Digital Watermarking
History, Challenges and the Future

Information Technologies for IPR Protection
Outline

• History
  – The first 50 years of electronic watermarking (Feb. 2002)

• Challenges
  – Why watermarking is nonsense? (Sep. 2002)
    • Questions and comments
  – WaCha ’06

• Future
  – Multimedia security: the 22 century approach (Nov. 2005)
History
Challenges
Why Watermarking is Nonsense

- Watermarking
  - Insertion of a mark in an object $x$ to form a marked object $x_m$ and the use of a detector $p()$ to determine whether or not a target object $y$ is a valuable version of $x_m$
  - Goal of watermarking
    - To contrive a method of marking and a detector such that
      - $p(y)=1$ if and only if $y$ is a valuable version of $x_m$
• The idea case
  – Let $V$ be the set of all $y$ that are valuable versions of $x_m$
  – Let $P$ be the set of all $y$ for which $p(y)=1$ (protected objects)
    • The two sets shall be very nearly coincident
    • They can never be precisely coincident…since
      – $p(x)=0$
Why Watermarking is Nonsense (cont.)

- **Existing Watermarking Schemes**
  - Contriving a marking method and a detector such that
    - \( p(y) = 1 \) if and only if \( y \) is close to \( x_m \)
  - The detector makes its decision by thresholding some real-valued function of \( y \),
    - e.g. \( d(y, x_m) \)
      - Euclidean distance
      - Weighted distance based on perception

- **An attacker must either**
  - Advance a large distance or
  - Determine the secret direction \( x - x_m \) and advances a small distance
• An obvious flaw in this approach
  – It assumes that the set $V$ (all valuable variations) lie in this single neighborhood
  – However, it is not clear that $V$ even forms a connected set
  – There exist valuable versions of $x_m$ that are unlikely to be close to it in any perceptually weighted metric
These variations are unlikely to be close to $x_m$ in any perceptually weighted metric.
Why Watermarking is Nonsense (cont.)

- Expanding $P$ until it encompasses the neighbors that need to be protected is not possible
  - This will include many objects not related to $x_m$
Why Watermarking is Nonsense (cont.)

- **P** shall be a union of non-connected neighbors
- **However, two problem arise**
  - **Complexity**
    - Assume a suite of **M** transformations, each of which can take **K** states exist, and every one of them maps \( x_m \) to a valuable version
    - If they can be cascaded, \( M^K \) neighborhoods the detector must protect
  - **Uncertainty**
    - Listing all possible transformations is not possible
    - No value of **M** can we ever be confident that this is secure
Why Watermarking is Nonsense (cont.)

- Disallowing certain transformations?
  - Giving up the hard problem in favor of a tractable but irrelevant one
- It is challenging to find a context in which a scheme that would not withstand left to right flips or shirt color changes would be useful
- “close to” $\neq$ “valuable version of”
- This fact compromises essentially all efforts so far
Four thousand holes in Blackburn, Lancashire
And though the holes were rather small
They had to count them all
Now they know how many holes
it takes to fill the Albert Hall

- “A Day in the Life, “ The Beatles
Comments from Pierre Moulin

• Herley’s results are obtained based on inaccuracies, omissions, ambiguities and assumptions. In particular:
  – the notion of “valuable set” is ill-defined
  – The claims about the detector’s computational complexity are unsubstantiated
  – The critical role of cryptographic keys is ignored
  – The complexity of the attacker’s problem is ignored
  – The required level of security in many applications is more benign than in standard cryptography applications
Valuable Sets

• What are the valuable sets worthy of protection?
  – Defined by the system designer, not by the attacker
  – System designers don’t need to announce what the valuable sets are
  – Application dependent
    • Inadmissible transformations may be reasonable for certain applications
• There is often no need to perform a search in the first place, and even if one is needed, there generally exist efficient search methods
  – Many schemes compute a single test statistic and do not require estimation of the unknown channel
  – Computationally feasible search may be sufficient
    • A search over geometrical operation can be organized as a tree
• Security cannot be achieved without cryptographic keys

$k$: a length $n_k$ binary key
$x$: the original object in object space $X$
$x_{mk}$: $F(x, k)$ the watermarked version of $x$ using $k$
$y$: the test image available to the watermark detector
$p(y, k)$: the binary value returned by the watermark detector
$V(x_{mk})$: the valuable set of $x_{mk}$
$P_k = \{y: p(y, k) = 1\}$: the acceptance region of the detector
$P_k^c = \{y: p(y, k) = 0\}$: the rejection region of the detector

$$P_D = 2^{-n_k} \sum_k \int d x p(x) \int_{y \in P_k \cap V(x_{mk})} p(y | x_{mk}) d y$$  \hspace{1cm} (1)$$

$P_k$ covers $V(x_{mk})$ densely

$$P_F = 2^{-n_k} \sum_k \int_{P_k} p(x) d x.$$  \hspace{1cm} (2)$$

$P_k \ll X$
$p_k$: cross-hatched areas
$V(x_{mk})$: shaded areas
• Blind detector does not use $x_m$, non-blind detector use $x$ to re-synchronize
• I won’t turn my TV set 90° to hack a video, e.g., in the context for DVD protection
• About the number of transformations $M$
  – for certain applications $M$ is small (e.g. $M=0$ for medical images),
  – The signal can be rendered transform invariant
  – For applications where $M$ is too big, using watermarking is nonsense
• Watermarking Virtual Lab (WAVILA)
  - ECRYPT Network of Excellence in Cryptography
    • Initiative launched on 1st February 2004
    • To intensify the collaboration of European researchers in information security, and more in particular in cryptology and digital watermarking
  - To bring watermarking and perceptual hashing to a higher degree of maturity such that when they are applied in "real" applications there is a clear idea of what these technologies can and, maybe equally important, cannot do.

<table>
<thead>
<tr>
<th>WVL1</th>
<th>Theoretical analysis of watermarking security</th>
<th>UVIGO</th>
</tr>
</thead>
<tbody>
<tr>
<td>WVL2</td>
<td>Development of practical systems</td>
<td>Univ. of Geneva</td>
</tr>
<tr>
<td>WVL3</td>
<td>Benchmarking</td>
<td></td>
</tr>
<tr>
<td>WVL4</td>
<td>Perceptual hashing</td>
<td></td>
</tr>
</tbody>
</table>
WaCha ‘05 and WaCha ‘06

• WaCha ’05
  – Effectiveness of Geometrical Attack
  – Cryptographically Secure Fingerprinting

• WaCha ’06
  – Is knowledge of watermarking algorithm useful for watermark removal?
  – How does the output of the optimal watermarking algorithm look like?
• Blind Newton sensitivity attack (BNSA)
  – An attack without any knowledge of the watermarking algorithm, but the access to the detector to return YES or NO
  – It has been shown to be able to attack basically all watermarking algorithm
  – For test images in BOWS contest, 56 dB is achieved
  – It can also guess the key and the embed image
• BOWS (Breaking Our Watermarking System)
  - Organized by University of Siena, Italy
  - Algorithm provided by I. Cox and M. Miller
  - Two phase
  - Remove the watermark from the watermarked images and retains a PSNR over 30dB

• BOWS Contest II
  - Organized by University of Geneva, Switzerland
  - Can you estimate the secret keys?
  - To gain more insight in robustness, sensitivity and information leakage

• Conclusion remarks
  - Knowledge of the digital watermarking algorithms result in shorter time in attack
• **Challenge 2**
  - Focus on benchmarking
    - Capacity, transparency and robustness
    - Peaks
      - Naïve steganography
      - Secure steganography
      - Watermarking
  - Benchmarking is useful but not for the very end product
  - A standard database of test sets shall be created
  - The algorithm will eventually be broken, choose the one with minimum cost (?)
  - Digital watermarking still has a long way to go
Future
• In the US, a considerable increase in funding in all security-related areas, including data hiding
• Interest has shifted from the core technology to the application
  – E.g. broadcast monitoring companies
  – Digimarc vs. RSA
• Watermarking has been oversold
  – None of the methods was robust enough for copyright protection
• Unrealistic expectations are heaped on watermarking by media companies and academics
  – Watermarking is economically valuable under realistic assumptions
Declining Interest in Watermarking

• Watermarking is too young
  – Time to do decent research and development to answer many open problems
  – Security-related technology is notoriously slow to get adopted
• The size of watermarking market is limited
• The attention should not be concentrated on the first application of data hiding - security applications (IPR and DRM-related)
  – Strong need of the content provider
  – Underestimating the security aspect
Declining Interest in Watermarking

• Discovering new applications beyond the circle of security
  - Multimedia transmission error concealment
  - A good tool for transitioning from one standard to another, especially when not all infrastructure of a particular standard can be easily upgraded
• Don’t be obsessed with commercial applications…==“
• Against the “why watermarking is non-sense paper”
  – The “valuable set” is ambiguously defined
  – How can an cryptographer be sure that an unforeseen attack cannot emerge?
    – Who cares if we cannot achieve absolute security?
      • Computational security can be achieved but serious efforts must be made

• Watermarking is not cryptography
  – The fact that a technology can be circumvented does not mean customers will circumvent it
Non-Security Oriented Applications of Watermarking

• For copyright protection and tracing applications, it can be viewed as a tool to increase the threshold of illegal actions

• Data hiding without malicious adversary seems to be ready for adoption
  – Watermarking vs. an ugly barcode

• Upgrade of legacy channel
  – E.g. from analog TV to digital TV
Non-Security Oriented Applications of Watermarking

• Metadata delivery for standards do not include means for such information exchange
• Error concealment