EasyChair Preprint

№ 11565

Multidimensional Trace Search in IT Forensics

Thomas Hrdinka

December 18, 2023

MULTIDIMENSIONALE SPURENSUCHE IN DER IT-FORENSIK

Thomas Hrdinka

Zivilingenieur, Universität Wien
Ocwirkgasse 22, 1210 Wien, AT
thrdinka@zth.at; http://www.zth.at

Schlagworte: ***Strafverfahren, Beweismittel, IT-Forensik***

Abstract: *Wie IT-Spuren welche mit einem Täter in Verbindung gebracht werden können zu bewerten sind lassen sowohl die Normen als auch die Rsp. offen. Auch bezüglich technische Normen existiert kein Ansatz, IT-Spuren systematisch zu bewerten, sodass ein Zusammenhang mit dem Täter, dem Tatort und der Tatzeit hergestellt oder ausgeschlossen werden kann. Aus der Literatur ist jedoch ein durchaus brauchbarer Ansatz bekannt, wo mit Hilfe einer 7-stufigen Bewertungsskala, die auf die Kombination von Wahrscheinlichkeit und Manipulationssicherheit der Spuren eingeht. Für die Bestimmung der Schuld oder Unschuld des Angeklagten muss unter der Beachtung der freien Beweiswürdigung mit an Sicherheit grenzender Wahrscheinlichkeit feststehen, dass kein Freispruchgrund vorliegt. Hingegen wird im Zivilprozess nach der überwiegenden Wahrscheinlichkeit gewertet. Generell ist die Beurteilung von IT-Spuren hins. Schuld oder Unschult eine vage Angelegenheit, die nur mit viel Verantwortung und Sachverstand bewerkstelligt werden kann, um einen Bias möglichst hintanzuhalten. Unterstützende Werkzeuge, Methoden oder Verfahrensweisen fehlen vollkommen. Diese Arbeit versucht die 7-stufige Bewertungsskala mehrdimensional zu erweitern, sodass ein systematischer Ansatz zur Spurenbewertung in der IT geschaffen wird.*

1. Einleitung

Zwecks Bewertung digitaler Spuren schlug CASEY[1] eine diskrete Bewertung in Klassen, mit Berücksichtigung von Manipulationsmöglichkeiten vor und zwar von C0 (fehlerhaft) bis C6 (sicher). Als Instrument kann solcherart Klassifikation Ermittlern die Möglichkeit bieten ihre Schlüsse die auf digitalen Beweisen beruhen zu formalisieren. Ein Vorteil dieser Bewertungsmethodik besteht darin, dass die Bestimmung des Beweiswerts der Beweismittel relativ flexibel ist. Gem. § 14 StPO[2] wird die freie Beweiswürdigung normiert, denn ob Tatsachen als erwiesen festzustellen sind, hat das Gericht auf Grund der Beweise nach freier Überzeugung zu entscheiden; im Zweifel stets zu Gunsten des Angeklagten oder sonst in seinen Rechten Betroffenen. Für die Bestimmung der Schuld oder Unschuld des Angeklagten muss mit an Sicherheit grenzender Wahrscheinlichkeit feststehen, dass kein Freispruchgrund vorliegt. Dieser liegt gem. § 259 Z. 3 leg.cit. u.a. dann vor, wenn der Tatbestand nicht hergestellt oder nicht erwiesen sei, dass der Angeklagte die ihm zur Last gelegte Tat begangen habe, oder dass Umstände vorliegen, durch die die Strafbarkeit aufgehoben oder die Verfolgung aus anderen als den unter Z. 1 und 2 angegebenen Gründen ausgeschlossen ist.

Im Zivilverfahren gilt hingegen nach h.M. ein Beweismaß nach der Wahrscheinlichkeitstheorie. Die behauptete Tatsache muss mit hoher Wahrscheinlichkeit wahr sein, und es genügt die überwiegende Wahrscheinlichkeit.[3]

---

1 CASEY: Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. Academic Press, 2011.

2 Strafprozessordnung, BGBl. 631/1975, i.d.g.F.

3 Vgl. 2 Ob 185/91i, RS 0110701 : „*Das Regelbeweismaß der ZPO ist die hohe und nicht eine an Sicherheit grenzende Wahrscheinlichkeit; eine solche ist nur in den Fällen eines erhöhten Regelbeweismaßes erforderlich.*"

Für einen Schuldspruch im Strafverfahren sind somit nach der Methode von CASEY folgende zwei Stufen jedenfalls relevant:

- Fast sicher C5: Beweise aus mehreren unabhängigen Quellen, die vor Manipulationen geschützt sind, stimmen überein. Es gibt jedoch kleine Unsicherheiten wie Datenverlust.
- Sicher C6: Die Beweise sind manipulationssicher und/oder weisen eine hohe statistische Konfidenz auf.

Der Nachteil dieser Methodik ist, dass sie keinerlei Bezug zum Täter (Verdächtiger, Beschuldigter oder Angeklagter gem. § 48 Abs. 1 StPO) herstellt. Auch werden für einen Schuld- oder Freispruch wesentliche Faktoren wie gem. der Handlungstheorie die Tatzeit § 67 Abs. 1 StGB[4] außer Acht gelassen, wobei die Tat zu der Zeit begangen wurde, da der Täter gehandelt hat oder hätte handeln sollen. Es kommt somit nur auf den Zeitpunkt des Handelns an. Die Kombination von Zeit und Ort der Vornahme der Tathandlung oder der Unterlassung wird durch die Einheits- oder Ubiquitätstheorie[5] bestimmt. Der inländische Tatort „*liegt gem. § 67 Abs. 2 StGB im Sinne der geltenden Einheitstheorie vor, wenn der Ort, an dem der Täter gehandelt hat oder hätte handeln sollen oder ein dem Tatbild entsprechender Erfolg ganz oder zum Teil eingetreten ist oder nach der Vorstellung des Täters hätte eintreten sollen, im Inland liegt.*"[6]

2. Fallbeispiel

Ein jüngstes Fallbeispiel soll die vorher beschriebene Problematik verdeutlichen.

Bei einem wg. § 207 Abs. 1, § 212 Abs. 1 Z. 2, § 208 Abs. 2 und § 207a Abs. 3 2. Fall StGB Angeklagten wurden im Ermittlungsverfahren 66 Datenträger sichergestellt, darunter USB-Sticks, externe Festplatten, PCs und Mobiltelefone. Die aufwändigen Untersuchungen der Kriminalpolizei resultierten schließlich in 8 positiven Datenträgern, wo kinderpornographisches Material gefunden wurde. Der Angeklagte bestritt nicht den Besitz der Datenträger, jedoch zeigte er sich hins. der Verantwortung der Speicherung des inkriminierten Materials als nicht geständig. Im Hauptverfahren stellte sich schließlich u.a. die zentrale Frage, ob technischerseits festgestellt werden kann, dass die auf den sichergestellten Datenträgern vorgefundenen Dateien, die pornographische Darstellungen mündiger und unmündiger minderjähriger Personen beinhalten (Lichtbilder und Videos) vom Angeklagten auf diese Datenträger heruntergeladen und gespeichert bzw. dort gesichtet, verwendet oder bearbeitet wurden. Mit anderen Worten, sollte erst im HV der Zusammenhang des inkriminierten Datenmaterials mit dem Täter bewiesen werden.

Im Polizeibericht über die gefundenen und gesichteten Dateien waren jedoch – bis auf wenige Ausnahmen – keine Metadaten wie Dateiname, Zeitstempel, bzw. die Art und Weise woher diese Dateien stammen enthalten. Die Ermittler extrahierten, sichteten und klassifizierten zig-tausende Bild- und Videodateien in monatelanger Arbeit, ohne eine Untersuchung, wie und wann dieses Material auf die Datenträger kam, und die wichtigste Frage, ob dieses überhaupt mit dem Tatverdächtigen in Verbindung gebracht werden kann. Die Polizei und StA schienen vermutlich davon auszugehen, dass der Besitz eines Datenträgers alleine genügt, um auch in die Verantwortung der Speicherung von inkriminiertem Datenmaterial zu übernehmen.

Die IT-forensische Untersuchung der verbliebenen 8 inkriminierten Datenträger ergab, dass die Ermittler offenbar mit Hilfe von „Carving" auf dieses Material stießen, was die fehlenden Metadaten der Dateien erklärt. Carving stellt eine Methode dar, um Dateien auf Speichermedien ohne die Hilfe des Dateisystems zu identifizieren und wiederherzustellen. Dazu wird der Datenstrom des Speichermediums nach charakteristischen Mustern oder anderen typischen Kopfdatenstrukturen bekannter Dateiformate durchsucht. Da keine Metadaten dieser Dateien mehr vorhanden sind, werden von der Carvingsoftware generische Dateinamen mit fort-

4 Strafgesetzbuch, BGBl. 60/1974 i.d.g.F.

5 KIENAPFEL, HÖPFEL, KERT, AT15 E 12 Rz. 6; OEHLER, Internationales Strafrecht Rz. 246; AMBOS, Internationales Strafrecht § 1 Rz. 17.

6 OGH 12 Os 111/06z.

laufenden Nummern vergeben. Die Ergebnisse sind in den meisten Fällen einwandfrei wieder hergestellte Bilder und Videos.

Das Endergebnis der umfangreichen technischen Untersuchungen war schließlich, dass das inkriminierte Material zweier Datenträger keinesfalls dem Angeklagten zuordenbar war. Offenbar, und das entspricht auch seiner Aussage, erhielt er diese PCs von Dritten, was im HV technisch belegt wurde, denn von diesen – obwohl schon formatierten und überschriebenen – Datenträgern konnten zahlreiche weitere Dokumente wie Word oder PDFs gecarved werden, die definitiv nicht vom Angeklagten stammen konnten, was auch die Tatzeit betrifft. Word, PDF und Bilddateien enthalten selbst Inhalts- und Metadaten, welche auf eine Urheberschaft und Zeitstempel hinweisen. Zu genau diesen Zeiten war der Angeklagte aber weder im Besitz dieser PCs noch konnte die Tatzeit mit ihm in Verbindung gebracht werden, da er damals wg. anderer Delikte in Haft war.

Weitere 4 positive Datenträger, welche er ebenfalls von Dritten übernommen hatte, konnten aber aufgrund bereits vernichteter Metadaten weder dem Angeklagten zugeordnet noch nicht zugeordnet werden. Lediglich für die Inhalte von 2 (von 66 sichergestellten) positiven Datenträgern musste sich der Angeklagte letztlich verantworten. Die Aufwände für diese Untersuchungen waren erheblich; eine effizientere und zielgerichtete Vorgehensweise wäre jedenfalls geboten gewesen.

3. Vorschlag für eine multidimensionale Spurensuche und Bewertung

Anhand des vorangegangen beispielhaften Falls, muss nun die Vorgehensweise der Ermittlungsbehörden hinterfragt werden. Es macht nämlich wenig Sinn auf allen Datenträgern alle automationsunterstützt gefundene inkriminierte Bilder einzeln zu sichten. Es wäre zweckmäßiger den Prozess so umzukehren, dass eine detaillierte manuelle Sichtung erst dann vorgenommen wird, nachdem erst anhand von Stichproben eine mit an Sicherheit grenzender Wahrscheinlichkeit Zuordenbarkeit zum Tatverdächtigen hergestellt oder aufgrund einer Unwahrscheinlichkeit ausgeschlossen wurde. Auch könnte man das Beweismaß in diesem Stadium der Ermittlungen auch auf eine überwiegende Wahrscheinlichkeit hin reduzieren. Jedenfalls könnte die Bewertungsmethode nach CASEY i.V.m. den Dimensionen Täter, Tatort und Tatzeit helfen, bereits im Ermittlungsverfahren wertvolle Zeit und Kosten zu sparen, ohne, dass Gefahr bestünde, ungenau oder unpräzise zu ermitteln.

Folgender Prüfansatz soll dazu die Grundlage bilden:

1. Suche automationsunterstützt alle Spuren. Wenn dies nicht möglich ist, dann müssen manuell – und dazu ist viel Erfahrung nötig – zielgerichtet Samples gewonnen werden. In diesem Schritt besteht die Gefahr eines unzulässigen Erkundungsbeweises.[7]
2. Kann mindestens eine der Spuren dem Verdächtigen zugeordnet werden? Bspw. anhand von Inhalten, Profilen, Useraccounts, …?
3. Kann mindestens eine der Spuren einer Tatzeit zugeordnet werden? Wenn nicht, kann aus der Erfolgszeit die Spur bis zur Tatzeit (bspw. anhand von Log-Dateien) weiter verfolgt werden?
4. Ist der Tatort relevant? Kann dieser bspw. anhand von Netzwerkprotokollen wie IP-Adressen festgestellt werden?
5. Kann der Tatverdächtige mit Tatort und Tatzeit in Verbindung gebracht werden?
6. Gibt es von den relevanten Spuren gleichartige, d.h. redundante?
7. Wie waren die Spuren gegen Manipulation geschützt?

7 Vgl. OGH 18.02.1986, 10 Os25/86, RS0099841: „*Von einem Erkundungsbeweis kann nur dann gesprochen werden, wenn Ermittlungen veranlasst sollen, um die Frage zu klären, ob (überhaupt) von bestimmten Beweisen eine Förderung der Wahrheitsfindung zu erwarten ist, oder ob überhaupt Beweismittel auffindbar sind, deren Heranziehung der Wahrheitsfindung dienlich sein können.*“

3.1. Algorithmus zur Spurensuche

Mit Hilfe der Rechtsvisualisierung[8] soll nun der vorher beschriebene Algorithmus detaillierter graphisch visualisiert werden. SysML[9] Zustandsdiagramme bilden hier die Grundlage:

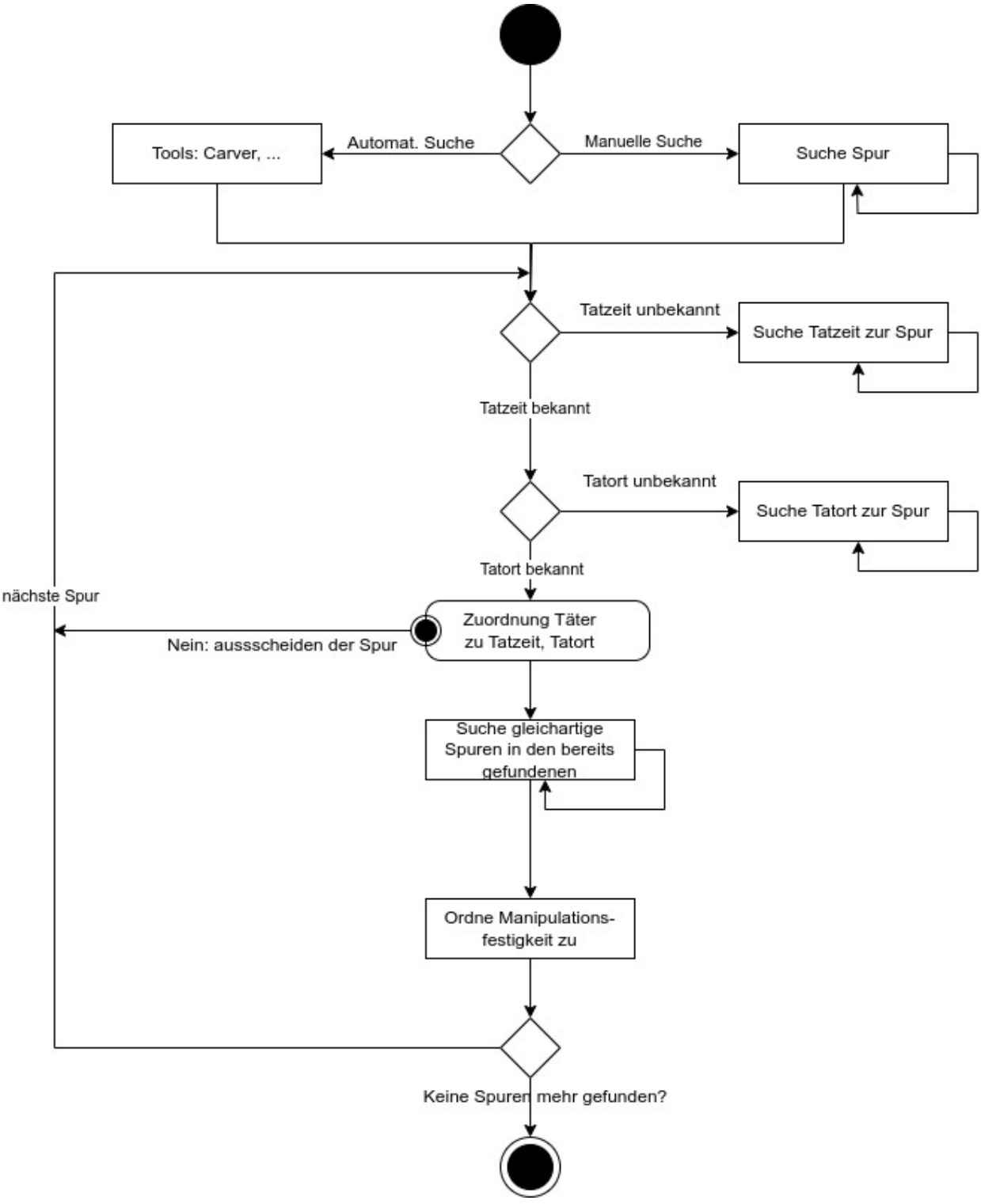

Abbildung 1: Algorithmus zur Spurensuche

8 Siehe Vorschlag HRDINKA: Rechtsvisualisierung IT-Forensischer Prozesse: Tagungsband des 26. Internationalen Rechtsinformatik Symposions IRIS 2023, S 437.

9 Systems Modeling Language der Object Management Group, https://www.omgsysml.org/.

3.2. Multidimensionale Spurenbewertung

Aufgrund der algorithmischen Vorgehensweise im vorherigen Schritt können nun die gefundenen Spuren einzeln kategorisiert und bewertet werden. Jede Spur hat eine Beziehung zu Dimensionen und beinhaltet als Attribute Messwerte. Dimensionen sind Tatzeit, Tatort, Täter und Manipulationsschutz, als Messwerte kommen die 7 Bewertungsstufen in Frage. Die Umsetzung als multidimensionaler Würfel, bspw. technisch durch ein multidimensionales Array, erscheint unpraktisch und nicht flexibel. Vielmehr bietet sich das aus dem relationalen Datawarehouse bekannte Star-Schema an, welches einerseits bezüglich der Dimensionen flexibel ist, und andererseits zumal nicht alle Kombinationen aller Dimensionen als Spur vorhanden sind, auch speicheroptimal. Ziel beim Star-Schema ist eine Denormalisierung der Datenbank, um den Lesezugriff zu optimieren. Dies steht zwar hier nicht im Vordergrund, jedoch eignet sich dieses Schema perfekt für eine multidimensionale Speicherung und verzichtet bewusst auf eine Normalisierung in die 3. NF.[10] Im Zentrum steht eine Faktentabelle welche die Fremdschlüssel zu allen umgebenden Dimensionentabellen beinhaltet. Die Messwerte, oder auch Fakten, sind Maßzahlen wie die 7 Bewertungsstufen. Die Dimensionen beinhalten Stammdaten von Personen oder Orten, jeweils mit Primärschlüsseln eindeutig identifiziert.

Als Darstellung eignen sich SysML Blockdiagramme:

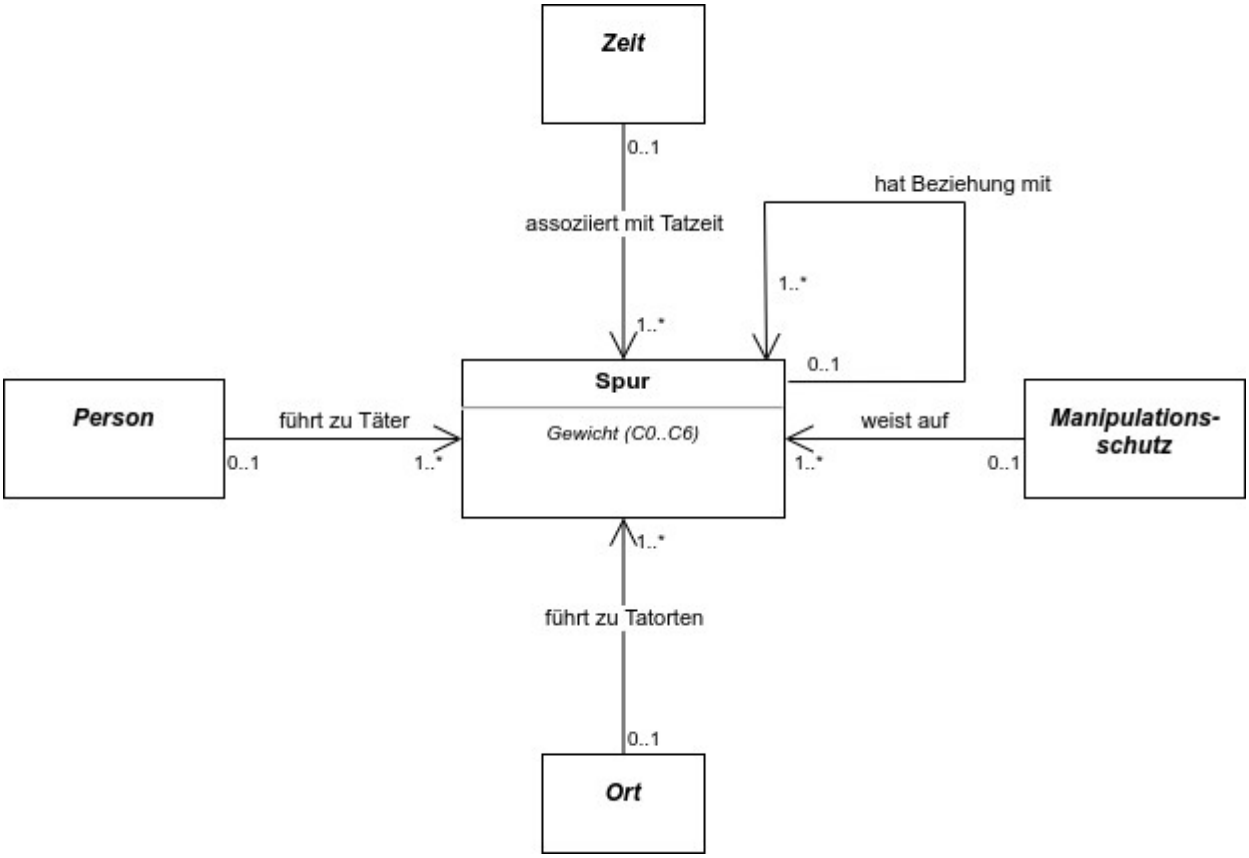

Abbildung 2: Star-Schema Modell einer Spurenbewertung

Im Unterschied zum Star-Schema im Datawarehouse wird hier zusätzlich die Möglichkeit geschaffen, dass Spuren untereinander Beziehungen haben, was schließlich zu einer Hierarchie führen kann. Die Befüllungsgrade der Dimensionen ist eher gering, da es meist nur eine oder wenige verdächtige Personen geben wird.

10 3. Normalform: Eine Relation erfüllt die 2. NF und es bestehen keine funktionalen Abhängigkeiten der Nichtschlüssel-Attribute untereinander (transitive Abhängigkeiten). Alle Nichtschlüssel müssen voll funktional abhängig vom Schlüssel sein.

3.3. Multidimensionales Beispiel

Echtdaten aus Fällen können hier aufgrund Vertraulichkeitsverpflichtungen und Datenschutz nicht gezeigt werden, und weiters weil eine Pseudonymisierung einen sehr hohen Aufwand bedeuten würde. Als Datengrundlage für ein Beispiel soll daher die bekannte Demo-Datenbank „HR“[11] von Oracle[12] dienen, die eine Firma darstellt.

Als Anfangsverdacht wird eine „mafiose Struktur“ angenommen, und aus diesem Grund werden die Personen, die Zeiten der Einstellung und die Orte des Arbeitsplatzes als Dimensionen geführt. Die Fakten sind das Gehalt, ein angenommenes Spurengewicht von C0 bis C6 (hier zufällig generiert). Weiters existiert auch ein Vorgesetzter. In Oracle SQL lässt sich aus HR das Star-Schema wie folgt erzeugen:

```
CREATE MATERIALIZED VIEW m_person as
       SELECT employee_id id, last_name, first_name, email
       FROM hr.employees;

CREATE MATERIALIZED VIEW m_ort as
       SELECT location_id id, city, country_name country
       FROM hr.locations JOIN hr.countries using(country_id);
ALTER MATERIALIZED VIEW m_ort ADD PRIMARY KEY (id);

CREATE MATERIALIZED VIEW m_zeit as
       SELECT employee_id id, extract(year FROM hire_date) YEAR,
              extract(MONTH FROM hire_date) MONTH, extract(DAY FROM hire_date) day
       FROM hr.employees;

CREATE MATERIALIZED VIEW m_fact as
       SELECT employee_id fk_person, employee_id fk_zeit, location_id fk_ort,
              PRIOR employee_id fk_chief, salary, round(dbms_random.value(0,6), 0) c
       FROM hr.employees JOIN hr.departments USING (department_id)
              JOIN hr.locations USING (location_id)
       START WITH hr.employees.manager_id IS null
       CONNECT BY hr.employees.manager_id = PRIOR employee_id;

ALTER MATERIALIZED VIEW m_fact ADD FOREIGN KEY (fk_person) REFERENCES m_person(id);
ALTER MATERIALIZED VIEW m_fact ADD FOREIGN KEY (fk_zeit) REFERENCES m_zeit(id);
ALTER MATERIALIZED VIEW m_fact ADD FOREIGN KEY (fk_ort) REFERENCES m_ort(id);
ALTER MATERIALIZED VIEW m_fact ADD PRIMARY KEY (fk_person, fk_zeit, fk_ort);
ALTER MATERIALIZED VIEW m_fact ADD FOREIGN KEY (fk_chief) REFERENCES m_person(id);
```

Abbildung 3: SQL Script zur Erzeugung eines Star-Schemas verdächtiger Personen

Das Ergebnis ist folgendes (hier mittels DBeaver[13]) reverse-engineerte ER-Modell:

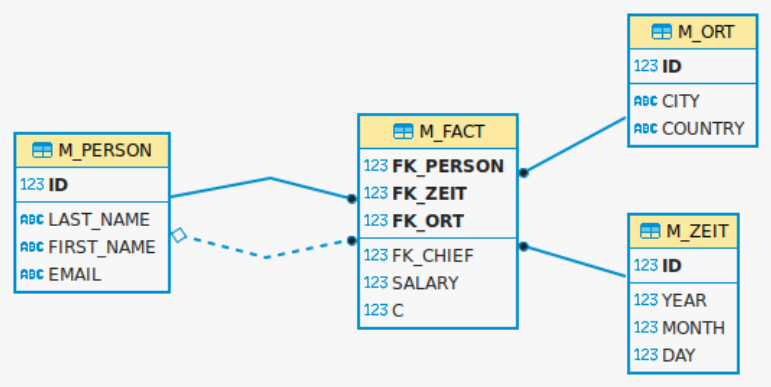

Abbildung 4: Reverse-engineertes ER-Modell des Beispiels

11 https://github.com/bbrumm/databasestar/tree/main/sample_databases/oracle_hr.
12 https://www.oracle.com/.
13 https://dbeaver.io/.

3.4. Auswertung des Beispiels

Die manuelle Suche nach Zusammenhängen zwischen den Spuren, bzw. auch gleichartige Spuren, ist in den meisten Fällen ein sehr hoher Aufwand, da einerseits Massendaten untersucht werden müssen, und andererseits die Spuren obwohl gleichartig nicht konsistent sein müssen. Bspw. kann eine IP-Adresse mehrere Darstellungsformen aufweisen wie 10.0.0.1 oder 10-0-0-1. Aufbauend auf der im vorherigen Schritt erstellten multidimensionalen Datenbank können nun Beziehungsgraphen automationsunterstützt aufgebaut werden, welche dann in einem Netzwerkdiagramm visuell dargestellt werden.

Knoten in einer Graphendatenbank würden Dimensionen darstellen, die Kanten die Beziehungen zwischen diesen. Gewichtungen ergeben sich aus der Häufigkeit der Beziehungen und der Gleich- oder Unterschiedlichkeit, gekennzeichnet durch den Manipulationsschutz.

Die Knoten im vorangegangenen Beispiel sind Mitarbeiter, die Größe der Knoten kennzeichnet deren Gehalt. Die Knoten haben untereinander Beziehungen „arbeitet für", umgesetzt als Attribut „fk_chief". Diese Kanten sind umso stärker als diesen ein Gewicht zugeordnet wurde, das sich aus einer (hier beispielhaft angenommenen) Spurenbewertung C0-C6 untereinander ergibt. Aus den Hierarchien und der Kantenstärke ergibt sich auch die Nähe im Graphen. Das Ergebnis stellt ein beispielhaftes „mafioses Netzwerk" dar. Technisch umgesetzt wurde dieses Beispiel mittels dem Open-Source Tool Gephi, indem die Daten aus dem Star-Schema importiert wurden.[14] Als Ergebnis ist ein Cluster rund um „Lex De Haan und Shelley Higgins" zu erkennen.

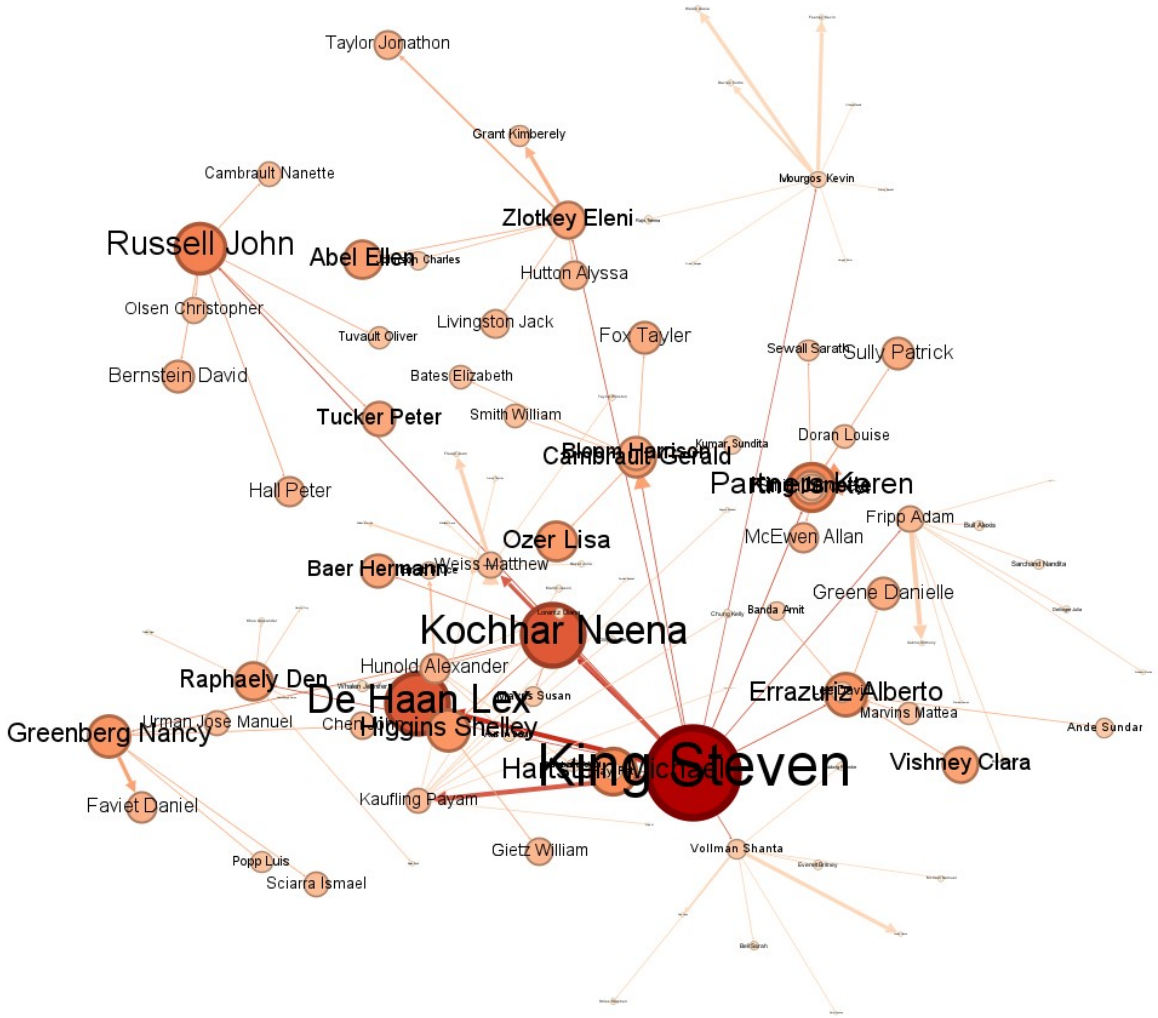

Abbildung 5: Beispiel einer Netzwerkanalyse am Sample „HR"

14 https://gephi.github.io/

4. Zusammenfassung und Ausblick

Das in dieser Publikation verwendete Beispiel ist natürlich gegenüber echten Fällen sehr einfach gestaltet. Es konnte jedoch gezeigt werden, wie solch eine Spurenbe- und Auswertung technischerseits durchgeführt werden kann. Die Methode des 7-stufigen Bewertungsmodells von Spuren wurde erweitert, sodass mehr Parameter berücksichtigt werden, und ein Zusammenhang der Spuren mit Tätern, Tatort und Tatzeit hergestellt werden kann.

Für die Bestimmung der Schuld oder Unschuld des Angeklagten muss mit an Sicherheit grenzender Wahrscheinlichkeit feststehen, dass kein Freispruchgrund vorliegt. Für einen Schuldspruch im Strafverfahren sind nach dieser Methode folgende zwei Stufen jedenfalls relevant: Fast sicher C5 (Beweise aus mehreren unabhängigen Quellen, die vor Manipulationen geschützt sind, stimmen überein) und Sicher C6 (Beweise sind manipulationssicher und/oder weisen eine hohe statistische Konfidenz auf).

Zuerst werden mit Erfahrung zielgerichtet Stichproben aus den inkriminierten Datenträgern gezogen, um eine mit an Sicherheit grenzender Wahrscheinlichkeit Zuordenbarkeit zum Tatverdächtigen herzustellen oder aufgrund einer Unwahrscheinlichkeit auszuschließen. Aufgrund der algorithmischen Vorgehensweise können die gefundenen Spuren einzeln kategorisiert und bewertet werden. Jede Spur hat eine Beziehung zu Dimensionen und beinhaltet als Attribute Messwerte. Als Dimensionen sind jedenfalls Tatzeit, Tatort, Täter und Manipulationsschutz zu nennen, als Messwerte kommen die 7 Bewertungsstufen in Frage. Mit Hilfe eines aus dem relationalen Datawarehouse bekannten modizierten Star-Schemas werden die erhobenen Spuren gespeichert, und stehen dann einer Analyse zur Verfügung. Aufbauend auf der multidimensionalen Datenbank können nun wie hier demonstriert Beziehungsgraphen aufgebaut werden, welche dann in einem Netzwerkdiagramm visuell dargestellt werden.

Diese Methodik ist bisher nicht publiziert worden und auch in keiner Software umgesetzt worden. Eine Möglichkeit wäre, bspw. für das auf dem Sleuth Kit[15] aufbauenden forensischen Tools Autopsy[16] ein Add-On Module zu implementieren, wo anhand der für dieses Tool zu Grunde liegenden relationalen Datenbank (das wäre SQLite[17] oder PostgreSQL[18]), die in dieser Publikation beschriebene Methodik und Auswertung umgesetzt wird.

15 https://www.sleuthkit.org/.
16 https://www.autopsy.com/.
17 https://sqlite.org.
18 https://www.postgresql.org/.