuitive and consistent way.

This research was possible by financial support of the Planning & Coordination Department of Mitsubishi Corporation.

References

[1] E. Duval, "Reuse of Educational Resources through Telematic Means (ARIADNE at HOME)", in Proceedings of ED-MEDIA '98, Freiburg, Germany, pp. 39-40, (1998).
[2] J.L. Encarnacao, et.al., "A concept and system architecture for IT-based life-long learning.", Computer & Graphics, Vol22, No. 2-3, pp 319-393 (1998)
[3] F. Graf, M. Schnaider, "IDEALS MTS - A Modular Training System for the Future", Proceedings of ED-Media '98, Freiburg, Germany, pp 486-492, (1998)
[4] F. Graf, M. Hausding, "An Open Framework for Monitoring Learning Success in WWW-based Training", Proceedings of EAEEIE 99, Capri, Italy, (1999)
[5] R. Hauber, T. Kopetzky, M. Mühlhäuser, "Lifecycle Support for Hypermedia Based Learning", in Proceedings of ED-MEDIA '98, Freiburg, Germany, pp. 531-536, (1998)
[6] J. Paaso, "Computer Based Teaching Technology for Software Engineering Education", dissertation at University of Oulu, Oulu, (1998).