Informatieblad

Vernietigen van digitale sporen met verschoningsgerechtigde informatie

Inhoudsopgave

Inleiding

Er wordt momenteel veel gesproken over het vernietigen van data uit forensische bewijsbestanden, onder meer data die onder het verschoningsrecht valt. Maar wat betekent vernietigen van data? Kan dat wel? En wat zijn de gevolgen daarvan voor je bewijs? In dit informatieblad beantwoorden we deze vragen.

Dit informatieblad bestaat uit twee delen. Voor een goed begrip van de mogelijkheden en beperkingen voor het vernietigen van digitale sporen, zijn beide delen even belangrijk.

Het eerste deel gaat over het vernietigen van verschoningsgerechtigd materiaal in de praktijk. We richten ons eerst op de vraag hoe verschoningsgerechtigde informatie in de data geïdentificeerd wordt. Vervolgens leggen we uit wat technisch mogelijk maar ook onmogelijk is en hoe het vernietigen van data beperkingen legt op forensisch onderzoek en contraexpertise. Ook gaan we in op het alternatief van ontoegankelijk maken van verschonings-gerechtigde informatie, ook wel ‘uitgrijzen’ genoemd.

Het tweede deel geeft de technische context. We beschrijven wat data eigenlijk is en hoe data wordt vastgelegd in forensische bewijsbestanden. We leggen op hoofdlijnen uit

hoe data gestructureerd is en hoe voor mensen leesbare digitale sporen (bijvoorbeeld e-mails, chatberichten, foto's en documenten) in data worden gecodeerd. Ook beschrijven we welke forensische uitgangspunten gelden bij het doen van digitaal onderzoek. Tot slot geven we een toelichting op de complexiteit van het vinden duplicaten van sporen.

1. VERNIETIGEN IN DE PRAKTIJK

1.1. Identificeren van verschoningsgerechtigde informatie

Het proces van vernietigen van verschoningsgerechtigde informatie begint met het identificeren van sporen die zulke informatie kunnen bevatten zoals e-mails, chatberichten en documenten. Identificatie gebeurt in het algemeen door te zoeken op specifieke termen zoals contactgegevens van verschoningsgerechtigden[1]. Dit gebeurt bijvoorbeeld wanneer een verschoningsgerechtigde aangeeft dat hij/zij informatie heeft uitgewisseld met degene van wie de data is veiliggesteld.

Om zoeken op termen mogelijk te maken, wordt inbeslaggenomen of gevorderde data vaak eerst ingelezen in forensische tools als *Hansken*[2] of *FTK Lab*[3]. Deze tools verwerken de data en maken sporen (e-mails, chatberichten, documenten, enz.) inzichtelijk en doorzoekbaar. Afhankelijk van de hoeveelheid data, duurt dit enkele uren tot soms zelfs enkele weken. Tijdens of na deze verwerking kunnen de sporen met verschoningsgerechtigde informatie dan worden gefilterd.

Filtering op basis van zoektermen vormt een goede basis voor het vinden specifieke sporen. Het betekent echter niet dat door te filteren altijd exact de gewenste sporen worden gevonden.

Bij het filteren is het soms wel en soms niet wenselijk dat ook samenhangende sporen worden meegenomen. Bij een e-mail is het bijvoorbeeld logisch dat ook de bijlagen meegenomen worden, ook als de zoektermen hier *niet* in voorkomen. Als de e-mail in een e-maildatabase zit, zal de zoekterm ook in de database staan. De hele database filteren (en dus vernietigen), is echter onwenselijk, want dan worden álle e-mails in die database gefilterd, dus ook die e-mails die eigenlijk *niet* vernietigd moeten worden. Uiteindelijk is vaak een menselijke beslissing nodig voor ieder specifiek geval.

Daarnaast kunnen altijd duplicaten voorkomen die niet eenvoudig als duplicaat herkend worden, bijvoorbeeld omdat een ze (minimaal) zijn aangepast of omdat ze op een andere manier op een andere plek zijn vastgelegd (bijvoorbeeld een afbeelding die als bijlage in een verschoningsgerechtigde e-mail is ontvangen en vervolgens via een chat is gedeeld). Zie hoofdstuk 2.8.2 voor een uitgebreide toelichting en voorbeelden van duplicaten.

Deze filtering voorafgaand aan het onderzoek kan alleen als alle data beschikbaar is en als de gebruikte forensische tools de hele structuur van de data inzichtelijk kunnen maken. Dit is meestal niet het geval. Er komen namelijk continu nieuwe apparaten op de markt. Ook verschijnen er dagelijks nieuwe apps en worden deze regelmatig geüpdatet. Data wordt hierdoor steeds op andere manieren gestructureerd. *Hansken*, *FTK Labs* en andere forensische tools ondersteunen daarom nooit alle datastructuren. Er is dus altijd kans op missers. Met nieuwere versies van tools en de inzet van digitaal experts kunnen (later) in een onderzoek extra sporen inzichtelijk worden die mogelijk verschoningsgerechtigde informatie bevatten. Zie deel 2 voor een uitgebreide toelichting over de complexiteit van datastructuren.

Ook nieuwe informatie in een lopend onderzoek kan leiden tot nieuwe sporen. Zo kunnen nieuwe gegevensdrager (en dus data) in het onderzoek beschikbaar komen, maar bijvoorbeeld ook wachtwoorden of digitale sleutels kunnen worden achterhaald waarmee versleutelde data inzichtelijk gemaakt kan worden die al eerder in het onderzoek beschikbaar was[4]. Deze nieuwe of ontsleutelde data kan mogelijk ook weer verschoningsgerechtigde informatie bevatten.

Los van een eventuele technische oplossing, is het identificeren van sporen met verschoningsgerechtigde informatie een arbeidsintensief proces.

---

[1] Zie het `Voorlopig Beleid Uitspraak Kort Geding Verschoningsrecht', beschikbaar op https://www.om.nl/documenten/richtlijnen/2022/04/19/voorlopig-beleid-uitspraak-kort-geding-verschoningsrecht.

[2] Zie https://hansken.nl/ voor informatie over Hansken.

[3] Zie https://exterro.com/ftk-lab voor informatie over FTK Lab, voorheen AD Lab.

[4] Beschikbare versleutelde data kan wel in Hansken of FTK Lab geladen worden, echter zal hier zonder sleutel geen inzichtelijke tekst of sporen uit komen.

1.2. Technisch mogelijkheden en beperkingen

Het daadwerkelijk vernietigen van verschoningsgerechtigde informatie is om de meerdere technische redenen soms wel en soms niet mogelijk.

Het is technisch mogelijk om delen van de data die is vastgelegd in bewijsbestanden te vernietigen. Dit vereist wel dat de data *geheel* inzichtelijk is, bijvoorbeeld een overzichtelijke verzameling e-mails[5].

Voor bewijsbestanden met complexe datastructuren zoals data uit smartphones en computers, is dit niet altijd mogelijk. Individuele sporen kunnen onderdeel uitmaken van een ondeelbaar geheel, bijvoorbeeld een bestand in een ZIP-archief in een e-mailbijlage in een e-maildatabase. Dit enkele bestand kan niet vernietigd worden zonder de hele ZIP en soms zelfs de hele e-mail of de hele e-maildatabase te vernietigen. Zie deel 2 voor een uitgebreide toelichting op complexiteit van datastructuren.

Dit geldt ook voor sporen die voorkomen in een bewijsbestand dat enkel door een commercieel product inzichtelijk kan worden gemaakt, bijvoorbeeld een *UFDR report*[6]. De interne structuur van zo'n bewijsbestand kan dan niet worden aangepast, waardoor het hele bewijsbestand ondeelbaar is.

Om zulke sporen uit een onderzoek te verwijderen, moet het ondeelbare geheel verwijderd worden. Als de overige sporen behouden moeten blijven, dan moeten deze in een andere datastructuur in een nieuw bewijsbestand geplaatst worden, waarbij de bewijsketen voor die sporen wel in stand moet blijven. Ook moet dit bewijsbestand met de meest gebruikte tools verwerkt kunnen worden. Hiervoor is momenteel geen gestandaardiseerde oplossing.

1.3. Gevolgen van vernietigen voor het digitale bewijs

Digitaal-forensisch onderzoek vindt normaalgesproken plaats op een kopie van de data, vastgelegd in een bewijsbestand. Om zeker te zijn dat zo'n kopie bij het kopiëren niet verandert, worden controles uitgevoerd op basis van *bestandskenmerken*[7]. In principe worden bewijsbestanden nooit aangepast, zeker niet in de eerste originele kopie van de data die uit een gegevensdrager is gehaald.

Het vernietigen van sporen betekent dat bewijsbestanden *wel* aangepast moeten worden. Dit betekent dat de integriteit van deze bewijsbestanden niet meer met de oorspronkelijke bestandskenmerken gewaarborgd kunnen worden. Het is mogelijk om van aangepaste bewijsbestanden nieuwe bestandskenmerken vast te stellen, deze in een proces-verbaal vast te leggen en voortaan de integriteit van kopieën op basis van deze nieuwe bestandkenmerken vast te stellen. De koppeling met de originele kopie loopt dan altijd via dit proces-verbaal.

Tijdens de opsporing en bewijsvoering gebruiken digitaal experts verschillende tools. Om resultaten te kunnen verifiëren en het onderzoeksproces te valideren, worden resultaten van meerdere tools met elkaar vergeleken. De sporen in nieuwe bewijsbestanden moeten dus zo beschreven worden, dat bestaande (forensische) tools met dit nieuwe bewijsbestand overweg kunnen. Hiervoor is momenteel geen oplossing waarbij metadata en bewijsketens van de sporen in stand blijven.

Veel forensische tools draaien op de computer van de digitaal expert, die daar dus ook toegang moet hebben tot (een kopie van) de bewijsbestanden. Hierdoor zijn vaak meerdere kopieën van zulke bestanden in omloop. Ook het plaatsen in grotere forensische tools zoals *Hansken* of *FTK Lab* zorgt voor nieuwe kopieën. Het vernietigen van sporen uit één kopie, leidt dus niet automatisch tot het vernietigen uit andere kopieën.

Het vernietigen van sporen zal dus moeten plaatsvinden vóórdat er gedetailleerd onderzoek plaatsvindt op de data. Dit staat echter haaks op het feit dat dit voortschrijdend onderzoek tot nieuwe sporen kan leiden, die eigenlijk al vernietigd hadden moeten zijn. In de praktijk zal het vernietigen van sporen kunnen leiden tot het meerdere keren moet aanpassen van alle kopieën waar digitaal experts mee werken. Eventuele verwijzingen in processen-verbaal naar (unieke) sporennummers kunnen dan niet meer kloppen, omdat de sporennummers mogelijk wijzigen bij het opnieuw verwerken met bijvoorbeeld *Hansken* of *FTK Lab*.

---

[5] Ook bij overzichtelijke verzamelingen e-mails is het vinden van duplicaten lastig, bijvoorbeeld als doorgestuurd e-mails zijn opgenomen in de tekst van andere e-mails. Zie hoofdstuk 2.8.2 van de bijlage.

[6] Een *UFDR report* is een rapport dat met Cellebrite UFED en Physical Analyzer gemaakt kan worden en met Cellebrite Reader worden ingezien. Zie https://cellebrite.org/ voor meer informatie.

[7] Zie Vakbijlage 'Forensisch gebruik van bestandskenmerken en bijbehorende hashalgoritmen', beschikbaar op https://www.forensischinstituut.nl/over-het-nfi/vakbijlagen-en-informatiebladen.

Alternatief is dat er met één set van de bewijsbestanden wordt gewerkt. Na vernietiging (door aanpassing of een nieuw bewijsbestand) moeten de oorspronkelijke bewijsbestanden dan worden verwijderd en vervangen door de aangepaste of nieuwe bewijsbestanden. Dit is dan een vorm van destructief forensisch onderzoek. Het is daarna ook niet meer mogelijk te controleren wat er precies vernietigd is.

Forensische onderzoek naar bijvoorbeeld authenticiteit of herkomst van specifieke sporen is vaak niet meer mogelijk omdat dit plaatsvindt op de originele data. Ook andere forensische onderzoek zoals terugvinden van verwijderde bestanden in de brondata kan niet altijd meer plaatsvinden, omdat de brondata mogelijk niet meer beschikbaar is.

Dit alles geldt ook voor de mogelijkheden voor contraexpertise.

1.4. Vernietigen versus ontoegankelijk maken

Momenteel worden bewijsbestanden niet gewijzigd. Data worden toegankelijk gemaakt aan zaakonderzoekers via forensische onderzoeksomgevingen als *Hansken* of *FTK Lab*. Op basis van filters worden sporen deels automatisch en deels handmatig ontoegankelijk gemaakt. Dit wordt ook wel '*uitgrijzen*' genoemd. Dit betekent dat de sporen bij zoekslagen niet gevonden en dus niet getoond worden aan de zaakonderzoekers.

'Uitgrijzen' heeft in de praktijk ook beperkingen:

- Bij 'uitgrijzen' met behulp van een onderzoeksomgeving, blijven de sporen beschikbaar via andere tools die direct op de bewijsbestanden werken. Een bewijsbestand kan bijvoorbeeld wordt ingeladen in *Hansken* of *FTK Lab* en via die omgeving deel worden 'uitgegrijsd'. Bij het openen van dit bewijsbestand met in een andere forenische tool, zullen de 'uitgegrijsde' sporen mogelijk toch inzichtelijk zijn;
- Sommige combinaties van sporen worden voor de eindgebruiker als samenhangend gezien (bijvoorbeeld een e-mail met bijlagen). Voor tools zijn dit echter mogelijk losse sporen. Afhankelijk van de gebruikte tools en de door de gebruiker gekozen instellingen, wordt hier bij het 'uitgrijzen' wel of geen rekening mee gehouden;
- Exporteren/analyseren van data in bovenliggende structuren (bijvoorbeeld een hele e-maildatabase) blijft soms mogelijk, waardoor 'uitgegrijsde' sporen toch weer in een onderzoek terecht kunnen komen;
- De problemen met het identificeren van sporen die onder het verschoningrecht vallen, gelden hier ook;
- De problemen met duplicaten die later in een onderzoek naar voren kunnen komen, gelden hier ook.

1.5. Conclusie

Het is technisch mogelijk om delen van de data die is vastgelegd in bewijsbestanden te vernietigen. Dit vereist wel dat de data *geheel* inzichtelijk is.

Digitaal-forensisch onderzoek vindt normaalgesproken plaats verschillende kopieën van de data. Het vernietigen van sporen zal dus moeten plaatsvinden vóórdat er gedetailleerd onderzoek plaatsvindt en meerdere kopieën in omloop komen. Alternatief is dat er met één set van de bewijsbestanden wordt gewerkt, wat een beperking legt op het gebruik van verschillende forensische tools.

Voor bewijsbestanden met meer complexe datastructuren zoals data uit computers of smartphones, is dit niet altijd mogelijk. Uit zulke data kunnen meestal niet de individuele sporen (e-mails, chatberichten, documenten) verwijderd worden, maar alleen grotere gehelen (hele databases). In die gevallen kan met een goed gedocumenteerd proces wel een deel van die onnodig verwijderde sporen behouden blijven. Er is momenteel echter geen oplossing zodat de bewijsketens en metadata van die sporen behouden blijft. Ook kunnen die sporen niet meer met alle tools onderzocht worden. Forensisch onderzoek naar bijvoorbeeld de herkomst of authenticiteit van sporen die op deze manier behouden zijn, is dan ook niet altijd meer mogelijk. Dit geldt ook voor contraexpertise.

Vernietigen en ontoegankelijk maken van digitale sporen hebben verder vergelijkbare beperkingen, zoals het vinden van duplicaten, het omgaan met onbekende datastructuren en voortschrijdend inzicht in techniek en binnen een zaakonderzoek. In beide gevallen is het goed identificeren van sporen met verschoningsgerechtigde informatie een arbeidsintensief proces. Ook kan geen garantie gegeven worden over de volledigheid van het werken met filters. Er kunnen altijd sporen worden gemist, of onterecht worden achtergehouden.

2. TECHNISCHE CONTEXT

2.1. Data

Digitale data zijn niet meer en niet minder dan (lange) reeksen *bits*, waarbij een bit een 0 of een 1 is. Een reeks van 8 bits wordt een *byte* genoemd. Dit is de kleinste eenheid waarin data worden vastgelegd. Tabel 1 geeft de naamgevingen van oplopende lengtes.

Tabel 1 naamgeving bits en bytes

1 bit	een 0 of een 1
1 byte	reeks van 8 bits, de eenheid voor data
1 kilobyte (kB)	reeks van 1.000 bytes
1 megabyte (MB)	reeks van 1.000.000 bytes (1.000 kB)
1 gigabyte (GB)	reeks van 1.000.000.000 bytes (1.000 MB)
1 terabyte (TB)	reeks van 1.000.000.000.000 bytes (1.000 GB)

Computers hebben momenteel een opslagcapaciteit van enkele terabytes. Smartphones hebben een capaciteit van tientallen gigabytes tot enkele terabytes. USB-sticks en 'geheugenkaartjes' heb je in alle maten, van enkele megabytes tot enkele terabytes.

Hexadecimale notatie

Voor het werken met en opschrijven van reeksen bytes is het niet handig om bytes op te schrijven als lange reeksen nullen en enen. De hexadecimale notatie wordt het meest gebruikt. Het hexadecimaal stelsel, ofwel 16-tallig stelsel gebruikt de karakters 0 tot en met 9, aangevuld met de zes letters A tot en met F. Met deze (2 x 2 x 2 x 2 =) 16 karakters kunnen reeksen van 4 bits worden beschreven:

0000 **0**	0100 **4**	1000 **8**	1100 **C**
0001 **1**	0101 **5**	1001 **9**	1101 **D**
0010 **2**	0110 **6**	1010 **A**	1110 **E**
0011 **3**	0111 **7**	1011 **B**	1111 **F**

Een byte (8 bits) kan worden beschreven met 2 karakters, zoals 4A voor de byte 0100 1010.

2.2. Brondata in bewijsbestanden

Brondata in een digitaal onderzoek is gedefinieerd als de dataverzamelingen die uit inbeslaggenomen *gegevensdragers*[8] zijn veiliggesteld of zijn gevorderd bij bijvoorbeeld providers.

Gevorderde data wordt door verschillende organisaties op verschillende manieren aangeleverd. Brondata in gegevensdrager wordt veiliggesteld in *bewijsbestanden*, ook wel *images* genoemd. Zo'n bewijsbestand bevat naast de veiliggestelde brondata vaak ook gegevens over het veiligstellen, zoals informatie over de gegevensdrager waar de data uit is gehaald, het tijdstip van veiligstellen en de betrokken personen. Er bestaan twee soorten bewijsbestanden: fysieke en logische.

Een *fysiek bewijsbestand*, ook wel *fysieke kopie* of *één-op-één-kopie* genoemd, bevat de exacte reeks bytes die op een gegevensdrager staat. Een fysieke kopie van een harde schijf waar 2 terabytes data op past, bevat dan ook dezelfde 2 terabytes data.

In sommige gevallen kan er geen fysieke kopie gemaakt worden van een gegevensdrager, bijvoorbeeld wanneer het apparaat (deels) is beveiligd of vercijferd. Dit is vaak het geval bij smartphones. Met behulp van (commerciële) apparatuur worden dan de beschikbare data uit het apparaat gehaald en vastgelegd in een *logisch bewijsbestand*, ook wel *logische kopie* of *report* genoemd. Zo'n kopie is dus typisch kleiner dan de opslagcapaciteit van de gegevensdrager. De mogelijkheden voor het maken van een logische kopie en de structuur van de logische kopie hangt af van zowel het soort en type gegevensdrager als van de gebruikte apparatuur waarmee de kopie wordt gemaakt.

Om de integriteit van brondata in bewijsbestanden te borgen, worden bij het veiligstellen *bestandskenmerken*[9] van de brondata berekend, ook wel *secure hashes* of *digests* genoemd. Deze bestandskenmerken zijn reproduceerbaar en worden vaak ook in het bewijsbestand vastgelegd.

2.3. Digitale sporen

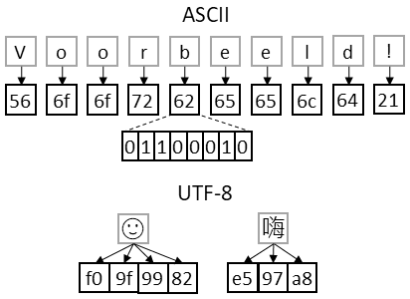
Brondata bevatten grote hoeveelheden gegevens, in forensisch onderzoek *digitale sporen* genoemd. Een deel van deze gegevens zijn door mensen gemaakt, voorbeeld foto's, documenten, e-mails en chatberichten. Een ander deel is technisch, bijvoorbeeld logbestanden of systeeminstellingen, en weer een ander deel bevat sensorregistraties zoals locatiegegevens of getelde stappen. Ook combinaties zijn mogelijk, bijvoorbeeld afbeeldingen met locatiegegevens.

[8] Zie Technische toelichting 'Terug naar de bestanden', beschikbaar op https://www.forensischinstituut.nl/over-het-nfi/vakbijlagen-en-informatiebladen.

[9] Zie Vakbijlage 'Forensisch gebruik van bestandskenmerken en bijbehorende hashalgoritmen', beschikbaar op https://www.forensischinstituut.nl/over-het-nfi/vakbijlagen-en-informatiebladen.

Al deze gegevens zijn op een of andere manier digitaal gecodeerd en gestructureerd, oftewel “vertaald naar reeksen nullen en enen”. De *codering* van de data –het vastleggen van gegevens in een reeks bytes– en de *structuur* van de data – hoe deze reeksen bytes gecombineerd worden– bepaalt dus wat de data betekent.

2.4. Datacodering

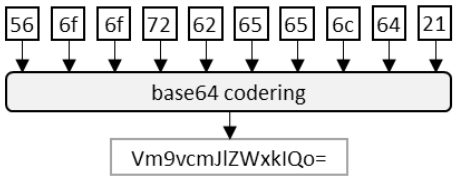


Figuur 2 Voorbeelden van ASCII en UTF-8 codering

Dezelfde gegevens kunnen op heel veel verschillende manieren in reeksen bytes worden vastgelegd.

Sommige coderingen, bijvoorbeeld ASCII[10], zijn relatief eenvoudig, waar iedere byte een normaal karakter (bijvoorbeeld een hoofdletter, kleine letter, cijfer of leesteken) codeert. Er zijn ook coderingen zoals UTF-8[11] waar ieder karakter in 1 tot 4 bytes wordt gecodeerd. Dit is nodig om niet-westerse karakters (zoals Chinees of Arabisch) te kunnen coderen, maar bijvoorbeeld ook alle emoticons. Figuur 2 toont een voorbeeld van codering van de tekst `voorbeeld!’, een smiley en een Chinees karakter.

Er zijn tal van coderingen voor meer specifiek gebruik. Bijvoorbeeld Base64[12], om willekeurige digitale data (reeksen bytes) om te zetten in ASCII (cijfers en letters en enkele speciale tekens, zie Figuur 1). Zo kan data worden vastgelegd in en overgedragen tussen systemen die alleen ASCII-tekens ondersteunen (bijvoorbeeld voor het opnemen van een afbeelding in een e-mail).



Figuur 1 Voorbeeld van base64 codering

2.5. Datastructuren

Naast de vele coderingen, bestaan er nog meer mogelijkheden om de gecodeerde gegevens te combineren.

2.5.1. Bestanden en bestandssystemen

Een gegevensdrager wordt standaard gestructureerd volgens een *bestandssysteem*[13], een basisstructuur om mappen en bestanden te beschrijven. Hiervoor is keuze uit tientallen structuren, bijvoorbeeld FAT32 (vaak gebruikt voor USB-sticks), NTFS (voor computers met Windows), Ext4 (voor Linux en Android) of APFS (voor Apple apparaten zoals iPhones en Macbooks). Deze bestandssystemen hebben een zogenaamde *bestandstabel* waarin de gegevens over de mappen en bestanden worden vastgelegd. Ook wordt hierin vastgelegd waar de data van ieder bestand op de gegevensdrager staat. Zulke gegevens over data wordt ook wel *metadata* genoemd.

Ieder bestand zelf heeft ook een structuur, ook wel *bestandsformaat* genoemd. Deze structuren kunnen relatief eenvoudig zijn (bijvoorbeeld een platte tekst of een logbestand in ASCII-codering), iets complexer (bijvoorbeeld een afbeelding of een document met opmaak) maar ook zeer complex (bijvoorbeeld een database met e-mails en bijlagen). Soms vormen meerdere bestanden samen weer één structuur (bijvoorbeeld een Multipart ZIP).

10 Zie bijvoorbeeld https://nl.wikipedia.org/wiki/ASCII_(tekenset) voor voorbeelden.

11 Bijvoorbeeld UTF-8, zie https://nl.wikipedia.org/wiki/UTF-8 voor meer informatie.

12 Zie bijvoorbeeld https://nl.wikipedia.org/wiki/Base64 voor meer informatie.

13 Zie sectie 3.3 van Technische toelichting 'Terug naar de bestanden’, beschikbaar op https://www.forensischinstituut.nl/over-het-nfi/vakbijlagen-en-informatiebladen.

Tabel 2 Database met contacten en berichten

contacten

ID	naam	nummer
1	Alice	0612345678
2	Bob	0623456789
3	Carol	0634567890

berichten

type	datum &tijd	contact ID	bericht
binnenkomend	2023-03-19 11:00:23	1	Goedemorgen! Hoe gaat het?
uitgaand	2023-03-19 11:02:18	1	Het gaat goed. Met jou?
uitgaand	2023-04-06 11:04:22	2	Vanmiddag kan ik niet

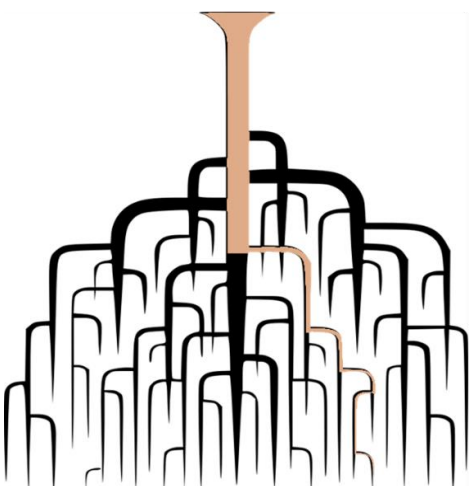
2.5.2. Databases

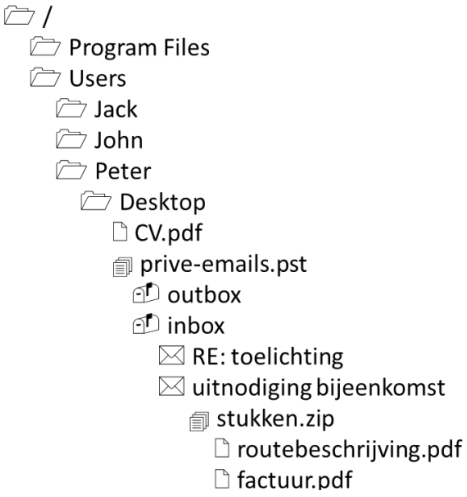
Veel programma's en apps leggen gegevens vast. Dit kan als losse bestanden (bijvoorbeeld de foto's gemaakt met een camera app). Vaak zijn dit echter gestructureerde gegevens, zoals contactpersonen, e-mails, chatcommunicatie of browsergeschiedenis. Zulke gegevens worden vastgelegd in *relationele databases*[14] die zijn opgebouwd uit tabellen. Een tabel bestaat uit rijen en kolommen, waarbij iedere rij een (deel van een) spoor vastlegt. Soms worden gegevens in verschillende tabellen aan elkaar gekoppeld. Zie bijvoorbeeld Tabel 2 voor een database met een tabel met contacten en een tabel voor de uitgewisselde berichten met die contacten.

Er zijn veel verschillende manieren om relationele databases digitaal te coderen. Veelgebruikte database-coderingen zijn MySQL, PostgrSQL en SQLite[15].

2.6. Sporenbomen

Sporen zelf kunnen andere sporen bevatten. Ieder bestand, bijvoorbeeld een e-maildatabase, is dus zelf een spoor, maar ook iedere e-mail en bijlage in zo'n database is een spoor met een eigen datastructuur. De data uit een gegevensdrager vormen op deze manier een grote boom met digitale sporen, die bestaat uit mappen met bestanden, databases met bijvoorbeeld contacten, e-mails of chatberichten, of bijvoorbeeld afbeeldingen met locatiegegevens. Figuur 3 toont een voorbeeld van een sporenboom. Een typische computer of smartphone bevat zo al snel enkele miljoenen sporen.

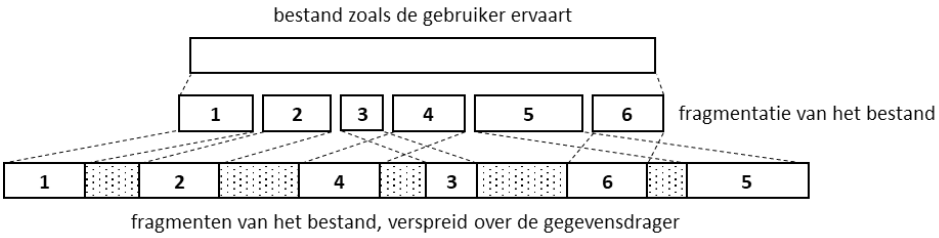




Figuur 3 Voorbeeld van een sporenboom

[14] Zie bijvoorbeeld https://nl.wikipedia.org/wiki/Relationele_database voor meer informatie.

[15] De afkorting SQL in de naamgeving van databases refereert naar Structured Query Language, de standaardtaal die gebruikt wordt om met gegevens in relationele databases te werken (schrijven, lezen, aanpassen, verwijderen). Zie https://nl.wikipedia.org/wiki/SQL voor meer informatie.



Figuur 4 Voorbeeld van fragmentatie

2.7. Forensisch onderzoek aan data

Om forensisch onderzoek te kunnen uitvoeren, worden de sporenbomen met forensische tools gereconstrueerd. Vervolgens wordt, op basis van hypothesen en vragen, onderzoek aan de sporen uitgevoerd. Bijvoorbeeld naar de vraag hoe een afbeelding op een smartphone terecht is gekomen of met welk apparaat een e-mail verstuurd is.

Voor beantwoording van onderzoeksvragen worden vaak veel verschillende sporen geanalyseerd en gecombineerd, bijvoorbeeld instellingen en log-bestanden van betrokken apps en apparaten, maar ook vergelijkbare sporen of sporen die net vóór of ná de te onderzoeken sporen zijn vastgelegd.

2.7.1. Forensische waarborgen

Een digitaal-forensisch onderzoek richt zich vaak op het beantwoorden van vragen over de herkomst, authenticiteit en betekenis van digitale sporen. Zo'n onderzoek moet transparant zijn, dat wil zeggen dat resultaten herleidbaar moeten zijn tot het bronmateriaal (chain of evidence) en dat duidelijk is welke onderzoekstappen precies hebben plaatsgevonden (chain of custody).

2.7.2. Datastructuren ontvlechten en data decoderen

De reconstructie van sporenbomen met forensische tools is vrijwel altijd onvolledig. Voor ieder spoor dat is vastgelegd, geldt namelijk dat om tot de data van dat spoor te komen, er meerdere lagen van decodering en ontvlechting van de datastructuren nodig is. Iedere stap in dit complexe proces kan bestaan uit het bij elkaar zoeken van verschillende fragmenten van de data, decompressie, decryptie en het toepassen van complexe algoritmes. Dit is niet altijd eenvoudig, te meer omdat niet alle gebruikte algoritmes bekend en/of beschikbaar zijn.

Fragmentatie

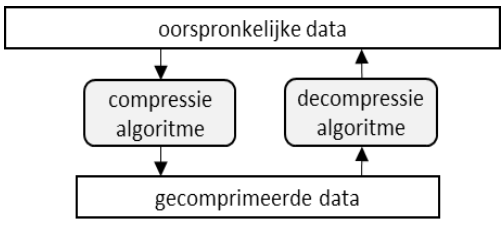
De inhoud van een gegevensdrager verandert continue door het gebruik, bijvoorbeeld bij het sturen van e-mails of chatberichten, bij het maken van foto's maar zelf ook al bij het maken van bewegingen. Als een apparaat 'aan' staat, kan er op de achtergrond ook van alles gebeuren zonder directe betrokkenheid van de gebruiker (bijvoorbeeld binnenkomende berichten of kloksynchronisaties).

De datastructuren waarin gegevens worden vastgelegd (bijvoorbeeld bestandssystemen en databases) zijn zo opgezet, dat gegevens snel en efficiënt worden weggeschreven op gegevensdragers. Data die samen één spoor vormt, kan hierbij worden opgeknipt en in verschillende delen in de datastructuur terecht komen, mogelijk ook in een andere volgorde (zie Figuur 4). Dit wordt *fragmentatie* genoemd. Dit geldt voor bestanden hoog in de sporenboom, maar bijvoorbeeld ook voor sporen in databases die over verschillende tabellen zijn verspreid.

Compressie

Omdat opslagruimte vaak beperkt is en het over een netwerk versturen van veel data veel tijd kost, is het efficiënt om gegevens in zo min mogelijk bytes te coderen, ofwel de data te *comprimeren*. Hiervoor bestaan tientallen compressiealgoritmes, veelgebruikt zijn ZIP voor bestandsarchieven of GZIP voor het efficiënt versturen van data. Dit zijn complexe algoritmes, waardoor bytes in de oorspronkelijke data niet te koppelen zijn aan bytes in de gecomprimeerde data.

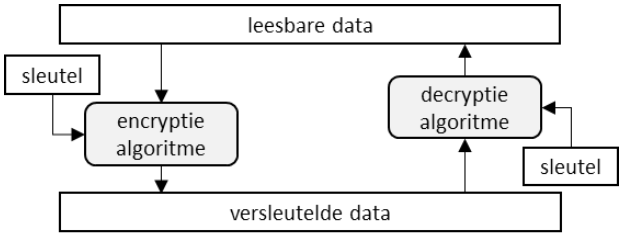
Compressie kan worden gebruikt binnen structuren –veel databases gebruiken compressie– maar ook vaak hoger in de sporenboom. Bestandssystemen kunnen zo worden ingesteld dat bijvoorbeeld alle bestanden, enkele bestanden of zelfs delen van bestanden gecomprimeerd worden opgeslagen.



Figuur 5 Voorbeeld van compressie

Encryptie

Soms is het wenselijk om data te beschermen. *Encryptie*, ook *vercijfering* of *versleuteling* genoemd, is een techniek waarbij de te beschermen data op zo'n manier wordt gecodeerd, dat er extra informatie nodig is om deze te decoderen. Deze extra informatie (bijvoorbeeld een wachtwoord of digitale sleutel) wordt dan ergens anders bewaard. Ook hiervoor zijn veel verschillende complexe algoritmes beschikbaar. Deze algoritmes zijn tegenwoordig zo sterk, dat zonder de extra informatie (wachtwoord of sleutel) decodering onmogelijk is.

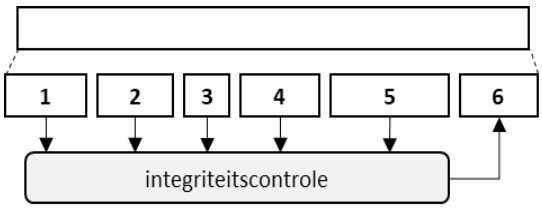


Figuur 6 Voorbeeld van encryptie

Net als compressie, kan encryptie worden gebruikt binnen structuren –bijvoorbeeld PGP-versleutelde e-mails– maar ook vaak hoger in de sporenboom. Hele schijven of bestandssystemen kunnen worden versleuteld. Zo kan encryptie tegelijkertijd worden toegepast op meerdere niveaus in de sporenboom.

Integriteitscontroles en herstelcodes

Complexe structuren, vooral databases, hebben interne integriteitscontroles en herstelcodes, zodat als er tijdens of na het wegschrijven op een gegevensdrager iets niet helemaal goed gaat, de vastgelegde gegevens toch nog kunnen worden hersteld.



Figuur 7 Voorbeeld van integriteitscontrole

Ook op hogere niveaus zijn controles mogelijk. In financiële administraties bijvoorbeeld, moeten de geregistreerde bij- en afschrijvingen in een database kloppen met de balans. Zulke controles kunnen onderdeel uitmaken van de algoritmes die worden gebruikt bij het coderen en decoderen van data.

Onbekende datastructuren

Veelvoorkomende structuren zijn over het algemeen gedocumenteerd en er is (forensische) software om de ze inzichtelijk te maken. Dit geldt ook voor veelvoorkomende coderingen. Deze software is dan publiek of commercieel beschikbaar, soms zonder en soms met de broncode.

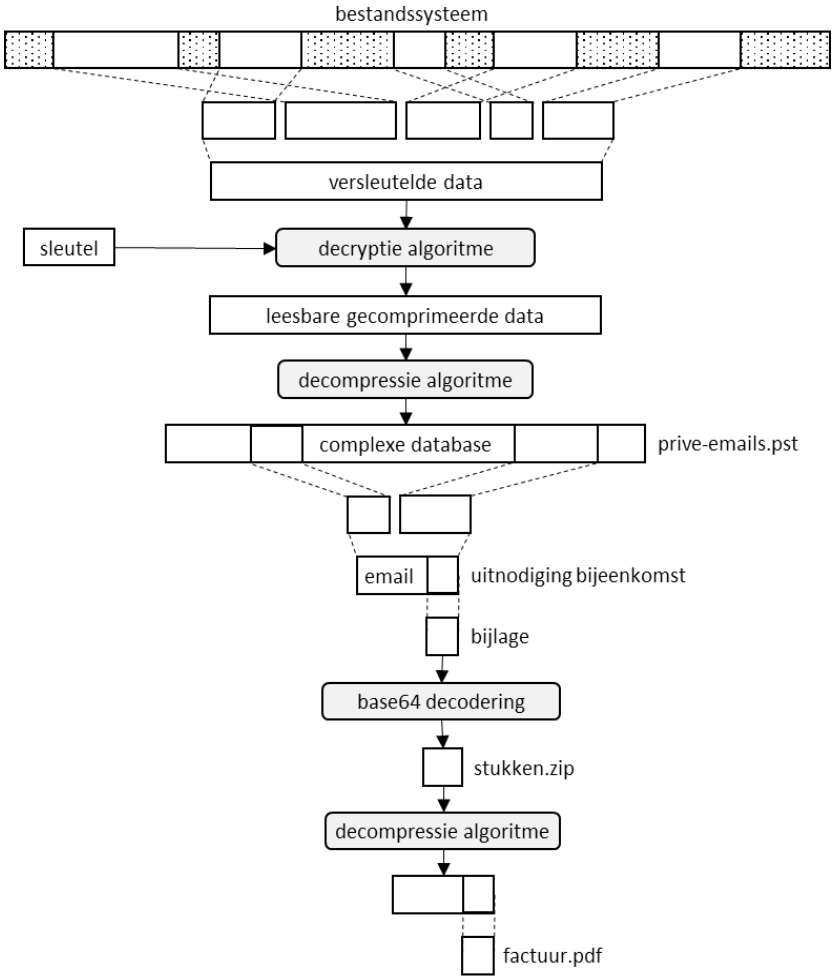
Er zijn echter ook veel structuren die niet gedocumenteerd zijn, bijvoorbeeld omdat ze door ontwikkelaars bedacht zijn voor het gebruik in één enkele app. Soms zijn ze ook bewust niet vrijgegeven om bijvoorbeeld security redenen of omdat er intellectueel eigendom op zit. In deze gevallen kan met (tijdrovende) experimenten en *reverse engineering* geprobeerd worden de coderingen en datastructuren te achterhalen. Dit kan bijvoorbeeld door bekende gegevens met de app vast te leggen, en vervolgens te kijken of en waar deze gegevens terug te vinden zijn in de door de app vastgelegde data. Dit lukt soms wel, soms niet, of slechts voor een gedeelte van de data. Het is eventueel ook mogelijk deze data in te lezen met de oorspronkelijke software waarmee de data gecreëerd is.

Onvolledige data

Een ander uitdaging is dat onvolledige data (gedeeltelijke datastructuren) ook potentieel interessante sporen kunnen bevatten. Denk bijvoorbeeld aan gedeeltelijk verwijderde e-maildatabases. Deze structuren zijn zelden te open met oorspronkelijke software, hiervoor zijn specialistische forensische tools en interpretatie door deskundigen/experts nodig.

Een voorbeeld

Figuur 8 laat zien hoe complexe geneste datastructuren werken op basis van het voorbeeld van een sporenboom in Figuur 3. Dit voorbeeld toont de complexe datastructuur en codering van slechts één spoor in een sporenboom met vaak miljoenen sporen.



Figuur 8 voorbeeld van geneste datastructuren

2.7.3. Verwijderde gegevens terughalen

De beschikbare opslagruimte van een gegevensdrager wordt zelden volledig gebruikt voor bestanden. Er is vrijwel altijd vrije ruimte over. Deze vrije ruimte, ook wel *unallocated space* genoemd, bestaat ook "gewoon" uit reeksen nullen en enen. Als deze ruimte is opgenomen in een (fysiek) bewijsbestand, dan kan deze ook doorzocht worden op bekende structuren en coderingen. Dit doorzoeken noemen we *carven*. Zo kunnen (delen van) verwijderde bestanden worden teruggevonden. Ook binnen bijvoorbeeld databases kan op deze manier worden gezocht naar verwijderde gegevens.

2.7.4. Onderzoek aan metadata

Zoals eerder toegelicht, bestaan sporen uit data (de inhoud van het spoor zelf) en metadata (informatie over het spoor). Deze metadata kan iets zeggen over het spoor zelf (bijvoorbeeld de breedte en hoogte van een afbeelding), kan afkomstig zijn uit de structuur waarin het spoor is aangetroffen (bijvoorbeeld de naam van een bestand uit een bestandstabel) maar kan ook gegevens over de omgeving van het spoor beschrijven (bijvoorbeeld dat een database hoort bij een specifieke app). Al deze informatie kan relevant zijn om onderzoeksvragen over bijvoorbeeld authenticiteit of herkomst te beantwoorden.

Met metadata kunnen sporen vaak ook aan elkaar gekoppeld worden. Op basis van metadata van e-mails, kan bijvoorbeeld een groot communicatienetwerken inzichtelijk gemaakt worden. Een ander voorbeeld van het gebruik van metadata, is het identificeren van levenspatronen op basis van locatie en -tijdregistraties, zodat vast kan worden gesteld wanneer iemand normaliter op zijn werk of juist thuis is.

2.8. Identificeren van sporen

Om sporen te vernietigen, moeten deze eerst worden gevonden. Dit kan bijvoorbeeld door te filteren op exclusielijsten met zoektermen zoals contactgegevens van verschoningsgerechtigden.

Inclusielijsten en exclusielijsten

Bij het filteren van sporen wordt de verzameling sporen in tweeën gesplitst, vergelijkbaar met een zeef: sporen die voldoen aan het filter blijven in de zeef achter, sporen die niet voldoen vallen erdoorheen.

Bij *black listing* worden de sporen die voldoen aan het filter weggegooid en blijft de rest over (die dus door de zeef vallen). Als dit gebeurt op basis van een lijst met zoektermen, wordt deze lijst een ex*clusielijst* genoemd.

Bij *white listing* blijven alleen de sporen die voldoen aan het filter over (die dus in de zeef achterblijven). Als dit gebeurt op basis van een lijst met zoektermen, wordt deze lijst een *inclusielijst* genoemd.

In beide gevallen bestaat de kans op vals-positieven en vals-negatieven.

Om selectie op basis van zoektermen mogelijk te maken, worden bewijsbestanden vaak eerst ingelezen in forensische tools als *Hansken* of *FTK Lab*. Deze tools verwerken de bewijsbestanden, waarbij de data wordt gedecodeerd en ontvlochten. Tijdens of na het verwerken, kunnen te vernietigen sporen dan worden geïdentificeerd.

Als er wordt gewerkt met exclusielijsten, zullen er vrijwel altijd sporen zijn die onterecht zijn tegengehouden of doorgelaten. Dit heeft meerdere oorzaken die nader worden toegelicht:

- Voortschrijdende techniek en zaakonderzoek;
- Het voorkomen van mogelijk aangepaste of anders gestructureerde duplicaten;
- De samenhangen tussen verschillende sporen.

2.8.1. Voortschrijdende techniek en zaakonderzoek

Er komen continu nieuwe apparaten op de markt. Ook verschijnen er dagelijks nieuwe apps en worden deze regelmatig geüpdatet. Data wordt hierdoor regelmatig op andere manieren gecodeerd en gestructureerd. Tools ondersteunen daarom nooit alle coderingen en structuren. Er is dus altijd kans op missers. Met nieuwere versies van deze of andere tools en de inzet van digitaal experts kunnen (later) in een onderzoek extra sporen inzichtelijk worden, waaronder mogelijk te vernietigen sporen.

Ook nieuwe informatie in een lopend onderzoek kan leiden tot nieuwe sporen. Wanneer nieuwe apparaten (en dus bewijsbestanden) in het onderzoek beschikbaar komen, maar bijvoorbeeld ook als nieuwe wachtwoorden of digitale sleutels worden achterhaald waarmee versleutelde data inzichtelijk gemaakt kan worden. Deze nieuwe of ontsleutelde data kan mogelijk ook weer te vernietigen sporen bevatten.

2.8.2. Duplicaten

Zowel binnen een bewijsbestand als in verschillende bewijsbestanden kan hetzelfde spoor voorkomen. Deze zijn vaak technisch verschillend maar voor de eindgebruiker hetzelfde.

Bij het normaal gebruik van een apparaat worden gegevens vaak op meerdere plaatsen opgeslagen. Als een foto bijvoorbeeld wordt gedeeld via een chat app, wordt deze verkleind en op een andere plek opgeslagen. Als een e-mailbijlage wordt geopend, wordt deze eerst opgeslagen als bestand in een bestandssysteem. Ook worden gegevens bijvoorbeeld vastgelegd in herstelbestanden van tekstverwerkers of transactielogs van databases. Dit gebeurt om te voorkomen dat gegevens verloren gaan of corrupt[16] raken als een app bijvoorbeeld vastloopt of een apparaat plotseling uitvalt.

Identieke data

De manier om te achterhalen of twee sporen identiek zijn, is door eenvoudigweg te controleren of de reeks bytes gelijk is. Omdat deze reeksen erg lang kunnen zijn, gebeurt dit in de praktijk door het berekenen en vergelijken van *secure hashes*. Deze manier van vergelijken werkt alleen als de gegevens op dezelfde manier gecodeerd en gestructureerd zijn.

Verschillende metadata

Een e-mail in de mailbox van de ontvanger is vanuit de gebruiker gezien dezelfde e-mail als die in de mailbox van de verzender. De inhoud van het bericht is identiek en kan meestal vergeleken worden op basis van *secure hashes*.

De metadata van deze e-mails kan echter verschillen, bijvoorbeeld omdat de e-mail bij de ontvanger informatie bevat over de afgelegde route over het internet, of omdat deze e-mail gescand is op spam of virussen. Als er meerdere ontvangers zijn, zal de metadata bij de verschillende ontvangers ook verschillen.

---

16 Corrupte bestanden zijn bestanden waarvan de structuur niet meer klopt, bijvoorbeeld omdat bij het opslaan iets fout is gegaan. Ook hele bestandssystemen kunnen corrupt raken, waardoor de bestanden niet meer gelezen kunnen worden.

Hetzelfde geldt voor sms'jes en chatberichten. Als de ontvanger een bericht krijgt van een onbekend persoon, staat er geen naam maar een telefoonnummer als afzender. Als de afzender in het adresboek staat onder een andere naam, wordt deze andere naam mogelijk bij het bericht vastgelegd.

Dit geldt ook voor bestanden, die op inhoud gelijk zijn, maar bijvoorbeeld door het kopiëren een andere naam hebben gekregen. Of als een bestand als e-mailbijlage is ontvangen en op een andere plaats is opgeslagen. De data van deze bestanden kunnen vaak wel op basis van *secure hashes* vergeleken worden.

Verschillende coderingen en structuren

Vergelijken wordt lastiger als de coderingen en structuren daadwerkelijk anders zijn. Een afbeelding die is gemaakt met een smartphone bijvoorbeeld, wordt bij het versturen via een chatbericht vaak verkleind. Deze afbeelding is hierna voor de gebruiker identiek, maar technisch gezien totaal verschillend. Om zulke sporen met elkaar te kunnen vergelijken zijn andere technieken nodig, bijvoorbeeld *perceptuele vingerafdrukken*, in het Engels *perceptual hashes*[17] genoemd. Afbeeldingen die vanuit de perceptie van de gebruiker sterk op elkaar lijken hebben vergelijkbare *perceptual hashes*. Deze techniek wordt vooral toegepast op multimedia-bestanden zoals foto's, video's en geluidsopnames.

Meerdere versies

Vooral voor documenten (maar ook voor andere sporen) geldt dat er vaak meerdere versies terug te vinden zijn in een bewijsbestand. Deze kopieën kunnen bewust gemaakt zijn door de gebruiker (bijvoorbeeld als backup), het gevolg zijn van een handeling met het document (bijvoorbeeld geprint of een kopie verstuurd als e-mailbijlage), of het gevolg van het gedrag van een systeem (bijvoorbeeld een herstelbestand). Deze documenten zijn vaak vergelijkbaar, maar niet identiek. Afhankelijk van de structuur en soort wijzigingen, kunnen meerdere versies teruggevonden worden.

Gedeeltelijk vergelijkbare sporen

Het kan voorkomen dat zowel de inhoud (data) als de metadata van twee sporen niet overeenkomen, maar deze sporen voor de gebruiker toch (deels) gelijk zijn. Een voorbeeld is een doorgestuurde e-mail. De inhoud van de doorgestuurde e-mail wordt dan in tekst de nieuwe e-mail opgenomen, vaak voorafgegaan door een kopje met metadata over de eerder e-mail, zoals het onderwerp, de afzender en de ontvangers. Voor de gebruiker is de doorgestuurde e-mail een kopie van het origineel. Vanuit de techniek gezien, is dit echter niet meer dan een stukje tekst in de nieuwe e-mail.

Zeker als bij het doorsturen gegevens zijn toegevoegd of aangepast, is het erg lastig om de doorgestuurde e-mail te herkennen als duplicaat.

2.8.3. Samenhangende sporen

Bij het identificeren van te vernietigen sporen, bijvoorbeeld op basis van zoektermen, is het soms wel en soms niet wenselijk dat ook onderliggende sporen in de sporenboom worden meegenomen. Bij een e-mail is het bijvoorbeeld logisch dat ook de bijlagen meegenomen worden, ook als de zoektermen hier *niet* in voorkomen. Als de e-mail in een e-maildatabase zit, zal de zoekterm ook in de database staan. De hele database vernietigen is echter onwenselijk, want dan worden álle e-mails in die database ook vernietigd, dus ook die e-mails die eigenlijk *niet* vernietigd moeten worden.

Hier kan op verschillende manieren mee omgegaan worden, waarbij uiteindelijke vaak een menselijke beslissing nodig is voor ieder specifiek geval.

Onbekende samenhang

Het kan voorkomen, dat sporen niet meer samenhangen, die voorheen wel samenhingen en door die eerdere samenhang wel onder het verschoningsrecht vallen. Een voorbeeld is het volgende scenario:

Een advocaat stuurt een e-mail (die dus onder het verschoningsrecht valt), naar een cliënt, die hem ontvangt op zijn smartphone. Deze e-mail bevat een bijlage, die dus ook onder het verschoningsrecht valt. Deze bijlage wordt geopend, waardoor een kopie van de bijlage in de Downloads *map terecht komt. De cliënt verwijdert de e-mail met bijlage, de gedownloade bijlage blijft achter. Dit bestand bevat zelf geen kenmerken waaruit blijkt dat deze van de advocaat afkomstig is.*

Als deze smartphone in een onderzoek wordt uitgelezen, dan wordt bij het toepassen van filters, deze gedownloade bijlage nooit gemarkeerd. Er kunnen dus sporen voorkomen, waarvoor geldt dat het op basis van de sporen zelf onmogelijk is om vast te stellen dat ze onder het verschoningsrecht vallen.

---

[17] Zie bijvoorbeeld https://en.wikipedia.org/wiki/Perceptual_hashing voor meer informatie over perceptual hashes.

Voor algemene vragen kunt u contact opnemen met de Frontdesk, telefoon (070) 888 68 88.

Voor inhoudelijke vragen kunt u contact opnemen met dr.ir. Harm van Beek, NFI-deskundige Forensische Digitale Technologie met specialisatie data-analyse van de afdeling Digitale en Biometrische Sporen.

telefoon (070) 888 6400.

Nederlands Forensisch Instituut

Ministerie van Justitie en Veiligheid

Postbus 24044 | 2490 AA Den Haag

Telefoon (070) 888 66 66

www.forensischinstituut.nl

april 2023.