Report on the DIMACS Workshop on Management of Digital Intellectual Property, April 17-18, 2000

Organizers:
Joan Feigenbaum, AT&T Labs
Dan Boneh, Stanford University
R. Venkatesan, Microsoft Research

Editorial Assistant
Sri Divakaran

1 Summary of Talks on Monday April 17, 2000

1.1 Efficient Distribution of Fingerprinted Content, Josh Benaloh, Microsoft Research

In this talk, Dr. Benaloh considered content protection in the context of An In-Flight Entertainment System and proposed fingerprinting as a mechanism for determining the source of the leak in the event the content is distributed illegally.

In this setting, Dr. Benaloh presented a method of fingerprinting protected content used by Boneh and Shaw and many others. In this method the digital content is divided into clips and two or more copies of each clip – each containing a distinct mark – are made. Each recipient of the protected content would then receive an individualized sequence of clips that could later be used to identify the recipient from any released content. For this method of fingerprinting, one of the ways to distribute marked content is to encrypt each copy of each clip with its own key and distribute the set of all copies of all clips over a common medium such as a CD, DVD, or broadcast channel. This reduces the problem of distributing individualized content to the problem of distributing individualized key sets. However, if the number of clips is large, even these key sets can be impractical to distribute individually. Also, this method of distribution collusion is still a problem. An alternative that Dr. Benaloh proposed is a method where each of these key sets can effectively be compressed into the size of a single key and then have a single encryption for each recipient. This method of reducing the key set size enables certain distribution options that would otherwise be precluded and also collusion does not provide access to additional clip keys. In this method the time to encrypt/decrypt each clip key is proportional to the number of clip keys and this method can be made efficient using amortization techniques.

Finally, Dr. Benaloh concluded his talk by mentioning that flexible finger printing methods are an important tool in content protection and they require large sets of keys and some of the techniques presented in this talk would help in making these methods efficient by reducing the bandwidth requirements for key distribution.

1.2 Watermarking, Tamper-Proofing and Obfuscation: Tools for Software Protection, Christian Collberg, University of Arizona and Clark Thomborson, University of Auckland

In this talk, Dr. Collberg classified the attacks on software intellectual property into three types, namely, reverse engineering, software piracy and tampering. It is believed that complete protection against these types of attack cannot be attained. However, software watermarking, tamper proofing and obfuscation have emerged as reasonable defenses against the above mentioned three types of attack.

Obfuscation attempts to transform a program into an equivalent one that is harder to reverse engineer. Tamper-proofing causes a program to malfunction when it detects that it has been modified. Software watermarking embeds a copyright notice in the software code to allow the owners to assert their intellectual property rights.

Dr. Collberg outlined the following :

- code obfuscation by lexical transformations, control transformations and data transformation;
- a thread model for watermarking (i.e., what constitutes a reasonable level of attack and what specific techniques the attacker employs);
- the different types of attack (additive attack, distortive attack, collusive attack);
- Static and dynamic watermarking techniques and dynamic graph watermarking;
- Tamper Proofing;

1.3 Divx, DPRM, and SDMI, Drew Dean, Xerox PARC

Divx was the first consumer product based on controlled use of digital content. Its cancellation suggests that while the technology is promising, there are a number of issues left in designing digital property right management systems that will have widespread success in the consumer market. This talk examined technical and policy factors that influence consumer acceptance of digital property rights management technology. In this talk, Dr. Dean discussed the technical and legal aspects of the security of these systems.

1.4 Long-Lived Broadcast Encryption, Juan Garay, Jessica Staddon and Avishai Wool, Bell Labs

Traditional broadcast schemes encrypt the digital content so that only privileged users can recover the content from the encrypted broadcast. The key material is held in a "tamper-resistant," replaceable, smartcard. A coalition of users may attack such a system by breaking their smartcards open, extracting the keys, and building "pirate decoders" based on the decryption keys they extract.

In this talk, Dr. Staddon introduced the notion of *long-lived broadcast encryption* as a way of adapting broadcast encryption to the presence of pirate decoders and maintaining the security of broadcasts to privileged users while rendering all pirate decoders useless. The long-lived broadcast encryption scheme for a system consisting of n users with r keys works as follows :

> To begin with, each user has r valid keys. As the knowledge of the compromised cards are received, the keys of those cards are rendered useless for recovering the broadcast messages. However, the keys that are not contained in the compromised cards are still valid and can be used to encrypt messages. After a certain number "d" (that will be determined later) of cards are compromised, the active users who no longer contain any valid keys are re-carded.

Long-lived broadcast encryption schemes are a more comprehensive solution to piracy than traitor tracing schemes, because the latter only seek to identify the makers of pirate decoders and don't deal with how to maintain secure broadcasts once keys have been compromised. When a pirate decoder is detected in a long-lived encryption scheme, the keys it contains are viewed as compromised and are no longer used for encrypting content.

In the first part of the talk, Dr. Staddon provided a formal framework for analyzing the various broadcast encryption schemes and then presented some results that demonstrate that the broadcast encryption scheme can :

- tolerate a high number of compromised cards despite the fact that it was designed to allow any set of m users to be excluded;
- tolerate a high number of compromised cards before any recarding of users is necessary;
- provide both theoretical and experimental evidence that this scheme achieves a steady state in which only a fraction of the cards need to be replaced in each epoch;

Finally, Dr. Staddon concluded by mentioning that more analysis is required to understand this scheme and also raised the following open question :

> can something meaningful be obtained by considering a proactive approach rather than a reactive approach ?

1.5 Digital Rights Management: Research Questions and Practical Implementations, Stuart Haber, InterTrust STAR Lab

The newly widespread availability of information in digital form raises a number of interesting questions about how to design a system for commercial transactions involving such information. In the non-digital world, intellectual-property regulations governed (and govern) many aspects of access to information; it remains to be seen how this will transfer to the digital world.

In the first part of his talk, Dr. Haber presented his thoughts about the requirements that one might desire of a system for commerce in digital intellectual property. Then, during the second half of the talk, Dr. Haber raised the following issues that are open for research :

1. What is a good business model for the Internet ?
2. What constitutes a good commerce architecture ?
3. How to build a protected processing environment ?
4. How do we analyze risk and manage fraud ?

5. How do we define a commerce control language ?

Finally, Dr. Haber concluded his talk by stating that with the growth of the Internet, there is an increasing need for building the necessary infrastructure to manage digital information and there are a lot of open issues.

1.6 Television and Tamper Resistance, Paul Kocher Cryptography Research, Inc.

In this talk Dr. Kocher addressed television piracy and motivated the topic by drawing parallels between the issues of television piracy and the piracy issues of digital intellectual property and how we can gain from the past experiences in combating television piracy.

Through the tremendous growth in digital TV, broadcast digital video, businesses have been losing more and more money because of piracy. So there is an increasing need to combat television piracy. Eliminating piracy is a complex problem that spans several disciplines like cryptography, risk management, law, entertainment, psychology, etc. Dr. Kocher then summarized the various approaches taken by by the television industry over the last 30 years and discussed the basic idea behind those approaches, their drawbacks and what we have learned from these drawbacks. In this context, to summarize a few, he presented the following schemes:

- the first CA-TV schemes
- 2nd generation analog schemes
- a few European schemes like DISCRET, OAK, SATPAL
- C-Band, Video Cipher II, Video Crypt, B Sky B

Dr. Kocher concluded his talk by stating that there is a need for new technical solutions for the problem of television piracy and solving this problem is important because it can save several millions of dollars for the cable industry.

1.7 Intellectual Property Protection for Software, David Krivoshik, Gibbons, Griffinger & Vecchione

In the past, laws were framed to protect physical ownership of property. Now, with the growth of Internet, digital content has become valuable. Since digital content is easy to copy and distribute, there is an increasing need to design laws for protecting digital intellectual property. In this context Dr. Krivoshik addressed the following issues :

- What are the legal issues involving in defining intellectual property rights that can meaningfully protect intellectual propery without restricting people's rights ?
- Why do the conventional laws on physical ownership not extend to intellctual property ?
- What types of content are patentable ?
- What are the problems with the existing patent laws and what needs to be done in order to come up with meaningful patent rules ?

1.8 Stronger Copyright Protection for Cyberspace: Desirable, Inevitable, and Irrelevant, Andrew Odlyzko, AT&T Labs - Research

Major revisions of copyright laws are being enacted around the world in response to strong demand from publishers, movie studios, and other "content providers." These revisions substantially restrict many traditional rights of users, such as the "first sale" doctrine. The push to strengthen copyright protection is resisted by scholars and librarians, who fear drastic curtailments in public access to information.

In this talk, Dr. Odlyzko argued that the the outcome of the ongoing battle will not matter much, and that specific provisions of copyright laws will be of minor importance. Technological, economic, and sociological factors will be the primary determinants of how information goods are sold. Content producers are likely to get stronger legal protection. However, they will probably find that they won't need it. On the one hand, effective electronic commerce will make customizable contracts much easier to arrange, so that content providers will be able to obtain any restrictions they wish through contract law. On the other hand, competition will make tight restrictions inadvisable. Netscape did not make its source code public because of any deficiencies in copyright laws, after all. Reduced barriers to entry and reduced costs do bring about a much more competitive market for most information goods.

Stronger copyright protection is desirable, primarily because it would enable a greater degree of price discrimination by content providers. It is also inevitable for several reasons. At one level, it reflects the political influence of content providers, which is visible in the rash of legislation around the world. At a more basic level, it reflects the needs of an information economy. The market for scholarly information is small. Even the trade press is not very large. The total of about $20 billion per year in book sales in the U.S. pales in comparison with the revenues of software companies, and especially with the total IT sector of the economy, which is estimated to be around $600 billion per year. The effective functioning of this much larger sector requires allowing flexible contractual arrangements. Therefore the government has limited options in using its power to decide which contracts to enforce, especially since in an online environment content producers will have great flexibility in packaging their goods. (The current antitrust lawsuit may establish that Internet Explorer is not an integral part of Windows 98, but it is clear that Microsoft can, if it chooses, make it an integral part of a later operating system.)

1.9 Large Scale Copy Detection, Narayanan ("Shiva") Shivakumar, Gigabeat

Currently, any small time cyber-pirate can make copies of music CDs and books available on the web in digital format to a large audience at virtually no cost. Content publishers such as Disney and Sony Records are therefore expected to lose several billions of dollars over the next few years in copyright revenues. In this talk Dr. Narayanan addressed this problem by proposing a copy detection system (CDS) where content publishers will register their valuable digital content. The CDS then crawls the web, compares the web content to the registered content, and notifies the content owners of illegal copies.

In the initial part of the talk, Dr. Narayanan presented the the two approaches to protect content: copy prevention and copy detection and discusses their shortcomings. Dr. Narayanan then discussed how to build a copy detection system that is accurate, scalable (e.g., to hundreds of gigabytes of data, or millions of web pages) and resilient to "attacks" (e.g., partial audio clips) from cyber-pirates. In this context, he then presented the following

- signature based approaches
- approaches based on iceberg strategies.

Finally, Dr. Narayanan concluded the talk by presenting the following:

- SCAM (Stanford Copy Analysis Mechanism);
- FRAUD (Finding Replicas of AUDio);
- DECIV (DEtecting CopIes of Video) for finding textual documents, audio clips and video clips on the web;

1.10 A Graph Theoretic Approach to Software Watermarking, Ramarathnam Venkatesan, Microsoft Research

The problem of software watermarking can be defined formally as follows :

> Given a program P, a secret key K, a watermark W, output another program Q that is functionally equivalent to P so that given Q and K, one can compute the watermark W using a recovery algorithm A (i.e. $A(K,Q) = W$), any simple transformation of program Q into another program $Q\prime$ does not remove the watermark W (i.e. $A(K,Q\prime) = W$) and without K one cannot find or detect whether the program Q is watermarked or not.

In the initial part of the talk, Dr. Venkatesan motivated the need for software watermarking and also mentioned the defficulty involved in defining this problem rigorously. In the later part of this talk, Dr. Venkatesan introduced program and flow graphs as a natural abstraction for software watermarking, where nodes represent the basic blocks of code and edges represent the control flow.

In this context, Dr. Venkatesan presented a simple scheme for software watermarking that would insert code that does not ineract with the variables of the original program. This scheme would not work because a simple flow analysis would eliminate the inserted code. Dr. Venkatesan concluded his talk by presenting a more sophisticated scheme that attempts to force the attackers into doing excessive visual examination. This scheme assumes local indistinguishability.

1.11 Passive Fingerprinting, Yacov Yacobi, Microsoft Research

In this talk Dr. Yacobi presented a fingerprinting scheme that improves on the Boneh-Shaw scheme in the following two ways :

- By merging a Direct Sequence Spread Spectrum (DSSS) embedding layer with the first Boneh-Shaw layer (the so called "Γ code"), effectively increasing the protected object size by about four orders of magnitude. As a result we have more than one order of magnitude improvement on the size of collusions that we can overcome.

- By replacing the "marking assumption" with a more realistic assumption, allowing random jamming on the so called "unseen" bits.

2 Summary of Talks on Tuesday April 18, 2000

2.1 The Effect of Anti-Circumvention Provisions on Security, Jonathan D. Callas, Counterpane Internet Security and Bruce Schneier, Counterpane Internet Security

One of the properties of digital Intellectual Property (IP) is that it can be easily reproduced, modified, and transferred. In response, IP owners have created new security technologies for controlling the digital works. Inevitably, this creates an opportunity for those who can circumvent those technologies.

Recent changes in copyright law attempt to address this on-going battle by prohibiting circumvention of these technologies. Unfortunately, this well-meaning provision does not tie the circumvention to infringement; circumvention is prohibited even if the technological measures prevent someone from something they are entitled to do. This leads to a number of unfortunate effects on development of security systems, including techniques that protect intellectual property itself.

In the first part of the talk, Dr. Callas presented the two basic approaches to the design of security models, namely the closed model and open model. In the closed design model, the security design is done in a closed group and often the design is reviewed by external people. This approach has the following advantages :

- it is difficult to break a closed design;

- security is maintained through obscurity;

- the design can be targeted towards specific needs;

- potentially less commitees

However, a closed design has the following dis-advantages

- team blindness can lead to poor security;
- easy to break the system once the design gets revealed;

In the open design model the security design is based on the principle of minimal secrets. The advent of international standards has resulted in more persons adopting the open model for designing security. In this model, the advantage is that there are very few surprises. However, this model has the following dis-advantages :

- Some problems cannot be solved without obscurity;
- there are fewer advantages the system has over its attackers;

So, in general it would be advantageous to mix both the open and closed design approaches.

In the second half of the talk, Dr. Callas outlined the following

- Intellectual property exists to benefit society and it is understandable that intellectual property owners want more protection and it is also understandable that society is skeptical of the desires and claims of these owners.
- The anti-circumvention provisons of the Digital Millenium Copyright Act considers circumventing a technological measure that effectively controls access to a protected information as a felony. However, encryption research, computer repair, reverse engineering and security testing are exempted from this provison. The anti-circumvention provisions have the following problems: (1) exceptions are defenses against anti-circumvention and not limitations. (2) Exceptions are tortuous. So this raises the issue of imbalance between the rights and responsibilities of makers and breakers.
- Making and breaking the security of a system is like a dance. If breaking is punished then makers become lax and also there will be no incentive to design good security.

Finally, Dr. Callas concluded his talk by suggesting the following solutions

- tie circumvention to infringement
- leave the security research alone

2.2 Distribution Chain Security, Glenn Durfee, Stanford University

Digital content distribution systems will enable business models in the near future that cannot be predicted today. In this paper, Dr. Durfee discussed a security problem that arises from the conflicting privacy and integrity goals of middlemen in digital distribution chains. In this context, he presented a cryptographic solution based on commitment schemes and zero-knowledge proofs of arithmetic relations.

A distribution chain is basically the resale or redistribution of a work. In a distribution chain, whenever there is a new sale or redistribution of a work, a new contract is attached to the existing contract. This new contract must maintain the old obligations and also not bypass the old security restrictions.

A naive approach to this problem is to either append the new terms to the old contract or encapsulate the old contract with a new contract. The problem with this approach is that the contract sizes will blow up and also the end user will also see the effect of every middleman.

Dr. Durfee then presented an approach that uses obfuscated contracts and verifies fairness using these obfuscated contracts. Dr. Durfee also presented some experimental results that demonstrated the practicality of his approach.

2.3 Copy-Protection Policies for Digital Images, Gregory L. Heileman and Carlos E. Pizano, University of New Mexico and Elisar Software Corporation

The growth of the internet has led to widespread availability of information and also has made it easy to reproduce information. So, it has become increasingly important to protect electronic information. In this talk, Dr. Heileman discussed the problem of protecting on-line image content. In this context, he surveyed the existing technologies like digital watermarking (both visible and invisible), client-side copy disabling (both at the device driver and browser levels), and secure image transfer protocols.

In the initial part of the talk, Dr. Heileman addressed the ease of use and the level of security provided by each of above mentioned technologies along with the assumptions implicit in each approach. In addition, Dr. Heileman also discussed the use of the above mentioned technologies for copyright protection and ownership verification of digital images.

In the later part of the talk, Dr. Heileman presented a widely used model for image-related commerce, where the image content must be made globally viewable to all who access the Internet. Then, he addressed the difficulties in this model in managing the intellectual copyrights and also the trade-offs associated with using particular combinations of the aforementioned technologies to enforce image copyright and usage terms under this model.

2.4 Image Hashing, Mariusz Jakubowski and Ramarathnam Venkatesan, Microsoft Research

Dr. Jakubowski presented new algorithms for hashing, or one-way image compression, and comparison of bitmapped images. The methods of Jakubowski and Venkatesan are based on random multiscale subdivision of images into regions, randomized rounding of intensity averages in those regions, and robust compression of the resulting vectors by error correction. As hashes useful for robust identification and comparison of images, the compressed vectors can replace watermarks. The Jakubowski-Venkatesan schemes work with images subjected to common distortions, including scanning, resizing, and resampling. Additionally, image hashes withstand anti-watermark transformations performed by software such as StirMark and unZign.

2.5 A Proven Secure Tracing Algorithm for the Optimal KD Traitor Tracing Scheme, Kaoru Kurosawa, Tokyo Institute of Technology, Mike Burmester, Royal Holloway, Yvo Desmedt, University of Florida

Dr. Desmedt presented a proven secure black box tracing algorithm for the Kurosawa-Desmedt one-time traceability scheme of Eurocrypt'98. It traces not only the traitors who use the Stinson-Wei/Boneh-Franklin attack but also any traitors. His result implies that the lower bounds of Kurosawa and Desmedt are tight and the scheme is optimum.

2.6 Defense against Man-in-the Middle in Client Server Systems with Secure Servers, Richard Lipton, Princeton University and Dimitrios Serpanos, University of Crete

The explosion in network services coupled with dramatic progress in network connectivity has increased the attack by computer viruses on digital content. So, there is a growing need for services that provide protection of Intellectual property.

In the initial part of the talk, Dr. Serpanos outlined the following :

- Conventional model for representing virus and anti-virus
- Structure of viruses
- Classification of viruses
- different methods for identifying viruses

During the later half of the talk, Dr. Serpanos presented a hardware-based solution for protecting digital content from viruess. In this solution, the applications are assumed to execute in a safe environment and the devices attached to the client behave as spies and have the the following characteristics :

- detect I/O activity
- have processor and memory for crypto computations
- simple and are of low cost
- are passive I/O devices

2.7 Summary of the National Research Council's "Digital Dilemma" Report: Findings, Non-findings, and Implications for the Technical Agenda Plus Open Questions Session, Clifford Lynch, Coalition for Networked Information, Joan Feigenbaum, AT&T Labs

The National Research Council under the advice of Federal Network Commitee initiated a study to address the issue of Intellectual rights in the context of emerging infra-structure. Based on this study, the National Research Council released a report entitled "The Digital Dilemma: Intellectual Property in the Information Age," that took a wide-ranging look at copyright in the digital environment. Both Dr. Lynch and Dr. Feigenbaum served on the committee that authored the report. Among other things, the report looks carefully at areas in which technology interacts with, and potentially restructures, markets for and uses of copyrighted works. Close attention was paid to both the potential pluses and the potential minuses of "technological protection mechanisms" for digital works.

In the first part of the talk, Dr. Lynch and Dr. Feigenbaum gave an overview of the findings of the report, including reccomendations for public policy, legislative action (or the lack thereof), and research. These form an important context for future research and development in protection mechanisms and other relevant technologies.

The key findings and recommendations of that report are as follows :

- Traditionally issues of copyrights have been handled by large corporations and attorneys. With the changes in distribution of information, intellectual property is getting into everyday lives of every person.
- Historically, the public and private spaces were different and were easy to distinguish one from the other. Now, the distinction between public and private spaces is not clear and also we have technologies that allow breach of ownership and private space.
- The impact of copying on business and everyday lives is not well understood.
- Historically, large corporations were involved in the licensing of products and services. Now, almost every segment of society in some form or the other is involved in the licensing of products and services. This raises a lot of concern due to unequal participants.
- Historically, intellectual rights was a legal issue. Now, it is a complex legal, business, psychological and social issue.
- Currently, we are in a volatile environment. In this situation, we do not need a slew of legislations whose implications are not well understood.
- There are lot of problems with the The Digital Millenium Copyright Act. The anti-circumvention part of 'The Digital Millenium Copyright Act' has been widely criticized since it it gives incentives for content owners to have a bad technological protection system.

- It is difficult to characterize fair use of digital content. There is a trade-off between what is considered a right use and what is considered as an acceptable defense.
- The issues of copyright legislation make it difficult for common man to follow law. So who takes the responsibility for educating the common man and how should this be done ?
- There are a lot of barriers in archiving cultural and historic information and this raises a lot of concerns.

The following were some of the issues that was not addressed in the report:

- The committee felt that intellectual rights covers a lot of complex areas. So, they wanted to focus on some specific areas. As a result of this, there was no discussion of global issues involved in protecting intellectual property except for software piracy.
- There is no discussion on trademarks and patents.

In the second part of the talk, there was an open question session. In this session the following issues / questions were raised :

- In the context of secure systems, how does one address the issues related to plaguarism ?
- Sometimes circumvention helps in protecting one's right. So when and how do we draw the boundary between what is right and what is defendable ?
- In the current context, how do we define publishing ?
- What is the distinction between copyrights and trademarks ?
- Theft of software can be as bad as theft of physical property. So, what type of legal and technical protection would be required to prevent software theft ?

2.8 Intellectual Property in the Information Age: Will Laws and Technology Destroy Public Libraries?, Barbara Simons, ACM

A few years ago Hollywood and the music industry discovered the Internet and realized, much to their horror, that the technology now exists to make arbitrary numbers of perfect copies of a digitized object. As a result, we have seen an explosion of legislative and treaty proposals. For example, the Digital Millenium Copyright Act, passed in 1998, attempts to outlaw devices and technologies that can be used to bypass copyright controls. This legislation has several bad features, among them the unintended side effect of making some legitimate computer security research illegal. It could even criminalize some techniques, such as reverse engineering, that were required to correct Y2K problems.

Both the legislation that is passed and the manner in which technology is implemented will have a major impact on the rights and responsibilities of creators and users of intellectual property. How will copyright be impacted? Will new laws and new technologies that protect intellectual property eliminate user rights of fair use and first sale? Are we moving from copyright protection of books and magazines on the net to contract law, and if so, what are the potential repercussions? The manner in which these questions are resolved will have a significant impact on our society.