

FORBILD

FORENSISCHE BILDERKENNUNG FINDET ILLEGALES MATERIAL

Ermittler bei Polizei und Staatsanwaltschaft müssen oft auf der Suche nach illegalem Bildmaterial große Mengen von Daten durchsuchen. Um die Massen von Bildern möglichst schnell und effizient automatisch durchkämmen zu können, benutzen sie Methoden und Werkzeuge der IT-Forensik. Das Fraunhofer SIT hat mit dem Projekt ForBild diese Methoden erweitert und präzisiert: Über sogenannte robuste Hashs können Bilder schnell und sicher identifiziert werden. Dieses ergänzende Verfahren hat das Fraunhofer SIT getestet und optimiert.

Polizeibehörden untersuchen große Datensammlungen oft mit IT-Werkzeugen, um illegales Bildmaterial aufzuspüren. Meist handelt es sich hier um kinderpornographische Bilder. Derzeit erkennt eine automatisierte Suche kinderpornographisches Bildmaterial nur dann zuverlässig, wenn eine identische Kopie dieses Bildes bereits in einer Datenbank hinterlegt ist. Über einen kryptographischen Vergleichswert, einen sogenannten Hash, kann die Software diese Kopie automatisch zuordnen und identifizieren.

Die Schwäche dieses Verfahrens ist, dass selbst minimale Veränderungen am Bild den Hashwert verändern. Wenn ein Nutzer ein Bild in ein anderes Datenformat umwandelt oder auch nur in einem Bildbetrachter-Programm öffnet und speichert, ist der Hash soweit



© Getty Images/Stockdisc

*Fraunhofer-Institut für Sichere
Informationstechnologie SIT*

*Kontakt:
Martin Steinebach
Rheinstraße 75
64295 Darmstadt*

*Telefon 06151 869-349
Fax 06151 869-224
martin.steinebach@sit.fraunhofer.de
www.sit.fraunhofer.de*

verändert, dass eine Identifizierung über dieses automatische Verfahren nicht mehr möglich ist. Das bedeutet, dass Polizeibehörden derzeit mit dieser Technik nicht in der Lage sind, bewusst verschleiertes illegales Bildmaterial aufzuspüren. Die Ermittler müssten also zusätzlich selbst die Datensammlungen sichten, um sicherzugehen, dass tatsächlich alle illegalen Bilder gefunden werden.

Eine Alternative beziehungsweise Ergänzung zu diesen etablierten Mechanismen sind sogenannte robuste Hashs. Diese Technik nutzt nicht die Datei-Eigenschaften zur Bilderkennung, sondern orientiert sich an der menschlichen Wahrnehmung: Wenn ein Bild für ein menschliches Auge identisch erscheint, ist auch der Vergleichswert identisch. Damit ignoriert das Verfahren bei der Identifikation einer Bilddatei mögliche Veränderungen der Größe, des Rauschfaktors oder des Dateiformats und konzentriert sich auf optische Übereinstimmungen. Dieser Mechanismus nennt sich robuster Hash, in Anlehnung an die kryptographischen Funktionen. Um einen robusten Hash zu erstellen, skaliert man ein Bild herunter auf 16 x 16 Pixel und wandelt es in 256 Graustufen um.



*Umwandlung des Originalbildes
in ein Graustufenbild von
16 x 16 Pixel*

Aus dieser Liste von Graustufen wird der Median ermittelt, also der Wert, der genau in der Mitte der von hell nach dunkel sortierten Graustufen liegt. Die Graustufen werden jetzt in Bits umgewandelt, dafür wird jedem Grau ein binärer Wert zugeordnet: Ein Grau unterhalb des Medians bekommt eine 0, eine Graustufe über dem Median eine 1. So werden die 256 Graustufen in 256 Bitwerte umgewandelt. Sie beschreiben das Bild ausreichend genau und sind robust gegenüber Skalierungen, Umspeicherungen und ähnlichen Veränderungen der Datei.



Umwandlung des Graustufenbilds in einen robusten Hash

Im Rahmen des Projekts ForBild – forensische Bilderkennung – hat das Fraunhofer SIT gemeinsam mit LSK Data Systems GmbH und der Technischen Universität Darmstadt diese robusten Hashverfahren getestet und einfache robuste Hashfunktionen so optimiert, dass sie als Ergänzung zu den üblichen kryptographischen Hashverfahren eingesetzt werden können.

Die robusten Hashs für die zu untersuchenden Bilder müssen sehr schnell zu berechnen sein, damit große Mengen von Bildmaterial möglichst zügig untersucht werden können. Zudem sollten sie Bilder sehr sicher identifizieren können, um eine niedrige Fehlerquote

zu erzielen. Denn jeder Fehlalarm muss manuell überprüft werden, was die Ermittler zusätzlich belastet. Gleichzeitig darf die Erkennungsrate nicht so niedrig sein, dass das Werkzeug für die Bildersuche unbrauchbar wird.

Das robuste Hashverfahren hat der ForBild-Projektpartner LSK Data Systems in ein CD-Lesegerät integriert. Damit können Ermittler einen Stapel CDs mit Bildmaterial, das sie überprüfen wollen, in das Gerät eingeben und die Daten automatisch überprüfen lassen. Auf der Suche nach einschlägigem Material werden die Bearbeiter so zeitlich und psychisch entlastet. Das Fraunhofer SIT empfiehlt, dieses Verfahren als Ergänzung zu den herkömmlichen Hashs bei der Verfolgung von kinderpornographischem Material heranzuziehen.

Das Fraunhofer SIT plant eine Erweiterung des robusten Hashverfahrens, um andere Datenträger und andere technische Strategien miteinbeziehen zu können. Zudem sollen auch Videos IT-forensisch auf illegale Aufnahmen hin untersucht werden können.



Die robusten Hashs dieser Bilder unterscheiden sich in nur fünf Bits. Das Verfahren zeigt deshalb an, dass diese Bilder identisch sind.