Von Bogenschützen und digitalen Wasserzeichen

Andreas Westfeld

Technische Universität Dresden
Institut für Systemarchitektur
mailto:westfeld@inf.tu-dresden.de

Zusammenfassung: Wasserzeichensysteme werden unter anderem verwendet, um digitale Medien zu schützen. Entfernt man beispielsweise aus einem raubkopierten Film das Wasserzeichen, soll er unbrauchbar werden. Wie robust eine eigens entwickelte Technologie namens „Broken Arrows“ tatsächlich ist, testet das EU-Expertennetzwerk ECRYPT derzeit im weltweit ausgeschriebenen zweiten BOWS-Wettbewerb. Es wird ein allgemeines Angriffsprinzip vorgestellt, das zum Sieg verhalf.

1 Der Wettbewerb

Es ist schwierig, digitale Wasserzeichensysteme zu evaluieren, da die Angriffe zum Zeitpunkt der Evaluation selten vorhersehbar sind. Das EU-Expertennetzwerk ECRYPT wählte daher einen „Monte-Carlo-Ansatz“, um mehr über die möglichen Angriffe auf digitale Wasserzeichensysteme zu erfahren. Der zweite BOWS-Wettbewerb („Brich unser Wasserzeichensystem“) fordert in drei Episoden Wissenschaftler aus aller Welt (in Anlehnung an den Wettbewerbsnamen liebevoll BOWmen, also Bogenschützen genannt) heraus, ein eigens für diesen Wettbewerb entwickeltes Wasserzeichensystem „Broken Arrows“ zu brechen [ECR07]. Dabei soll die gemessene Qualität des Mediums möglichst unbeeinträchtigt bleiben. In der ersten Episode, die von Mitte Juli bis Mitte Oktober 2007 lief, gelang dies nur zehn der weltweit 450 registrierten Teilnehmer. Mitglieder der TU Dresden, die als einzige deutsche Institution erfolgreich vertreten war, erreichten die Plätze 1, 3 und 10.

Aufgabe war es, aus drei Graustufenbildern (Landschaftsbilder mit 512×512 Bildpunkten, hohem Kontrast und einem großen Anteil an hohen Videofrequenzen, vgl. Abbildung 1) die unsichtbaren Wasserzeichen zu entfernen.

2 Andeutungen und Inspiration

Die markierten Bilder des Wettkampfs liegen im PGM-Format vor (portable greymap). Im Kopf der Bilddateien befindet sich der Kommentar „Generated by libit“. Wasserzeichen werden selten direkt in die Werte der Bildpunkte geschrieben. Um größere Robustheit zu erreichen, werden sie in einem transformierten Wertebereich aufmoduliert. In den meisten Fällen handelt es sich um den DCT- oder den Wavelet-Bereich (Scott Craver veröffentlich-

Abbildung 1: Die Bilder der ersten Episode des Wettkampfs: „Herbst“, „Schaf“ und „Erinnerung“

te dieses Rechercheergebnis während eines Vortrags auf der SPIE 2007 [CY07]). Nachdem schon im ersten BOWS-Wettbewerb der DCT-Bereich verwendet wurde [ECR06, Wes06], sollte im zweiten der Wavelet-Bereich an der Reihe sein. Die im Kommentar erwähnte libit, eine Bibliothek mit Funktionalität zur Signalverarbeitung, unterstützt genau eine zweidimensionale Transformation, die Wavelet-Transformation. Natürlich könnte es sich bei dem Kommentar um eine Finte handeln, jedoch sprechen die erzielten Ergebnisse eher dagegen.

Die Wavelet-Transformation halbiert die Auflösung des Bildes und gibt die dabei verlorengegangene Information in je einem Bild mit horizontalen (LH), vertikalen (HL) und diagonalen Waveletkoeffizienten (HH) wieder, welche als Subbänder bezeichnet werden. Die Transformation kann rekursiv auf das Bild mit der reduzierten Auflösung angewendet werden (vgl. Abbildung 2). Wenn wir die Koeffizienten der ersten und zweiten Zerlegungsebene auf null setzen, kann der Online-Detektor das Wasserzeichen nicht mehr erkennen.

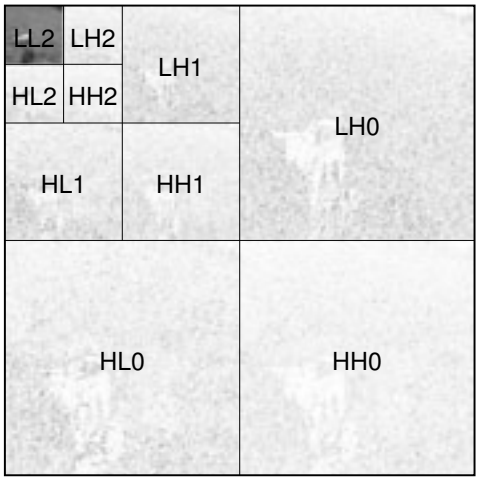

Abbildung 2: Waveletkoeffizienten für „Schaf“

Durch das Ausnullen wird das Bild unschärfer. Für einen erfolgreichen Angriff ist allerdings eine Bildqualität von mindestens 20 dB gefordert, welche zwar für „Herbst“ (20,11 dB) und „Schaf“ (21,41 dB) erreicht wird, nicht jedoch für „Erinnerung“ (18,61 dB). Diese Entdeckung hat offenbar auch Dominik Engel aus Salzburg gemacht, der sich am 9. September 2007 in der Ruhmeshalle verewigte mit 20,11 dB, 21,41 dB und 20,01 dB. Man kann bestimmte Angriffe also am erreichten Qualitätsnivau erkennen.

Gleichzeitig wird deutlich, dass das dritte Bild schwieriger zu brechen ist. Am 30. August 2007 trugen Jihane Bennour und Jean-Luc Dugelay sich mit dem bis dahin höchsten Wert für das dritte Bild in die Liste ein: 21,12 dB. Vom ersten BOWS-Wettbewerb ist bekannt, dass Jihane Bennour und Kollegen [BDM07] den Selbstähnlichkeitsangriff anwendeten [RDCD02], vgl. Abbildung 3. Allerdings ist die darin verwendete Bilddatenbank etwas unhandlich und schränkt die Leistungsfähigkeit willkürlich ein.

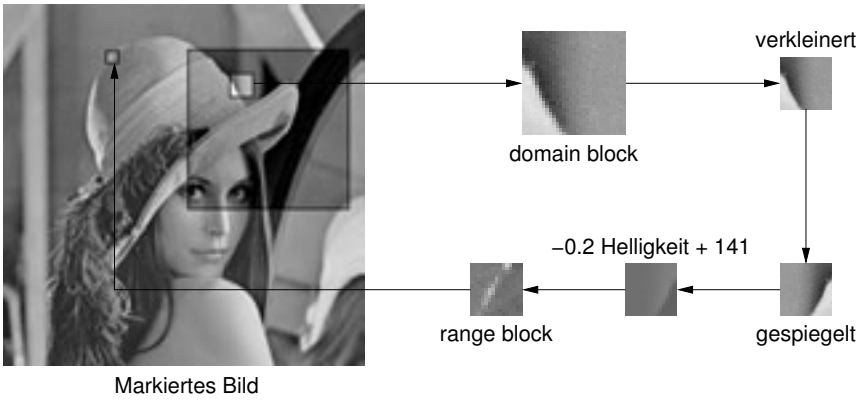

Abbildung 3: Selbstähnlichkeitsangriff nach [RDCD02]

3 Modellbasierter Angriff auf Wasserzeichen im Waveletbereich

Der vom Autor entwickelte Angriff lehnt sich insofern an den Selbstähnlichkeitsangriff [RDCD02] an, als dass Werte im Bild durch andere Werte aus der näheren Umgebung nachgebildet werden. Da örtlich dicht beieinanderliegende Werte in Bildern stark voneinander abhängen, die Elemente des Wasserzeichens jedoch nicht, lässt sich das Bild durch eine Schätzung aus der Umgebung nahezu erhalten, während das Wasserzeichen vollständig entfernt wird (vgl. Abbildung 4). Die lokalen Abhängigkeiten zwischen den Koeffizienten des Bildes werden für alle Subbänder außer LL2 jeweils in einem Regressionsmodell zusammengefasst und alle Koeffizienten mit dem dadurch bestimmten linearen Prädiktor abgeschätzt. Abbildung 4 zeigt die Koeffizienten für die Schätzung im Subband LH2. Jeder Koeffizient in LH2 wird geschätzt aus

- seinen unmittelbaren Nachbarn in LH2,

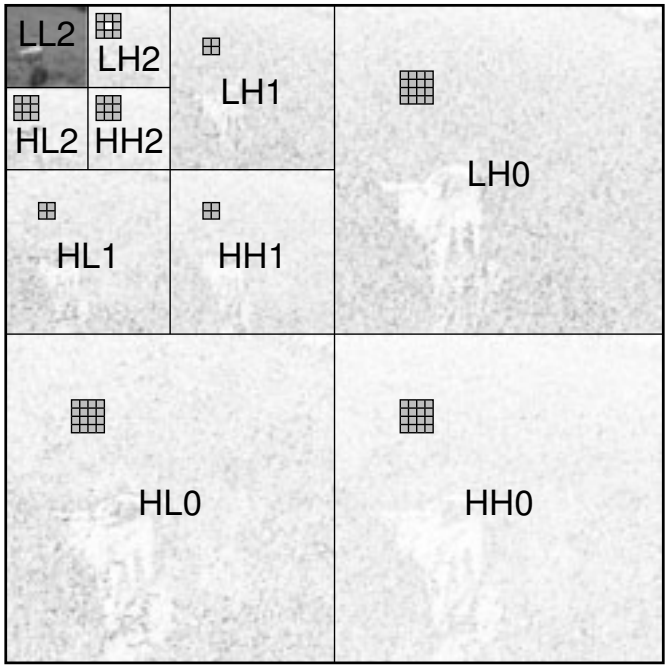

Abbildung 4: Modell für die Umgebungsschätzung der Wavelet-Koeffizienten in LH2

- seiner Entsprechung in den anderen Subbändern und deren unmittelbaren Nachbarn,
- seinen übergeordneten Entsprechungen der ersten und zweiten Zerlegungsebene (4 bzw. 16 je Subband).

Das Gleichungssystem für ein Subband in der dritten Zerlegungsebene hat also 4096 Gleichungen (für jeden Koeffizienten eine) mit 86 Unbekannten. Ein Gleichungssystem für ein Subband in der zweiten Zerlegungsebene hat 16384 Gleichungen mit 38 Unbekannten und für ein Subband in der zweiten Zerlegungsebene 65536 Gleichungen mit 26 Unbekannten. Die Qualität lässt sich noch erheblich steigern, indem bei der Regression und Schätzung nur die *Beträge* verwendet werden, das *Vorzeichen* jedoch vom Original übernommen wird. Tabelle 1 zeigt die Ergebnisse für den Angriff mit dem geschätzten Bild. Wesentlich bei diesem Angriff ist, dass er ohne Interaktion mit dem Detektor auskommt, d. h. die Schätzung kann für ein beliebiges mit „Broken Arrows“ markiertes Bild ausgeführt werden, danach ist das Wasserzeichen unlesbar.

Der Detektor erkennt das Wasserzeichen auch dann nicht, wenn der Angriff abgeschwächt wird – dadurch lässt sich die Qualität um ein reichliches Dezibel verbessern.

$$\text{abgeschwächtes Bild} = p \cdot \text{angegriffenes Bild} + (1-p) \cdot \text{markiertes Original}$$

Durch Intervallhalbierung lässt sich der optimale Parameter $p = 0 \cdots 1$ mit etwa 10 Detektoranfragen finden. Das Ergebnis nach dieser Abschwächung ist ebenfalls in Tabelle 1

Tabelle 1: Qualität (PSNR) für die Angriffe vor und nach Reduktion

	Mittelwert	Herbst	Schaf	Erinnerung
Nach Schätzung	22,87 dB	22,82 dB	24,60 dB	21,67 dB
Nach Abschwächung	24,30 dB	24,90 dB	25,45 dB	22,97 dB

verzeichnet. Die Schätzung wurde auch sowohl auf zwei Zerlegungsebenen eingeschränkt als auch auf vier ausgedehnt. Nach Abschwächung war die Qualität des angegriffenen Bildes für die Schätzung in drei Zerlegungsebenen am besten, so dass die Vermutung naheliegt, dass sich das Wasserzeichen in den ersten drei Zerlegungsebenen befindet.

4 Ausblick

Während der ersten Episode waren täglich nur 30 Detektoraufrufe je Teilnehmer gestattet. Damit sind Sensitivitätsangriffe, die das Wasserzeichen durch eine Vielzahl von Aufrufen abschätzen ausgeschlossen. Sensitivitätsangriffe sind Gegenstand der zweiten Episode. Während der dritten Episode ist die Zahl der Detektoraufrufe erneut beschränkt. In ihr soll aus einer großen Anzahl von markierten Bildern die Information des Wasserzeichens ermittelt und für einen Angriff ausgenutzt werden.

Literatur

[BDM07] Jihane Bennour, Jean-Luc Dugelay und Frederico Matta. Watermarking Attack (BOWS Contest). In Edward J. Delp III und Ping Wah Wong (Hrsg.), *Security, Steganography and Watermarking of Multimedia Contents IX (Proc. of SPIE)*, S. 18–1–18–6, San Jose, CA, January 2007.

[CY07] Scott Craver und Jun Yu. Reverse-Engineering a Detector with False Alarms. In Edward J. Delp III und Ping Wah Wong (Hrsg.), *Security, Steganography and Watermarking of Multimedia Contents IX (Proc. of SPIE)*, S. 0C–1–0C–10, San Jose, CA, January 2007.

[ECR06] ECRYPT. BOWS, Break Our Watermarking System, 2006. Online verfügbar unter `http://lci.det.unifi.it/BOWS`.

[ECR07] ECRYPT. BOWS-2, Break Our Watermarking System, 2nd Edition, 2007. Online verfügbar unter `http://bows2.gipsa-lab.inpg.fr`.

[RDCD02] Christian Rey, Gwenaël Doërr, Gabriella Csurka und Jean-Luc Dugelay. Toward Generic Image Dewatermarking? In *IEEE International Conference on Image Processing ICIP 2002*, Bd. 2, S. 633–636, New York, NY, USA, September 2002.

[Wes06] Andreas Westfeld. Lessons from the BOWS Contest. In *Proc. of ACM Multimedia and Security Workshop 2006, MM&Sec06, Geneva, Switzerland*, S. 208–213, New York, September, 26–27 2006. ACM Press.