Die Enkel von Locard

Roman Povalej, Heiko Rittelmeier, Johannes Fähndrich, Silvio Berner, Wilfried Honekamp & Dirk Labudde

Informatik Spektrum
Organ der Gesellschaft für Informatik e.V. und mit ihr assoziierter Organisationen

ISSN 0170-6012

Informatik Spektrum
DOI 10.1007/s00287-021-01393-5

Springer

Informatik Spektrum
https://doi.org/10.1007/s00287-021-01393-5

HAUPTBEITRAG

Die Enkel von Locard

Analyse digitaler Spuren in der forensischen Informatik

Roman Povalej[1] · Heiko Rittelmeier[2] · Johannes Fähndrich[3] · Silvio Berner[4] · Wilfried Honekamp[5] · Dirk Labudde[6]

Angenommen: 10. Juli 2021

Zusammenfassung
Die seit Jahrhunderten verwendeten Methoden in der Forensik basieren auf der Annahme eines Austausches von Materie und Mustern. Durch die Digitalisierung sind diese Annahmen nur noch eingeschränkt gültig und werden hier erweitert und diskutiert. In dem Zusammenhang ist es erforderlich, den Spurenbegriff grundlegend zu überdenken. Gleichzeitig werfen der ständige technische Fortschritt und die immer größer werdende Flut von auszuwertenden Daten die Ermittlungsbehörden immer wieder zurück. Dieser Entwicklung ist nur durch Automatisierung Herr zu werden. Verfahren der Künstlichen Intelligenz können und werden die Ermittlungsbehörden zukünftig dabei zunehmend unterstützen.

Die Welt ist immer stärker vernetzt. Jeder hat (mindestens) ein digitales Gerät, mit dem er/sie im Internet surft, sich mit Mitmenschen austauscht, Daten hoch- und herunterlädt und vieles mehr. Insofern ist es nicht verwunderlich, dass auch der digitale Tatort immer stärker in den Vordergrund rückt. Die forensische Analyse digitaler Spuren trägt immer mehr zur Aufklärung von IT-Vorfällen oder gar Straftaten bei. Jeder Täter unterliegt auch in der digitalen Welt Locards Austauschprinzip. Das Zusammenspiel zwischen digitalem Tatort, digitaler Spuren, gerichtsverwertbarer Sicherung von Beweisen unter Zuhilfenahme von intelligenten Assistenzsystemen bei Ermittlungsarbeiten sowie der Nachvollziehbarkeit der erzeugten Ergebnisse ist für die Enkel Locards eine Herausforderung, sodass auch neue Wege des Denkens und Handelns zu beschreiten sind.

✉ Roman Povalej
roman.povalej@polizei.niedersachsen.de

Heiko Rittelmeier
heiko@rittelmeier.de

Johannes Fähndrich
johannesfaehndrich@hfpol-bw.de

Silvio Berner
silvio.berner@polizei.sachsen.de

Wilfried Honekamp
wilfried.honekamp@hochschule-stralsund.de

Dirk Labudde
labudde@hs-mittweida.de

[1] Polizeiakademie Niedersachsen, Nienburg (Weser), Deutschland

[2] Nüdlingen, Deutschland

[3] Hochschule für Polizei Baden-Württemberg, Villingen-Schwenningen, Deutschland

[4] Hochschule der sächsischen Polizei (FH), Rothenburg, Deutschland

[5] Hochschule Stralsund, Stralsund, Deutschland

[6] Hochschule Mittweida – FoSIL, Mittweida, Deutschland

Locard

Seit fast 100 Jahren bildet das Locard'sche Prinzip die Grundlage für die forensische Fallarbeit. Es beschreibt die materielle Übertragung von möglichen Spuren zwischen Tatbeteiligten [1, S. 139]:

> Überall dort, wo er geht, was er berührt, was er hinterlässt, auch unbewusst, all das dient als stummer Zeuge gegen ihn. Nicht nur seine Fingerabdrücke oder seine Fußabdrücke, auch seine Haare, die Fasern aus seiner Kleidung, das Glas, das er bricht, die Abdrücke der Werkzeuge, die er hinterlässt, die Kratzer, die er in die Farbe macht, das Blut oder Sperma, das er hinterlässt oder an sich trägt. All dies und mehr sind stumme

Published online: 03 September 2021

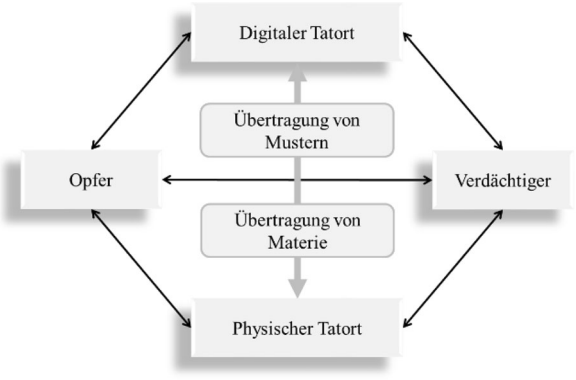


Abb. 1 Erweiterung von Locards Austauschprinzip unter der Annahme, dass der digitale und reale Tatort durch Spuren miteinander in Verbindung stehen. (eigene Abbildung, nach Dewald und Freiling [3])

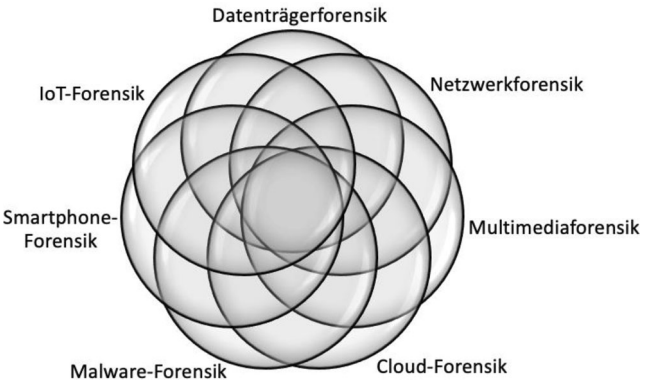


Abb. 2 Spezialgebiete der Forensik und deren resultierende Schnittmenge

> Zeugen gegen ihn. Dies ist der Beweis, der niemals vergisst. Er ist nicht verwirrt durch die Spannung des Augenblicks. Er ist nicht unkonzentriert, wie es die menschlichen Zeugen sind. Er ist ein sachlicher Beweis. Physikalische Beweismittel können nicht falsch sein, sie können sich selbst nicht verstellen, sie können nicht vollständig verschwinden. Nur menschliches Versagen, diese zu finden, zu studieren und zu verstehen, kann ihren Wert zunichte machen.

Die Wirkung (z. B. eines Tatwerkzeugs) ist in den Veränderungen an den Objekten messbar. Dies kann eine reine materielle Veränderung als auch ein auswertbares Muster sein. Der Übergang von der Übertragung von Materie auf die Übertragung von Mustern („transfer of traits“) (siehe [2]) ermöglicht eine formale Erweiterung dieses Prinzips (siehe auch Abb. 1):

1. Die *Übertragung von Materie („physical transfer“)* ist die Annahme, dass ein Objekt unter wohldefinierter Einwirkung eines anderen Objekts zerteilt wird. Dabei werden Einzelteile von einer Quelle auf ein Ziel übertragen. Bei physischer Einwirkung auf ein Rechensystem können z. B. Datenträger oder Teile davon am Tatort zurückbleiben.
2. Die *Übertragung von Mustern („transfer of traits“)* ist die Annahme, dass charakteristische Eigenschaften von einem Objekt auf ein anderes übertragen werden, ohne dass notwendigerweise Materie ausgetauscht wird. Die charakteristischen Eigenschaften lassen sich als Muster definieren.

Durch die Digitalisierung und die schnelle Einführung von neuen Technologien ergibt sich nicht nur die Frage nach der Allgemeingültigkeit dieses Prinzips, sondern eine zwingend notwendige Erweiterung auf die digitale Fallarbeit. Mittlerweile ist es wissenschaftlich unstrittig, dass auch in der digitalen Welt keine Interaktion ohne Spuren möglich ist. Experten – *die Enkel Locards* – müssen also wissen, in welchen Zusammenhängen welche Spuren entstehen, wie diese zu finden sind und wie lange diese beweiskräftig gesichert werden können. Ergänzend beschäftigt sich das noch relativ junge Fach der „Forensischen Informatik“ mit den folgenden Fragestellungen:

- Wie sichert man digitale Spuren so, dass sie während des Untersuchungsprozesses vor Veränderungen geschützt sind?
- Wie bereitet man die Spuren anschließend so auf, dass sie vor Gericht als Beweismittel dienen können?

Als Synonyme werden in der Literatur „Digitale Forensik“ und „IT-Forensik“ verwendet. Ziel dieses wissenschaftlichen Bereichs ist die Suche nach digitalen Spuren jeder Art, deren Sicherung und schließlich die Aufbereitung der Daten, sodass das Gericht am Ende zu einem sachgerechten Urteil kommen kann. Letztendlich ist das Ziel jeder forensischen Untersuchung die gerichtsverwertbare Präsentation der Ergebnisse einer wissenschaftlichen Untersuchung. Dem weiten Themenfeld [4] ist es geschuldet, dass es mittlerweile mehrere Spezialgebiete der digitalen Forensik gibt. Die wichtigsten sowie deren Überschneidungen untereinander sind der Abb. 2 zu entnehmen. Daneben gibt es noch etliche weitere Spezialisierungen, die an dieser Stelle jedoch keinen verallgemeinernden Charakter tragen. Darüber hinaus gibt es noch die Möglichkeit der Einteilung in Post-Mortem- und Live-Forensik sowie entsprechend der untersuchten Technologien [5].

Die Frage jedoch ist, ob eine solche Einteilung, die sich in den letzten Jahren etabliert hat, unter Berücksichtigung der technischen Entwicklung überhaupt noch sinnvoll ist. Sollte sich eine moderne Einteilung nicht vielmehr an den digitalen Spuren selbst orientieren, wobei die Quelle von untergeordnetem Interesse ist?

Locard digital: Tatort und Spuren

Sowohl in der analogen als auch in der digitalen Welt suchen Ermittelnde an einem Tatort nach Spuren und Beweisen, um dadurch den Tathergang rekonstruieren zu können, sodass ein vernünftiger Zweifel am Tatgeschehen nicht möglich ist (siehe auch [6, 7]). Nach Casey [8] basieren „digitale Spuren“ auf Daten, die in Computersystemen gespeichert oder übertragen wurden. Digitale Spuren, die durch ihre physikalische Eigenschaft, wie z. B. Magnetisierung oder Ladezustand, bestimmt werden können, werden als digitale Daten bezeichnet. Entsprechend werden digitale Spuren zuerst identifiziert, lokalisiert, extrahiert und gesichert, bevor sie in eine verständliche und interpretierbare Form gebracht werden können. Dabei müssen alle Spuren und Beweise gerichtsverwertbar gesichert werden.

Um die Bedeutung von digitalen Spuren während eines investigativen Prozesses zu verstehen, bedarf es einer eindeutigen Klärung des Begriffs „digitaler Tatort“ und der Verortung von „digitalen Spuren“. Folgendes Szenario kann als Illustration dienen und zeigt die Verschmelzung von analoger und digitaler Forensik:

Bei einem Beschuldigten wird ein PC sichergestellt. Der PC verfügt über weitere Peripheriegeräte (Monitor, Webcam, Lautsprecher, externe Festplatten u. ä.), welche ebenfalls der IT-Forensik übergeben werden. Auch wenn der PC als Träger digitaler Spuren gilt, kann dieser in der realen Welt verortet werden und ist Teil eines physischen Tatorts bzw. Ereignisortes. Parallel dazu befindet sich der PC im LAN oder im WLAN des Ereignisortes.

Auch wenn das Locard’sche Prinzip für digitale Spuren gilt, ist es erforderlich den PC einem konkreten Nutzer und dann einer verdächtigen Person zuzuordnen. Hier kann die Stärke der Authentifizierung (Login-Daten) auf dem PC bzw. auf anderen mobilen Endgeräten genutzt werden, um einen Nutzer einer konkreten Person (hier der verdächtigen Person) zuzuordnen. Ist diese Zuordnung erfolgt, können weitere mobile Endgeräte des Tatverdächtigen dem Ereignisort zugefügt werden. Der digitale Tatort ist durch die Verortung der Geräte Teil des realen (physischen) Tatorts geworden. Aus der Analyse der mobilen Endgeräte können zusätzlich noch Bewegungsdaten oder Verkehrsdaten abgeleitet werden. Durch die Verwendung aller Spuren kann eine Verbindung zwischen dem Opfer bzw. den Opfern und dem möglichen Tatverdächtigen hergestellt werden (siehe auch Abb. 3).

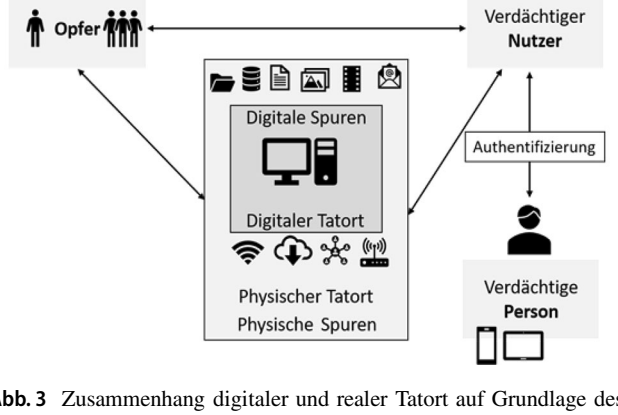


Abb. 3 Zusammenhang digitaler und realer Tatort auf Grundlage des Locard’schen Prinzips über die Verortung digitaler Geräte

Das Aufkommen an digitalen Spuren auf dem PC ist sehr heterogen, in Bezug sowohl auf die Formate als auch auf die zu verwendenden Werkzeuge. Diese reichen von Textdokumenten, Bildern, E-Mails, Videos über Log-Dateien des jeweiligen Betriebssystems bis hin zu Artefakten auf Ebene der genutzten Dateisysteme. Weitere Spuren können sich auf dem Router oder in der Cloud befinden. Analog zur klassischen Forensik braucht die digitale Forensik Werkzeuge und Vorgehensmodelle. Der PC des Verdächtigen ist ein realer Spurenträger, auf dem die digitalen Spuren als physische Spuren (Magnetisierung, dem Ladezustand oder elektromagnetische Wellen) enthalten sind. Erst durch geeignete Werkzeuge bzw. andere Anwendungen ist eine Interpretation von digitalen Spuren möglich. Mit Interpretation ist hier das Verstehen und die Einschätzung dieser Spuren gemeint [9].

Auch wenn digitale Spuren aus sehr unterschiedlichen Quellen stammen, haben sie wohldefinierte Eigenschaften und Gemeinsamkeiten. Auf der Grundlage der Eigenschaften ist es möglich, sie zu kategorisieren. Betrachtet man die Verwendung von IT-Systemen aus der Sicht von Nutzern so wird deutlich, dass Daten bei jedem Schritt und während Datenverarbeitungen anfallen. Somit ist es unumstritten, dass eine Nutzung von IT-Systemen ohne das Hinterlassen von Datenspuren nicht möglich ist. Spuren entstehen an verschiedenen Orten und können in lokale und nichtlokale Spuren klassifiziert werden. *Lokale Spuren* entstehen auf dem Gerät selbst. Somit erfolgt die Sicherung in der Regel von dem Gerät. Zur Klasse der Geräte zählen IT-Systeme oder Datenträger (unter anderem Computer, Smartphones, Digitalkameras bzw. USB-Sticks, SD-Karten, DVDs). Zur Klasse der lokalen Spuren gehören die Inhaltsdaten einer Datei (z. B. doc-/sqlite-Datei), Dateinamen, Windows Registry, Log-Dateien, temporäre Dateien und auch Backups. *Nichtlokale Spuren* entstehen nicht auf dem Gerät selbst. Sie entstehen an einem anderen Ort, etwa bei Diensten wie Webservern, Mailservern oder sozialen Netzwerken. Auch Daten in einer Cloud, bei Internet Service Providern oder Mobilfunkanbietern gehören dieser Klasse an. Beispiele sind die Inhaltsdaten bei Onlinediensten, Daten aus Überwachungskameras, des Mobilfunkanbieters (wie Standort- und Kommunikationsdaten) oder Verkehrsdaten bei Internet Service Providern.

Eine wichtige Eigenschaft von digitalen Spuren ist die *Flüchtigkeit*. Sie besagt, wie lange Daten unverändert erhal-

ten bleiben. Daraus ergeben sich besondere Aufgaben an die Prozesse in der digitalen Forensik. Die Prozesse müssen effektiv und schnell ablaufen, um einen Verlust von Spuren zu verhindern und Manipulationen zu vermeiden. Das Konzept der Flüchtigkeit enthält 3 Stufen: Persistente, semipersistente und flüchtige Spuren. *Persistente Spuren* bleiben über einen langen Zeitraum, auch ohne permanente Stromversorgung, auf dem Speichermedium erhalten. *Semipersistente Spuren* sind bei aktiver Stromversorgung über einen langen Zeitraum im System gespeichert. *Flüchtige Spuren* sind auch bei einer permanenten Stromversorgung nur kurzzeitig verfügbar. Die Einteilung nach Flüchtigkeit stellt den zeitlichen Rahmen und die Reihenfolge der Sicherung (Order of Volatility) von digitalen Spuren dar.

Elektronische Daten befinden sich in der Regel persistent auf Datenträgern. Die Daten werden für den wahlfreien Zugriff in Dateisystemen organisiert. Dateisysteme sind die häufigste Quelle digitaler Spuren. Wenn Daten gerade verarbeitet werden, müssen sie im flüchtigen Speicher gehalten werden. Zur Übermittlung über Netzwerke werden Daten in der Regel in kleine Pakete zerlegt und von einem Knotenpunkt zum nächsten transferiert, bis sie am Ziel ankommen.

Ein anderes Konzept der forensischen Informatik beschäftigt sich mit der Fragestellung der Vermeidbarkeit. Man unterscheiden nach technisch vermeidbaren („non-essential“) und technisch unvermeidbaren („essential“) digitalen Spuren. Unvermeidbare Spuren sind im Vergleich zu vermeidbaren Spuren für die Funktionsweise essenziell notwendig. Da eine Manipulation bei technisch unvermeidbaren digitalen Spuren mit einem höheren Aufwand verbunden ist, ist die Aussagekraft dieser Spuren auch höher als bei den technisch vermeidbaren Spuren. Dies machen die unvermeidbaren Spuren für den Forensiker besonders vertrauenswürdig und sie besitzen in der Folge eine höhere Beweiskraft [3].

Im forensischen Kontext ergeben sich aus den Eigenschaften von digitalen Spuren unerbittliche Anforderungen an Werkzeuge und Anwendungen für die Sicherung und Analyse. Dazu gehört die Chain of Custody [10]. Dies bedeutet, dass der Nachweis über den Verbleib und die Bearbeitung einer digitalen Spur, ab dem Zeitpunkt der Erfassung, lückenlos erbracht werden muss. Diese Forderung bedingt strikte technische und organisatorische Forderungen an den gesamten digitalen Ermittlungsprozess.

Eine wichtige Aufgabe der Forensik besteht darin, die bestmöglichste Tathergangsrekonstruktion auf Basis der vorhandenen Spuren zu erstellen. Die Tathergangsrekonstruktion soll als ein hypothesengetriebener Prozess geführt werden. Aus digitalen Spuren lassen sich in der Regel sogenannte Raum-Zeit-Muster ableiten. Diese eignen sich für eine Rekonstruktion des Tathergangs. Die Größe der Zeit kann direkt mit den Zeitstempeln der digitalen Spuren verknüpft werden, was in Form einer Timeline der Ereignisse dargestellt werden kann. Die Raumdimension kann zum einem durch die Verortung der Geräte hergestellt werden und zum anderen aus dem Prozess der Authentifizierung der Personen.

Locard angewandt: Digitale Ermittlungsarbeit

Es gibt schon heute quasi keinen einzigen Kriminalfall mehr, in dem digitale Spuren keine Rolle spielen. Sogar im Zusammenhang mit „gefühlt“ vollständig analogen Straftaten wie Sexualdelikten, Einbrüchen oder Raubüberfällen werden digitale Spuren zur Ermittlung herangezogen. Beispiele hierfür sind Funkzellendaten und Aufnahmen aus digitalen Überwachungssystemen.

Im polizeilichen Kontext spielen im Wesentlichen 3 unterschiedliche Untersuchungsziele eine Rolle:

- Identifizierung der für die Untersuchung relevanten Spuren
- Nachweis des vorgeworfenen Sachverhalts/Delikts oder Entkräftung der Vorwürfe
- Generierung von weiteren Ermittlungsansätzen

Diese Ziele sind für jede Untersuchung im Bereich der digitalen Forensik relevant. Das erste Ziel dient vorwiegend der Reduktion der Daten auf diejenigen, die für die Untersuchung von Bedeutung sind. Fallabhängig ist dies meist ein kleiner Anteil der vorhandenen Dateien (unterer einstelliger Promille- bis Prozentbereich). Die Relevanz für den Fall kann sich – abhängig von der konkreten Fragestellung – aus einer Vielzahl möglicher Aspekte ergeben:

1. *Vorhandensein der Daten*
   Dies gilt vor allem dann, wenn es sich um Daten handelt, die sich auf dem Untersuchungsobjekt überhaupt nicht befinden dürften. Ein Beispiel sind kinderpornografische Daten. Da im Strafgesetzbuch der Besitz unter Strafe gestellt wird, ist mit dem Vorhandensein derartiger Dateien der Besitz – und damit in der Regel das Delikt – nachgewiesen.
2. *Nachweis eines Sachverhalts bzw. der Beteiligung an einem solchen*
   Dies wird beispielsweise dadurch ermöglicht, dass sich die handelnden Personen bei der Tatausführung filmen. Beispiele aus der Praxis: Videos von einer Schlägerei, Bilder von den Tatverdächtigen mit Drogen/Geld/Waffen, Brandstifter posieren mit Feuerzeug vor Rauchsäule etc.
3. *Nachweis der Anwesenheit an einem bestimmten Ort zu einer bestimmten Zeit*
   Indizien für die Anwesenheit an einem bestimmten Ort können einerseits aus den Aufnahmen selbst entstehen

(z. B. Gebäude oder Gegenstände im Hintergrund), andererseits aus den Metadaten der Mediendateien. Beispiele für Letzteres sind regelmäßig Geoinformationen, die in den EXIF/IPTC/XMP-Daten innerhalb oder im Zusammenhang mit der Dateien zu finden sind.

In der Praxis stellt schon das erste Ziel den Ermittler vor immense Herausforderungen, die derzeit nur mit erhöhten Personalressourcen zu lösen sind (siehe auch [11–13]). Oftmals beinhalten Fälle unglaubliche Mengen an Mediendateien, allein innerhalb einer beliebigen Chat-Anwendung. Üblicherweise werden massenhaft Bilder und Video in Chats geteilt. Dazu kommen noch große Anzahlen an Audiodateien, da viele Nutzer Sprachnachrichten einer textuellen Kommunikation vorziehen. Die Bewertung des Inhalts dieser Nachrichten allerdings kann derzeit – mangels in der Praxis nutzbarer semantischer Erkennung – fast ausschließlich manuell durch Ermittler vorgenommen werden. Hier fehlen zweifelsfrei zuverlässige Lösungen, z. B. in Form intelligenter Assistenzsysteme.

Genauso problematisch ist die Prüfung von Bildern und Videos auf deren Inhalt. Selbst klar definierbare Objekttypen (Fahrzeuge, Waffen, Drogen, Geld, Symbole verbotener Organisationen usw.) sind heute nur rudimentär automatisiert auswertbar; die Ergebnisse von Inhaltsprüfungen auf Basis aktueller künstliche-Intelligenz(KI)-unterstützter Systeme können lediglich als Anhalt und Grobprüfung gelten und sparen letztendlich nur wenig Ressourcen. Noch schlechter ist die Bilanz, wenn „unscharfe“ Kriterien dazukommen, die für den Fall von Relevanz sind. Beispiele hierfür sind die Differenzierung zwischen erlaubter Pornografie und solcher, die einer verbotenen Kategorie angehört. Die Kriterien sind – sofern das Alter der handelnden Personen nicht klar bekannt ist – nicht immer absolut trennscharf und damit auch einem unterstützenden KI-System nicht in einer ausreichenden Deutlichkeit antrainierbar.

Erschwert wird das alles noch dadurch, dass es in der kriminalistischen Praxis so etwas wie „den Standardfall“ nicht gibt. Jeder Fall ist anders, praktisch kommen Aspekte dazu, die den Fall einzigartig machen. Diese in eine technische Lösung einfließen zu lassen, ist eine der größten Herausforderungen.

Als Quintessenz läuft es derzeit in (viel zu) vielen Fällen darauf hinaus, dass Medieninhalte manuell geprüft und bewertet werden müssen, da diese von aktuell vorhandener Technik nur unzureichend in der notwendigen Qualität verarbeitet und bewertet werden können. Es scheint in den genutzten Anwendungen reichlich Optimierungspotenzial vorhanden zu sein. Ein Blick auf die aktuellen Arbeitsweisen soll das verdeutlichen:

In der Regel werden Datenträger heute „forensisch“ gesichert, das heißt die Speichermedien werden – sofern möglich – komplett auf unterster Datenebene als Image gesichert („bitgenaue Kopie“) und in Form einer Containerdatei (meist im Encase-Witness-Disc-Image-Format) gespeichert. Zur Verhinderung von unerkannten Änderungen werden schon bei der Sicherung Hashwerte der Daten gespeichert, die später mit den Daten abgeglichen werden können. Meist werden Speichermedien, die Daten enthalten, derzeit physikalisch sichergestellt. Das bedeutet in der Praxis, dass die Ermittler die Computer, Festplatten oder Mobilfunkgeräte im Original mitnehmen. Anschließend wird von den Asservaten eine Kopie erstellt, die als neues Original der Daten gilt. Von dieser wird eine Arbeitskopie angefertigt, mit der die weitere Untersuchung durchgeführt wird. Datenschutz und Datensicherheit nehmen bei den Nutzern, direkt und indirekt, eine immer größere Rolle ein. Dies stellt somit die Ermittler vor erheblichen Schwierigkeiten. In immer mehr Fällen ist bereits die Erstellung einer Datenkopie eine Herausforderung. Abhängig vom System scheitert bereits solch ein Versuch aufgrund fehlender Schnittstellen, vollständiger Verschlüsselung und teilweise proprietärer Zugriffsmethoden. Von einer weiterführenden Aufbereitung der Daten kann dann keine Rede mehr sein.

Wenn die Datensicherung erfolgreich war, ist die nächste Herausforderung derzeit die ständig wachsende Datenmenge, die auf den Systemen gespeichert ist. Beispielsweise findet man in immer mehr Verfahren schon Smartphones mit Speichergrößen jenseits der TB-Grenze. Die wachsenden Speichergrößen bringen mit sich, dass es für die Nutzer kaum noch die Notwendigkeit gibt, ihr System „aufzuräumen“ und nicht mehr erforderliche Daten zu löschen (was aber aus forensischer Sicht nicht nur Nachteile mit sich bringt). Die Menschen werden zu „digitalen Messies“.

Aus der Erfahrung heraus bedingen fehlende praktisch nutzbare automatische Verfahren (wie auch der rechtliche Rahmen), dass viele Auswertevorgänge immer noch überwiegend manuell durchgeführt werden müssen. Die sichergestellten Datenmengen verhindern aber, dass deren Auswertung allein mit „mehr Personal“ garantiert werden kann.

Auf den Systemen finden sich Daten aus einer Vielzahl unterschiedlicher Quellen, die keiner nachvollziehbaren Struktur folgen. Beispiele hierfür sind Daten aus sozialen Netzwerken, Gruppenkommunikation und Daten aus Telefonüberwachungen (hier sowohl klassische Sprachaufzeichnungen als auch Netzwerkverkehr). Erschwert wird die Situation durch Kommunikation, die keinen nachvollziehbaren Regeln folgt (Slang, Emoticons, Digitalbilder mit eingebetteter Schrift, Kommunikation in Fremdsprachen, Sprachnachrichten in unterschiedlichsten Dialekten etc.) in Verbindung mit diversen inhaltlichen Kriterien. Dies stellt die automatisierte Auswertung – auch auf Basis von KI – derzeit vor teilweise nicht überwindbare Hindernisse. Dazu kommt noch die zeitliche Komponente, da viele Ermittlungen aus unterschiedlichen Gründen unter einem erheblichen zeitlichen Druck stehen.

Zum Schluss 2 kleine Beispiele aus der Praxis

Im Rahmen einer Ermittlung wegen Brandstiftung mit einem Millionenschaden werden 2 Tatverdächtige befragt. Die beiden wurden in zeitlichem Zusammenhang mit dem Brand von Zeugen in der Nähe gesehen, allerdings in einem öffentlichen Bereich. Beide bestreiten die Tat. Im Rahmen der Vernehmung wird bei einem der Täter ein Smartphone sichergestellt. Es soll noch während der laufenden Vernehmung ausgewertet werden um eventuelle Erkenntnisse in die Vernehmung einfließen zu lassen. Die Analyse ergab, dass keine Daten im direkten Zusammenhang mit dem Brandort (örtlich/zeitlich) existieren, die einen Hinweis auf die Täterschaft liefern, dafür aber um den betrachteten Zeitraum herum hunderte fremdsprachliche Nachrichten, die auf die Schnelle nicht übersetzt werden können. Die Wende bringt ein Foto, das einen der Tatverdächtigen zeigt, wie er mit einem Feuerzeug und einer Flasche Wein in den Händen ca. einen Kilometer vom Tatort entfernt vor der großen Rauchsäule „posiert". Das Bild wird ausgedruckt und dem vernehmenden Beamten zugeleitet. Im Kontext der laufenden Vernehmung führt die Konfrontation mit dem Bild zu einem Geständnis eines der Tatverdächtigen. Der „vollständig analoge" Fall wurde letztendlich durch digitale Forensik gelöst.

Im Rahmen eines Ermittlungsverfahrens in einem Mordfall konnte die Leiche trotz intensiver Suche unter anderem mithilfe von Spürhunden und Hubschrauberüberwachung nicht gefunden werden. Im Zuge der Ermittlungen konnten die Beamten jedoch einen Tatverdächtigen ermitteln, welcher sich aufgrund der Auswertung von Funkzellendaten in der Nähe des Tatorts befunden haben könnte. Aufgrund der Funkzellengröße, in dem sich das Smartphone des Tatverdächtigen eingebucht hatte, fiel es den Ermittlern jedoch schwer, den tatsächlichen Ablageort der Leiche zu lokalisieren. Im Rahmen der Vernehmung fiel den Beamten auf, dass der Tatverdächtige einen sogenannten „Fitnesstracker" am Arm trug. Diesen konnten die Beamten sichern. Die für die weiteren Ermittlungen wichtigen GPS-Daten des Trackers konnten in der Cloud des Anbieters gesichert werden. Aufgrund der exakten Daten zu Laufstrecke, Höhenmeter etc. konnte die Strecke welche der Tatverdächtige zum Ablageort der Leiche zurückgelegt hatte rekonstruiert und die Leiche gefunden werden.

Die Locard'sche Intelligenz: KI und Ermittlungsarbeit

Im Jahre 2025 werden einer Schätzung zur Folge 55 Mrd. Geräte online sein [14]. Ein Tatort ohne digitale Spuren ist dann kaum noch denkbar. Die so anfallenden Daten, die relevant für eine Ermittlung sind, auszuwerten, ist zunehmend nur mit maschineller Unterstützung möglich. Die Aufgaben und Herausforderungen in der digitalen Forensik, bei der Menschen im Mittelpunkt stehen, sind Teil des wissenschaftlichen Diskurses [15]. Einen Teil davon stellt die Unterstützung der Ermittelnden durch Assistenzsysteme dar. Dabei ist die Nachvollziehbarkeit der erzeugten Ergebnisse eine der notwendigen Herausforderungen. Mit der steigenden Verwendung von intelligenten Systemen wächst das Bedürfnis der Nachvollziehbarkeit sogar noch [16].

Die Entwicklungen der letzten Jahre haben gezeigt, dass meist die Heterogenität der zu verarbeitenden Spuren und deren Datenfehlern, wie inkorrekte oder veraltete Information, Inkonsistenzen oder fehlende Werte sowie die Menge der irrelevanten Daten, problematisch ist [17]. Das Fehlen von automatischer Datentyperkennung, z. B. durch Entropieanalysen [18], und der Mangel an ontologischer Integration, wie beispielsweise Data Property Klassifikation [19], und damit das Verständnis für die Bedeutung unstrukturierter Daten, macht dies zu einer hochgradig manuellen Arbeit.

Methoden der künstlichen Intelligenz auf forensische Untersuchungen werden bisher nicht nur aus technischen, sondern auch aus rechtlichen Gründen noch nicht eingesetzt [20]. Typische Anwendungen sind beispielsweise automatisches Profiling von Verdächtigen (z. B. mittels Social Media oder Open Source Intelligence), Fahrzeugidentifikation (z. B. automatische Nummernschilderkennung), Kryptowährungsanalysen oder automatische Erkennung von kinderpornografischen Schriften [21].

Mit der Zunahme der Nutzung weiterer Kommunikationskanäle wie Instant Messaging [22] ist die zu analysierende Menge an Spuren, weit über menschliches Vermögen hinaus, angewachsen. Durch die vereinfachte Verwendung von Anonymisierungstechniken entstehen dabei neue Herausforderungen wie die Verwendung von Autorenbestimmungsmethoden [23]. Bei klassischen Medien wie E-Mail wird dies schon seit Jahren beforscht [24], findet jedoch noch kaum Anwendung im forensischen Kontext. Auch Ansätze zur Verwendung von maschinellem Lernen in der Forensik sind schon länger Teil des wissenschaftlichen Diskurses [25].

An vielen Stellen in einer Ermittlung können Methoden der künstlichen Intelligenz die Arbeit erleichtern, auch wenn der Prozess einer Ermittlung zwischen mehreren Personen wechselt. Hier könnten Fehler vermieden und automatisierbare Prozessschritte durch maschinelles Lernen abgebildet und in Zukunft automatisch übernommen werden. Die Interaktion zwischen Forensikern und Ermittlern ist dabei je nach Kontext neu zu definieren. Ein Versuch, diesen Prozess zu formalisieren und zu analysieren, wurde in [26] unternommen. Dabei werden in [26] verschiedene Hilfestellungen vorgestellt (Trilogy und Sentinel, Coplik, Forensic Led Intelligence System und Crime Investigation Decision

Support System) und deren Probleme und Grenzen diskutiert. Leider sind Sprachmodelle wie BERT, GPT-3 und Bildmodelle wie Image GPT-3 [27], AlexNet oder VGG16 [28] noch nicht in forensischen Anwendungen integriert.

Die Enkel Locards – quo vadis?

Die Zusammenführung von digitalen und analogen Spuren, um die forensische Aufgabe zu erfüllen, stellt eine große Herausforderung für die moderne interdisziplinäre Forensik dar. Nicht nur durch die Menge der in Ermittlungsverfahren erhobenen Spuren, sondern auch durch die Verwertbarkeit müssen neue zielführende Werkzeuge entwickelt und evaluiert werden. Immer mehr Leute sind mobil digital unterwegs und geben immer mehr Informationen (bewusst oder unbewusst, letztlich aber freiwillig) über sich in Systeme ein, die diese in großen Datenbanken verarbeiten, zu Profilen zusammenführen und selbstständig Schlüsse ziehen. Computersysteme werden unser Leben zunehmend durchdringen. Mit der fortschreitenden Verbreitung von IoT-Devices (siehe Entwicklungen in den Bereichen Smarthome und Industrie 4.0) sowie Big Data und Cloud-Computing werden neue Angriffsvektoren geschaffen und die Kriminalitätsrate wird hier entsprechend ansteigen. Gleiches gilt für die computergestützte Einflussnahme auf Märkte und Personen. Cybercrime-as-a-Service ermöglicht es mehr und mehr auch Computerlaien, Straftaten zu begehen [29].

Der ständige technische Fortschritt wirft allerdings die Ermittlungsbehörden immer wieder zurück. Kaum ist eine Technik durchdrungen und es haben sich Standards in der IT-Forensik etabliert, wird diese Technik zunehmend disruptiv abgelöst. Dieses lässt sich am Beispiel der Speicherung von Daten auf der Festplatte verdeutlichen. Während Daten auf klassischen Magnetspeichern so gut wie immer wiederhergestellt werden können, wenn sie nur gelöscht und nicht überschrieben wurden, so geben die Controller heutiger SSD gelöschte Daten einfach nicht mehr heraus, wenn TRIM (Funktion stellt spezielle Verbindung des Betriebssystems zum SSD-Controller her) aktiviert wurde. Hier steht die IT-Forensik vor neuen Herausforderungen, Wege zu finden, die Daten trotzdem auszulesen. Controllermanipulation und Chiptransplantation stecken hier erst in den Kinderschuhen (siehe auch [30–32]).

Der immer größer werdenden Flut von auszuwertenden Daten, und somit auch des Informationsgehalts ist nur durch Automatisierung Herr zu werden. Verfahren der künstlichen Intelligenz können und werden die Ermittlungsbehörden zukünftig dabei zunehmend unterstützen. Dabei ist unter anderem die multimodale Verarbeitung von Daten eine eigene Herausforderung. Dazu gehörten beispielsweise die Objekterkennung und somit die Verbindung zwischen bildlichen und textlichen Darstellungen. Die semantische Analyse von Bildern oder Videos wird dabei bis heute beforscht [33, 34]. Image GPT ist dabei ein aktuelles Beispiel, wie anhand von Bildern einem System Bezeichner, auch Labels genannt, angelernt werden können. Dieses System kann mittels One-Shot-Learning Objekte auf Bildern erkennen, ohne diese vorher gesehen zu haben [35].

Viele Bereiche der KI-Forschung können Anwendung in der Forensik finden. Leider ist diese Verbindung noch nicht soweit etabliert, dass die wissenschaftliche Community hier ihre Evaluationen sucht. Datensätze, Problemstellungen und Anwendungsszenarien könnten und sollten geschaffen werden, damit mehr der neuen Methoden Anwendung finden. Speziell für den Anwendungsbereich der Ermittlungen ist jedoch, dass die in der Forschung entwickelten Prototypen jeweils auf ihre Erklärbarkeit und Nachvollziehbarkeit geprüft werden müssen. Ausreichend Verständnis der verwendeten Methoden ist notwendig, um sicherstellen zu können, dass hier kein Fehler in der Klassifikation entstanden ist. Bei den meisten sogenannten „Blackbox-Verfahren“ wie großen neuronalen Netzen, funktioniert eine Nachvollziehbarkeit jedoch nur bedingt [36]. Einer Anwendung dieser „Blackbox-Verfahren“ sind in einem Ermittlungsverfahren mehrere Hürden in den Weg gestellt: Sie müssen die richtige Erkenntnis erzeugen, dann aber auch rechtlich vor Gericht standhalten, indem erklärt werden kann, warum dieser Methode und ihrem Ergebnis getraut werden kann. Hier entsteht ein interessanter interdisziplinärer Forschungsbereich zwischen Informatik und Rechtswissenschaften.

Funding Open Access funding enabled and organized by Projekt DEAL.

Literatur

1. Locard E (1930) Die Kriminaluntersuchung und ihre wissenschaftlichen Methoden. Kameradschaft, Berlin, S 139 (frz. Originalausgabe 1920: Locard, E: L’enquête criminelle et les méthodes scientifiques)

2. Inman K, Rudin N (2000) Principles and practice of criminalistics: the profession of forensic science. CRC, Boca Raton
3. Dewald A, Freiling FC (Hrsg) (2015) Forensische Informatik, 2. Aufl. Books on Demand, Norderstedt
4. Brinson A, Robinson A, Rogers M (2006) A cyber forensics ontology: creating a new approach to studying cyber forensics. Digit Investig 3:37–43
5. Stoyanova M, Nikoloudakis Y, Panagiotakis S, Pallis E, Markakis EK (2020) A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Commun Surv Tutor 22(2):1191–1221
6. Gletschertraum (2007) Beweislehre – Der Beweis (Definition). https://www.gletschertraum.de/Kriminalistik1/DerBeweis(Definition).html. Zugegriffen: 14. Mai 2021 (Lehrmaterialien zur Kriminalistik I)
7. Momsen C, Hercher N (2014) Digitale Beweismittel im Strafprozess – Eignung, Gewinnung, Verwertung, Revisibiltät. In: Die Aktzeptanz des Rechtsstaats in der Justiz – 37. Strafverteidigertag 2014, S 173–196
8. Casey E (2011) Digital evidence and computer crime. Forensic science, computers, and the Internet, 3. Aufl. Academic Press, Waltham
9. Labudde D, Spranger M (Hrsg) (2017) Forensik in der digitalen Welt: Moderne Methoden der forensischen Fallarbeit in der digitalen und digitalisierten realen Welt, 1. Aufl. Springer Spektrum,
10. Lone AH, Mir RN (2019) Forensic-chain: blockchain based digital forensics chain of custody with PoC in Hyperledger Composer. Digit Investig 28:44–55
11. Garbers N (2019) Erkennung inkriminierter Bilder: Neuronale Netzarchitekturen und Hautanteilfilter im Vergleich. In: Honekamp W, Kühne E (Hrsg) Polizei-Informatik 2019, S 83–99
12. Mayer F, Steinebach M (2018) Unterstützung bei Bildsichtungen durch Deep Learning. In: Honekamp W, Bug S (Hrsg) Polizei-Informatik 2018, S 66–78
13. Schulze R (2016) LiDaKrA – Linked-Data Kriminalanalysesystem für die Ermittlungsunterstützung. In: Honekamp W, Mielke J (Hrsg) Polizei-Informatik 2016, S 31–34
14. Ouerfelli FE, Barbaria K, Zouari B, Fachkha C (2020) Prevention of DDoS attacks in IoT networks. International Conference on Advanced Information Networking and Applications (AINA 2020). Springer, Cham, S 1239–1250 https://doi.org/10.1007/978-3-030-44041-1_106
15. Sunde N, Itiel ED (2019) Cognitive and human factors in digital forensics: problems, challenges, and the way forward. Digit Investig 29:101–108. https://doi.org/10.1016/j.diin.2019.03.011
16. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115
17. Garfinkel S (2012) Lessons learned writing digital forensics tools and managing a 30TB digital evidence corpus. Digit Investig 9:80–89
18. Conti G, Bratus S, Shubina A, Sangster B, Ragsdale R, Supan M, Lichtenberg A, Perez-Alemany R (2010) Automated mapping of large binary objects using primitive fragment type classification. Digit Investig 7(Supplement):3–12. https://doi.org/10.1016/j.diin.2010.05.002
19. Glimm B, Horrocks I, Motik B, Stoilos G (2010) Optimising ontology classification. In: International semantic web conference. Springer, Berlin, Heidelberg
20. Rademacher T (2020) Artificial intelligence and law enforcement. Regulating artificial intelligence. Springer, Cham, S 225–254
21. Raaijmakers S (2019) Artificial intelligence for law enforcement: challenges and opportunities. IEEE Secur Priv 17(5):74–77
22. O'Day DR, Ricardo AC (2013) Text message corpus: applying natural language processing to mobile device forensics. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW).
23. Iqbal F, Debbabi M, Fung BCM (2020) Artificial intelligence and digital forensics. In: Machine learning for authorship attribution and cyber forensics. Springer, Cham, S 139–150
24. De Vel O (2000) Mining e-mail authorship. Workshop on Text Mining, ACM International Conference on Knowledge Discovery and Data Mining (KDD'2000).
25. McClendon L, Meghanathan N (2015) Using machine learning algorithms to analyze crime data. Mach Learn Appl 2(1):1–12. https://doi.org/10.5121/mlaij.2015.2101
26. Giles O, Chapman B, Speers J (2020) Forensic intelligence and the analytical process. Wiley Interdiscip Rev Data Min Knowl Discov 10(3):e1354
27. Chen M, Radford A, Child R, Wu J, Jun H, Dhariwal P, Luan D, Sutskever I (2020) Generative pretraining from pixels. https://cdn.openai.com/papers/Generative_Pretraining_from_Pixels_V2.pdf. Zugegriffen: 14. Mai 2021 (International Conference on Machine Learning)
28. Zhang Y, Fu H, Dellandréa E (2017) Adapting convolutional neural networks on the shoeprint retrieval for forensic use. In: Chinese Conference on Biometric Recognition. Springer, Cham https://doi.org/10.1007/978-3-319-69923-3_56
29. Honekamp W (2018) Cybercrime: Aktuelle Erscheinungsformen und deren Bekämpfung. In: Lange HJ, Model T, Wendekamm M (Hrsg) Zukunft der Polizei. Trends und Strategien. Springer VS, Wiesbaden, S 47–59
30. Attingo Datenrettung (2020) Gefahr für gelöschte Daten: TRIM-Befehl bei Solid State Drives. https://www.attingo.de/blog/gefahr-fuer-geloeschte-daten-trim-befehl-bei-solid-state-drives/. Zugegriffen: 14. Mai 2021
31. O&O Software (2019) Benutzerhandbuch O&O SafeErase 12. https://www.oo-software.com/de/docs/usersguide/oose14.pdf. Zugegriffen: 14. Mai 2021
32. Voges H (2013) Workshop Windows 8.1 – Verwaltung und Fehlerbehebung. https://www.netz-weise-it.training/images/dokus/Handout_Wordshop_Windows_8.1.pdf. Zugegriffen: 14. Mai 2021
33. Cho J, Lu J, Schwenk D, Hajishirzi H, Kembhavi A (2020) X-LXMERT: paint, caption and answer questions with multimodal transformers. 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).
34. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020) End-to-end object detection with transformers. In: Vedaldi A, Bischof H, Brox T, Frahm JM (Hrsg) Computer vision – ECCV 2020. Lecture notes in computer science, Bd. 12346. Springer, Cham, S 213–229 https://doi.org/10.1007/978-3-030-58452-8_13
35. Chen M, Radford A, Sutskever I (2020) Image GT. https://openai.com/blog/image-gpt/. Zugegriffen: 14. Mai 2021
36. Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R (Hrsg) (2019) Explainable AI: interpreting, explaining and visualizing deep learning, 1. Aufl. Lecture notes in computer science, Bd. 11700. Springer, Cham

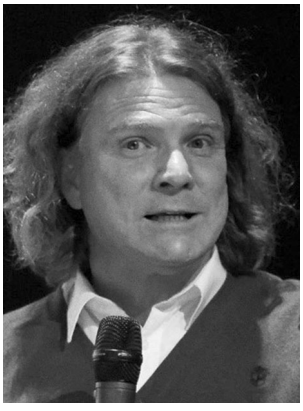
Roman Povalej

Silvio Berner

Heiko Rittelmeier

Wilfried Honekamp

Johannes Fähndrich



Dirk Labudde