Szenario-basierte Testverfahren zur Zertifizierung von Wasserzeichen

Michael Arnold, Christoph Busch und Martin Schmucker

{arnold,busch,mschmuck}@igd.fhg.de

Abstract: Verschiedenste Algorithmen zum Markieren digitaler Daten wurden in den letzten Jahren veröffentlicht und entwickelt, um die Daten mit nicht wahrnehmbaren Kennzeichnern zu versehen. Unterschiede in den Algorithmen sind zum Teil nur marginal, oder sie unterscheiden sich in ihren Grundprinzipien.
Eine objektive Bewertung existierender Algorithmen ist schwierig und kostenintensiv. Deshalb ist diese Bewertung sowohl für kommerzielle als auch für wissenschaftliche Organisationen nicht durchführbar. Dadurch wird die Analyse neuer Ansätze verzögert. Ein automatisches Testverfahren ist somit nicht nur hilfreich, sondern auch notwendig, um objektive Vergleiche und Analysen zu ermöglichen, die für die Verbesserung und die Neuentwicklung eine wichtige Grundlage darstellen.
Wir beschreiben in diesem Beitrag die Unterschiede von Wasserzeichenalgorithmen und ihre Evaluierung am Beispiel verschiedener Angriffe. Desweiteren weisen wir auf die Notwendigkeit eines objektiven, nachvollziehbaren und allgemein anerkannten Testverfahrens[1] hin. Aufgrund der Vielzahl existierender Algorithmen und Angriffe argumentieren wir für die Notwendigkeit eines szenario-basierten Ansatzes.

1 Hintergrund

Der Begriff “digitale Wasserzeichen” ist mehrdeutig. Er beschreibt ein Forschungsgebiet, eine konkrete Methode oder ihre Implementierung und auch den eingebetteten Inhalt. In diesem Artikel befassen wir uns mit Testverfahren für die Implementierungen von Wasserzeichenverfahren. Um die Verwechslungsgefahr zu verringern, verwenden wir im folgendem die Begriffe “Verfahren” für die Testverfahren, “Algorithmus” und “Methode” für die Wasserzeichenverfahren, und “Wasserzeichen” für die eingebettete Information.
Wasserzeichenmethoden verstecken bestimmte Informationen in bestehenden Daten. Eine Abstraktion ist folglich, daß Wasserzeichenmethoden einen Übertragungskanal in einem bestehenden Medium erschließen. Anforderungen an diesen Übertragungskanal, wie z.B. seine Kapazität[2] oder seine Störungssicherheit[3] ergeben sich durch die konkrete Anwendung. Wasserzeichen können wahrnehmbar oder nicht wahrnehmbar sein.

[1] Wir gehen davon aus, daß ein allgemein anerkanntes Testverfahren ein Mindestmaß an Praxisrelevanz aufweisen muß, welche folglich auch ein Mindestmaß an Tauglichkeit mit einschließt.
[2] Die Kapazität beschreibt die Menge an Information, die versteckt werden kann.
[3] Anstelle des Begriffes “Störungssicherheit” verwendet man bei Wasserzeichen den Begriff “Robustheit”.

Methoden zur Einbettung nicht wahrnehmbarer Wasserzeichen gehören zu der Klasse der steganographischen Methoden, die Inhalte in anderen Medien verstecken. Sie sind mit der Suche nach einer Nadel im Heuhaufen vergleichbar. Allerdings ist im Falle von steganographischen Methoden die Nadel in kleine Eisenspäne aufgeteilt und nur der legitime Nutzer hat Kenntnis über die Lage der Eisenspäne und den entsprechenden Bauplan. Das Ziel ihrer Entwicklung liegt hauptsächlich in der unbemerkten Kommunikation. Dies bedeutet für digitale Daten, dass keine Analyse eine Aussage über das Vorhandensein eines möglicherweise eingebetteten Inhaltes treffen kann. Im Gegensatz dazu kann in Szenarien, in denen Wasserzeichenmethoden genutzt werden, die Kenntnis über eine Kommunikation - also die Existenz einer nicht wahrnehmbaren Information - vorhanden sein. Diese Kenntnis hat Auswirkungen auf die Robustheit der eingebetteten Information, z.B. sollte sie unautorisiert nicht entfernbar sein.

Durch die Anforderung der Nichtwahrnehmbarkeit müssen Wasserzeichenmethoden individuell für die jeweiligen Datentypen entwickelt werden. Der Schwerpunkt bisheriger Entwicklungen befaßte sich mit Audio, Bilder und Video. Gegenwärtige Forschung umfaßt auch weitere Medientypen wie geometrische Modelle, Musiknoten oder natürliche Sprache.

Wie schon oben angedeutet, sind Anforderungen an Wasserzeichenmethoden durch die jeweiligen Anwendungsszenarien definiert. Diese werden in Abschnitt 2 erläutert. In Abschnitt 3 beschreiben wir Angriffe auf Wasserzeichenmethoden. Die verschiedenen Prinzipien von Wasserzeichenalgorithmen werden in Abschnitt 4 vorgestellt. Anforderungen an Testverfahren werden in 5 erläutert. Eine Zusammenfassung und ein Ausblick in Abschnitt 6 schließen diesen Artikel.

2 Anwendungsszenarien für Wasserzeichenalgorithmen

Anforderungen an Wasserzeichenalgorithmen werden durch den jeweiligen Anwendungsfall bestimmt. Die verschiedenen individuellen Anwendungen lassen sich zu sogenannten Anwendungsszenarien zusammenfassen (siehe [CMB02] und [ASW03]). Diese schließen ein:

- Im *Urheberrechtsschutz* dienen Wasserzeichen als Nachweis des Urhebers.
- *Transaktionswasserzeichen*[4] speichern Informationen über durchgeführte Transaktionen.
- Beim *Monitoring* wird ein Übertragungskanal (typischerweise ein Radio- oder Fernsehsender) auf die Existenz eines Wasserzeichens geprüft. Dabei kann z.B. festgestellt werden, welches Audio-, Fernseh- oder Werbestück gerade gesendet wird.
- *Authentisierung* und *Integritätsschutz* digitaler Daten ist ebenfalls mit digitalen Wasserzeichen möglich.

[4] Transaktionswasserzeichen werden auch “active fingerprintings” genannt

- *Metadaten* oder allgemein das Bereitstellen eines Übertragungskanals in digitalen Daten wird genutzt, um Metainformationen mit dem Inhalt direkt zu verknüpfen.

Neben der Nichtwahrnehmbarkeit und der Robustheit ergeben sich noch weitere spezifische Anforderungen. Für das Monitoring ist die Echtzeitfähigkeit des Wasserzeichendetektors zwingend. Eine sehr hohe Kapazität ist besonders für das Einbetten von Metadaten wichtig. Hingegen ist die "Zerbrechlichkeit" für den Integritätsschutz von enormer Bedeutung: Manipulationen werden in dieser Anwendung durch die Veränderung des Wasserzeichens identifiziert.

3 Angriffe auf Wasserzeichenalgorithmen

Wie im vorigen Abschnitt beschrieben, gibt es verschiedene Anwendungsmöglichkeiten für Wasserzeichenalgorithmen. Daraus resultieren (Sicherheits-) Anforderungen, wie z.B. die Robustheit der eingebetteten Wasserzeichen. Angriffe auf die Sicherheit von Wasserzeichenalgorithmen können einerseits bewußt erfolgen, mit der Absicht ein eingebettetes Wasserzeichen zu entfernen. Anderseits können aber auch Verarbeitungsoperationen ein Wasserzeichen unbeabsichtigt beeinflussen.

Angriffe haben immer einen Kontextbezug. Innerhalb eines bestimmten Anwendungsszenarios treten manche Verarbeitungs- oder Angriffsoperationen mit höherer Wahrscheinlichkeit auf. Dabei beschränkt sich die Sicherheit eines Wasserzeichens aber nicht nur auf dessen Lesbarkeit. Folgende Konsequenzen eines Angriffs können identifiziert werden (siehe [ASW03]):

- Wasserzeichen können nicht detektiert werden.
- Falsche Wasserzeichen werden detektiert.
- Wasserzeichen werden ohne Autorisierung detektiert.

3.1 Angriffe die eine Detektion fehlschlagen lassen

Angriffe können auf verschiedenen Arten die Detektion von Wasserzeichen erschweren oder verhindern:

- durch das Entfernen oder Manipulieren von eingebetteten Wasserzeichen.
- durch eine Desynchronisierung des Wasserzeichendetektors und des Wasserzeichenträgers.

Das Entfernen eines Wasserzeichens ist die naheliegendste Form eines Angriffs. Dies kann auch unbewußt passieren. Zu den möglichen Angriffen gehören:

- *Signalverarbeitungsoperationen*, wie z.B. Komprimierung und Filtern von Audio- oder Bilddateien, sind alltägliche Verarbeitungsoperationen.
- Bei *algorithmusspezifischen Angriffen* werden Informationen über das Prinzip des verwendeten Wasserzeichenalgorithmus genutzt, um eingebettete Wasserzeichen zu entfernen. Diese Angriffe wurden z.B. von Craver et al. [CWL+01] durchgeführt.
- *Verschwörungsangriffen* mißbrauchen verschieden markierte Daten eines Originals, um ein Wasserzeichen zu entfernen, ähnlich dem arithmetischen Mittel. Eine entsprechend komplementäre Methode erlaubt das Schätzen eines Wasserzeichens bei verschiedenen Originalen, die mit dem gleichen Wasserzeichen versehen wurden.
- Angriffe unter Verwendung eines *Orakels* nutzen das Ergebnis eines Wasserzeichendetektors, um dessen Entscheidungsgrenze in einem mehrdimensionalen Raum zu bestimmen (vgl. [LvD98] und [KLvD98]. Dieses Wissen wird dann für einen Angriff genutzt.

Eine subtilere Art des Angriffs ist die Desynchronisierung des Wasserzeichenträgers und des Wasserzeichendetektors. Dabei ist der Wasserzeichendetektor nicht mehr in der Lage, ein vorhandenes Wasserzeichen zu lesen.

- Bestimmte globale und lokale *Transformationen*, wie Warping der Zeitachse bei Audio [CWL+01] oder Warping von Bildern [KP99], zerstören das Wasserzeichen nicht, führen aber häufig dazu, daß der Detektor das Wasserzeichen nicht mehr finden kann.
- "*Scrambling*" Angriffe vertauschen Teile des Originals, die sich von der Wahrnehmung her sehr ähnlich sind [PK02].

3.2 Angriffe die falsche Wasserzeichen detektieren lassen

Diese Angriffe simulieren die Existenz eines ursprünglich nicht eingebetteten Wasserzeichens.

- Bei dem "*Copy*"-Angriff [KVH00] wird ein Wasserzeichen in markierten Daten, wie z.B. Audio oder Bild, geschätzt und auf andere Daten übertragen.
- Als *Overwatermarking* bezeichnet man das wiederholte Markieren von Daten. Leidet die Qualität des Ergebnis nicht drastisch unter einem Overwatermarking, so kann ein weiteres Wasserzeichen eingebettet werden, das ursprünglich nicht vorhanden war.
- Der "*Deadlock-Angriff*"[CMYY96] versucht, ein Wasserzeichen in markierte Daten "hinein zu interpretieren", um sich damit ein Original zu erzeugen.

3.3 Angriffe die eine unerlaubte Detektion ermöglichen

- "*False alarm*" Angriffe führen zu der Detektion eines Wasserzeichens, obwohl diese Detektion nicht autorisiert ist. Dies kann z.B. bei Zugriff auf einen Detektor durch das Finden eines geheimen Schlüssels erfolgen. Bei diesem Angriff ist die Dimension des Schlüsselraumes und die Trennbarkeit zwischen markierten und nicht markierten Daten von entscheidender Bedeutung.

4 Prinzipien von Wasserzeichenalgorithmen

Neben der Vielzahl der unterschiedlichen anwendungsspezifischen Angriffen gibt es auch eine Vielzahl an existierenden Wasserzeichenalgorithmen. Unterschiede können gegeben sein durch die Einbettungsart, Detektionsmetriken, fehlerkorrigierende Codes, Wahrnehmungsmodelle, uvm.

Ein Anhaltspunkt über die Vielfalt ist durch die Anzahl der Publikationen auf diesem Gebiet gegeben. Neben den verschiedenen Forschungsalgorithmen existieren auch eine Reihe kommerzieller Derivate. Im folgenden geben wir einen kurzen Überblick der Charakteristika existierender Verfahren für Audio- und Bildaten.

Wasserzeichenalgorithmen übertragen ein schmalbandiges Signal (das Wasserzeichen) in einem breitbandigen Kanal (dem Wasserzeichenträger). Sie sind daher eng mit den Spreizband-Übertragungstechniken[5] verwandt. Diese Übertragung ist in Abbildung 1 dargestellt.

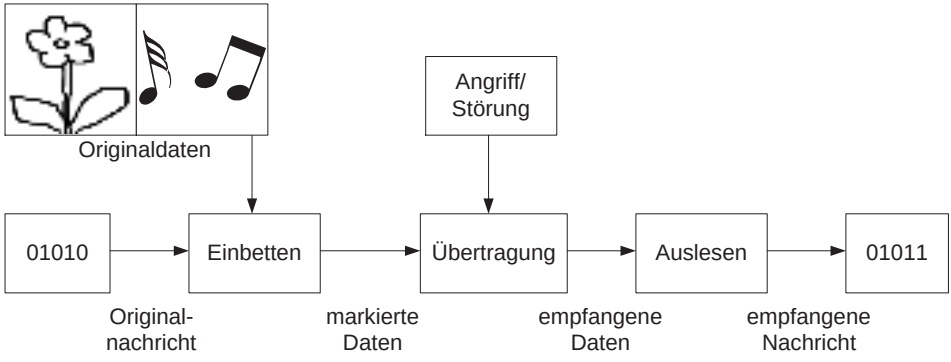

Abbildung 1: Schema des Einbettens und Auslesens eines Wasserzeichens in ein Bild oder in eine Audiodatei. Unabsichtliche oder gezielte Angriffe, wie z.B. Nachbearbeiten des Bildes, können die Lesbarkeit des Wasserzeichens beeinträchtigen. In diesem Fall ist die empfangene Nachricht verschieden von der Originalnachricht.

Die entwickelten Algorithmen können z.B. durch das zugrunde liegende Einbettungsver-

[5] engl.: "spread spectrum communication"

fahren unterschieden werden:

Das *Einbetten im Zeit- oder Ortsraum* ist dadurch gekennzeichnet, daß die nötigen Änderungen zum Einbetten des Wasserzeichens direkt am Datenmaterial vorgenommen werden. Dabei findet keine Konvertierung in einen Transformationsraum (wie z.B. in den Frequenzraum) statt.

Im Gegensatz zum Einbetten im Orts- oder Zeitraum wird beim *Einbetten im Frequenzraum* das Datenmaterial in den Frequenzraum transformiert, um dann die Frequenzinformationen zu modifizieren. Zu den am häufigsten genutzten Transformationen gehören:

- die diskrete Kosinus-Transformation (DCT) (wie z.B. [BKZ98]) und die diskrete Wavelet-Transformation (DWT) (wie z.B. [KH98]) für Bilder
- die diskrete Fourier-Transformation (DFT) für Audio (wie z.B. [Arn00])

Für Bilder sind die Eigenschaften der DCT und Modifikation der DCT-Koeffizienten aufgrund ihrer Verwendung bei der JPEG-Komprimierung bekannt. Dabei ist die Wavelet-Transformation näher an die menschliche Wahrnehmung von Bildsignalen angelehnt als die blockorientierte DCT.

Weitere Transformationen, wie z.B. die Fourier-Mellin-Transformation oder die Cepstrum-Transformation, werden aufgrund ihrer spezifischen Eigenschaften auch für das Einbetten von Informationen genutzt. Daneben gibt es noch weitere Algorithmen, die sich an die Verfahren der Bildkomprimierung (fraktale Komprimierung) oder die Bild- und Signalverarbeitung (Detektion und Verändern von Merkmalen) anlehnen.

5 Testverfahren für Wasserzeichenalgorithmen

Obwohl eine umfassende Beschreibung existierender Wasserzeichenalgorithmen und ihrer Angriffe den Rahmen dieses Beitrages sprengen würde, zeichnet dieser kurze Einblick schon die Schwierigkeit bei der Bestimmung der Eigenschaften von Wasserzeichenalgorithmen ab: Nur umfangreiche Tests charakterisieren ein System vollständig. Dies ist allerdings sehr zeitaufwendig und wird daher, wenn überhaupt, sehr selten durchgeführt. Unter anderem können folgende Ursachen identifiziert werden:

- Die Bestimmung des Anwendungsszenarios einschliesslich der damit verbundenen Operationen und Parametern ist sehr komplex und alle Kombinationen können aus Komplexitätsgründen nicht beachtet werden.
- Die Tests müssen unter realitätsnahen Bedingungen erfolgen und die Ergebnisse miteinander vergleichbar sein. Dies bewirkt eine weitere Erhöhung gesamten Testaufwands.
- Die verwendeten Testdaten (Bilder, Audio, Video, ...) müssen repräsentativ sein sein. Sie müssen identifiziert werden und in einer Datenbank gepflegt werden, um auf einer breiten Datenbasis reproduzierbare Testresultate zu erhalten.

Durch die oben beschriebene Vielfältigkeit der Entwicklungen wird deutlich, warum Technologieentwickler, -anbieter und -benutzer einen zuverlässigen und allgemein anerkannten Vergleich – ein Testverfahren – von Algorithmen benötigen.

Da ein Entwicklungskriterium digitaler Wasserzeichen die Nichtwahrnehmbarkeit ist, ist ein wichtiger Bestandteil eines solchen Testverfahrens die Qualität des resultierenden Datenmaterials. Diese kann anhand subjektiver und objektiver Tests bestimmt werden. Allerdings stimmen objektive Qualitätsmetriken besonders für Bilder nicht mit der subjektiv wahrgenommen Qualität überein. Deshalb ist ein vollständig automatisiertes Testverfahren von Algorithmen insbesondere für Bilder und Video schwierig.

Existierende Testverfahren können als Orientierung für zukünftige Wasserzeichentestverfahren dienen:

- Zu den bekannten Testverfahren für die Leistungsfähigkeit von Prozessoren gehören die von SPEC[6] [SPE] entwickelten Benchmarks. Diese geben Auskunft über die anwendungsspezifische Leistung eines Prozessors.
- Die Leistungsfähigkeit graphischer Koprozessoren wird durch intensive Tests von kommerziellen Anbietern, wie z.B. 3D Mark [Dag03], bis hin zu der Bestimmung von Bildwiederholungsraten bestimmter Spiele- oder Demonstrationssoftware bestimmt.
- Für Virenscanner bietet EICAR [EIC] eine Testdatei an, die real existierende Viren "simuliert" und die es ermöglicht, die Leistungsfähigkeit von Virenscanner zu evaluieren.

Für Wasserzeichentestverfahren bietet sich ein SPEC-orientierter Ansatz an: Anforderungen an Wasserzeichenalgorithmen sollten aus praxisnahen Testszenarien abgeleitet sein. Aus Komplexitätsgründen können diese Testverfahren nicht alle Details beinhalten. Dennoch sollten die Testszenarien genügend Aussagekraft besitzen, um Anwendern aufgrund ihrer speziellen Sicherheitsanforderungen Abschätzungen der Leistungsfähigkeit von Wasserzeichenalgorithmen entsprechend ihren Anwendungsszenarien zu ermöglichen.

Ein wünschenswertes Resultat einer Evaluierung ist eine Zertifizierung ähnlich einer Zertifizierung im Sinne der "Common Criteria" (CC, [CC]): eine nachweisbare Bescheinigung des Sicherheitsnivaus des getesteten Systems ("system under test", SUT). Der Bereich der digitalen Wasserzeichen ist jedoch noch ein relativ junger Forschungsbereich, so daß eine CC-ähnliche Zertifizierung (noch) nicht angebracht erscheint - ganz abgesehen von den finanziellen Restriktionen, die sich durch den Testaufwand ergeben.

5.1 Parameter

In einem Testverfahren lassen sich folgende vier Parameter-Klassen identifizieren, die die Variablen eines Testverfahrens kategorisieren (siehe Abbildung 2):

[6] SPEC ist im wesentlichen ein Zusammenschluss verschiedener Hardwarehersteller.

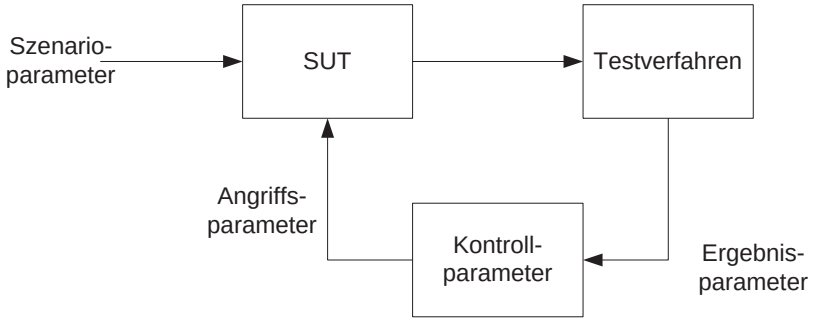

Abbildung 2: Die Variablen eines Testverfahrens können in Szenario-, Angriffs-, Kontroll- und Ergebnisparameter unterteilt werden. Die Ergebnisparameter haben Einfluß auf die Kontrollparameter. Ein Beispiel hierfür ist die variierende Einbettungsstärke in Abhängigkeit des Detektionsergebnisses.

- Zu den *Szenarioparameter* zählen Variablen, wie der Umfang einer Nachricht, das Ergebnis der Detektion (Wasserzeichen vorhanden oder Inhalt des Wasserzeichens) oder "blind detection" (Detektion ohne Verwendung des Originals). Der Umfang einer Nachricht ist z.B. ein anwendungsspezifischer Parameter wobei das Ergebnis einer Detektion vom System vorgegeben ist.
- Zu den *Kontrollparameter* zur Steuerung des Einbettungsverfahrens gehört zum Beispiel der zulässige Qualitätsverlust. Ein Kontrollparameter bei der Detektion ist ein Schwellwert, der über das Vorhandensein eines Wasserzeichens entscheidet. Diese Kontrollparameter sind typischerweise durch das Anwendungsszenario festgelegt.
- *Angriffsparameter* definieren die verwendeten Angriffe (und ihre Kombination). Diese sind durch das Anwendungsszenario bestimmt. Beispiele hierfür sind Komprimierung oder oben beschriebene Angriffe.
- *Ergebnisparameter* werden iterativ in das Testsystem eingespeist (Backpropagation), um z.B. Abhängigkeiten bestimmen zu können.

5.2 Anforderungen und existierende Testverfahren

Ein automatisches Testverfahren ("*benchmark system*" oder "*benchmark*") sollte folgende Kriterien erfüllen, damit seine Ergebnisse aussagekräftig sind:

- Die Unabhängigkeit der Ergebnisse muß gewährleistet sein. Dies impliziert die Entwicklung und Überwachung des Testverfahrens durch oder im Auftrag einer unabhängigen Organisation. Als Alternative kann die Beteiligung (fast) aller betroffenen Gruppen in Betracht bezogen werden.

- Die Ergebnisse eines Testverfahrens müssen nachvollziehbar sein. Eine wiederholte Testdurchführung darf nicht zu anderen Ergebnissen führen, was eine Protokollierung impliziert.
- Testszenarien müssen realen Anwendungsszenarien entsprechen. Das Problem hierbei ist, eine möglichst große Abstraktion aus Gründen der Komplexität zu erreichen, die es zuläßt, für den konkreten Anwendungsfall aussagekräftige Werte abzuleiten.
- Ein Benchmark muß erweiterbar sein: Neue Algorithmen und Angriffe sollten leicht integrierbar sein, um aktuelle Entwicklungen sowohl von Seiten der Wasserzeichenentwickler als auch von Seiten der Angreifer berücksichtigen zu können.
- Typische szenario-spezifische Parameter, wie Datenmaterial oder Nachrichtenlänge, müssen in festgelegten Standardszenarien definiert werden, um Anwendern allein durch die Veröffentlichung der szenario-spezifischen Ergebnisse ein Vergleich von Algorithmen ohne die Druchführung eines Tests zu ermöglichen.

Existierende Benchmarksysteme, wie z.B. StirMark [KP99], CheckMark [PVM+01] oder OptiMark [STN+01], erfüllen diese Anforderungen nur teilweise. Der Hauptkritikpunkt an den existierenden Testverfahren ist die Tatsache, daß sie nicht allgemein akzeptiert sind. D.h. sie beschreiben ein Wasserzeichensystem nicht verifizierbar und attestieren ihnen keine bestimmten Charakteristiken. Desweiteren ist das Datenmaterial nicht festgelegt. Sie entsprechen somit nicht der Anforderung einer allgemein anerkannten Zertifizierung.

5.3 Zertifizierung von Wasserzeichenalgorithmen am Beispiel von Certimark

In dem Projekt *Certimark* (siehe [Cer]) wurden die oben beschriebenen Nachteile existierender Benchmarksysteme erkannt und es wurde versucht, diese zu lösen: Im Gegensatz zu den oben beschriebenen Testverfahren war der Grundgedanke einen zertifizierten Benchmark für Wasserzeichentechnologien zu entwickeln. Certimark ist die Abkürzung für "Certification for Watermarking Techniques". In diesem von der Europäischen Union geförderten Konsortial-Projekt waren fünfzehn Partner beteiligt. Das Spektrum reichte von Universitäten, über Forschungsinstitute und KMUs[7], bis hin zu großen Firmen. Diese Gestaltung des Konsortiums und die Anzahl der beteiligten Entwickler bündelte nicht nur das wissenschaftliche Know-how auf diesem Gebiet sondern erzielte eine gewisse Unabhängigkeit des Benchmark-Verfahrens durch sein gemeinsames Design und seine verteilte Entwicklung. Die Architektur des entwickelten Benchmarks ist in Abbildung 3 dargestellt.

Die Funktionalität des entwickelten Testverfahrens beinhaltet die Verwaltung der Testdaten und der parametersierten Einbettungsprozesse und Angriffe. Der jeweils zu testende Wasserzeichenalgorithmus (System Under Test - SUT) kann dabei einer einzelnen Operation oder auch einer Sequenz von Angriffen ausgesetzt werden. Für die Bewertung der

[7] kleinere und mittlere Unternehmen

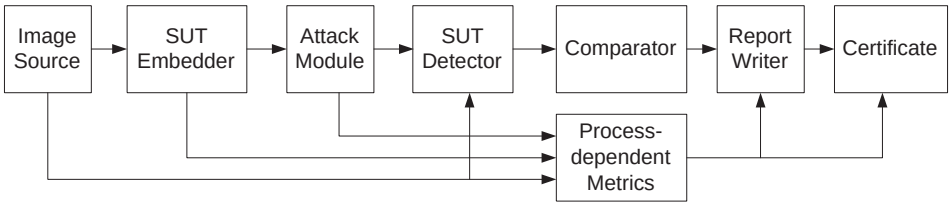

Abbildung 3: Die Architektur des Certimark-Testverfahrens.

Systemsicherheit eines Algorithmus ist es maßgeblich, ob nach den Angriffen die eingebettete Information noch ausgelesen bzw. detektiert werden kann. Dies wird im Certimark Benchmark im Comparator-Modul evaluiert.

5.4 Zertifizierung

Wie zuvor beschrieben, war die Zielsetzung des dargestellten Certimark-Systems ein vertrauenswürdiges, praxisrelevantes und offenes Testverfahren. Ein wichtiger Punkt ist die objektive und reproduzierbare Bewertung von Wasserzeichenalgorithmen.

Das Ergebnis der Bewertung wird in einem Zertifikat festgehalten. Dieses Zertifikat beinhaltet eine Beschreibung des Algorithmus (in Form eines Hashes) und der Testszenarien unter denen der Algorithmus getestet wurde, den Testzeitpunkt und die Testergebnisse.

Wichtig hierbei ist, daß die Testergebnisse das erfolgreiche Überstehen gegen Angriffe dokumentieren. Sie attestieren nicht eine absolute Sicherheit für bestimmte Anwendungsszenarien. Nur dadurch kann eine zukünftige Weiterentwicklung von Angriffen berücksichtigt werden: Das entsprechende Szenario muß durch neue Angriffe erweitert werden. Dieser Sachverhalt drückt sich dann z.B. durch eine Szenarioversionsnummer in dem Zertifikat aus.

Ein weiteres Kriterium ist die Unabhängigkeit der Entwickler und des durchführenden Testpersonals. Da die Durchführung des Benchmarks etwa bei anerkannten CC-Prüfstellen schon aus wirtschaftlichen Gründen nicht realisierbar war wurde im Rahmen des Competence Centers for Applied Security Technology (CAST) ein technisches Komitee für die Zertifizierung von Digitalen Wasserzeichen gegründet (CAST TCW, [CAS]). Dieses Komitee wurde von den Partnern des Certimark-Projektes initiiert. Weitere Vertreter des Themenfeldes können in diesem offenen Komitee mitwirken. Die Durchführung von Technologie-Evaluierungen kann auf diese Weise kostengünstig und mit der gebotenen Objektivität durch Mitarbeiter des Komitees durchgeführt werden.

6 Zusammenfassung und Ausblick

Dieser Beitrag stellt das Spektrum verschiedener Algorithmen zum Markieren von Daten und Angriffe dar. Ein objektiver nachvollziehbarer Vergleich existierender Wasserzeichenalgorithmen ist folglich ohne ein automatisches Testverfahren praktisch nur unter hohen personellem Zeitaufwand durchzuführen. Die Forschung benötigt Testverfahren, um intensive Tests durchzuführen, Schwachstellen zu identifizieren und bestehende Algorithmen zu verbessern. Anwender von Wasserzeichenalgorithmen benötigen Testverfahren als Hilfestellung bei der Bewertung von Algorithmen.

In diesem Zusammenhang sind Reproduzierbarkeit und Objektivität zwingend, was eine unabhängige Institution für die Entwicklung und die Durchführung eines Testverfahrens erfordert.

Das zugrundeliegende Testverfahren muß verschiedene Szenarien abbilden, und sowohl die Parameter zum Einbetten und das Datenmaterial also auch relevante Angriffe beinhalten.

Das Ergebnis eines anerkannten Testverfahrens ist ein Zertifikat, das einem Prüfsiegel entspricht und die Details des Tests, die Testergebnisse und den getesteten Wasserzeichenalgorithmus festhalten.

Bei der Erstellung zukünftiger Testverfahren wäre ein Zusammenschluß verschiedener Anbieter von Algorithmen und von bestimmten Anwendergruppen wünschenswert (ähnlich SPEC). Nur dadurch kann ein anerkanntes und aussagekräftiges Testverfahren geschaffen werden. Für zukünftige Entwicklungen stellt sich die Frage, in wie weit eine Standardisierung entsprechend DIN oder ISO von Nutzen wäre und die Entwicklung neuer Algorithmen davon profitieren kann.

Literatur

[Arn00] Michael Arnold. Audio Watermarking: Features, Applications and Algorithms. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2000)*, pages 1013–1016, New York, NY, USA, July 2000. IEEE Press.

[ASW03] Michael Arnold, Martin Schmucker, and Stephen Wolthusen. *Techniques and Applications of Digital Watermarking and Content Protection*. Artech House, Boston, MA, USA, 2003.

[BKZ98] Scott Burgett, Eckhard Koch, and Jian Zhao. Copyright Labeling of Digitized Image Data. *IEEE Communications Magazine*, 36(3):94–100, March 1998.

[CAS] CAST-TCW-Homepage, http://www.cast-forum.de/activities/tcw/overview .

[CC] CommonCriteria, http://www.commoncriteria.org.

[Cer] Certimark-Homepage, http://www.certimark.org.

[CMB02] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom. *Digital Watermarking*. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2002.

[CMYY96] Scott Craver, Nasir Memon, Boon-Lock Yeo, and Minerva Yeung. Can Invisible Watermarks Resolve Rightful Ownerships? Technical Report 20509, IBM Research Divison, Yorktown Heights, NJ, USA, July 1996.

[CWL+01] Scott A. Craver, Min Wu, Bede Liu, Adam Stubblefield, Ben Swartzlander, Dan S. Wallach, Drew Dean, and Edward W. Felten. Reading Between the Lines: Lessons from the SDMI Challenge. In *Proceedings of the 10th USENIX Security Symposium*, Washington D.C., USA, August 2001.

[Dag03] Maneesh Daghat. 3DMark03 - Next Generation 3D Benchmarking. Technical report, Future Mark Corporation, 2003.

[EIC] EICAR, http://www/eicar.org.

[KH98] Deepa Kundur and Dimitrios Hatzinakos. Digital Watermarking using Multiresolution Wavelet Decomposition. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2969–2972, Seattle, WA, USA, May 1998. IEEE Press.

[KLvD98] Ton Kalker, Jean-Paul M. G. Linnartz, and Marten van Dijk. Watermark Estimation through Detector Analysis. In *Proceedings of the International Conference on Image Processing*, pages 425–429. IEEE Press, October 1998.

[KP99] Martin Kutter and Fabien A. P. Petitcolas. A Fair Benchmark for Image Watermarking Systems. In Ping Wah Wong and Edward J. Delp, editors, *Proceedings of Electronic Imaging 1999, Security and Watermarking of Multimedia Contents*, pages 226–239, San Jose, CA, USA, January 1999. SPIE.

[KVH00] Martin Kutter, Sviatoshlav Voloshynovskiy, and Alexander Herrigel. The Watermark Copy Attack. In Ping Wah Wong and Edward J. Delp, editors, *Proceedings of Electronic Imaging 2000, Security and Watermarking of Multimedia Contents II*, pages 371–381. SPIE, January 2000.

[LvD98] Jean-Paul M. G. Linnartz and Marten van Dijk. Analysis of the Sensitivity Attack against Electronic Watermarks in Images. In David Aucsmith, editor, *Information Hiding: Second International Workshop*, volume 1525 of *Lecture Notes in Computer Science*, pages 258–272, Portland, OR, USA, 1998. Springer-Verlag.

[PK02] Fabien A. P. Petitcolas and Darko Kirovski. The Blind Pattern Matching Attack on Watermark Systems. In *Proceedings 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3740–3743, Orlando, FL, USA, May 2002. IEEE Press.

[PVM+01] Shelby Pereira, Sviatoslav Voloshynovskiy, Maribel Madueño, Stéphan Marchand-Maillet, and Thierry Pun. Second Generation Benchmarking and Application Oriented evaluation. In Ira S. Moskowitz, editor, *Information Hiding: 4th International Workshop*, volume 2137 of *Lecture Notes in Computer Science*, pages 340–353, Pittsburgh, PA, USA, October 2001. Springer-Verlag.

[SPE] SPEC - standard performance evaluation corporation, http://www.spec.org.

[STN+01] V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, and I.Pitas. A Benchmarking Protocol for Watermarking Methods. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'01)*, pages 1023–1026, Thessaloniki, Greece, October 2001. IEEE Press.