Australian Journal of Forensic Sciences, Volume 55, Issue 5, September 2023, pp. 689-698
DOI:10.1080/00450618.2022.2071988

Considering 'technically possible' alternative meanings for data traces found during a digital forensic examination.

Dr Graeme Horsman
School of Computing, Engineering & Digital Technologies
Teesside University
graeme.horsman@googlemail.com
ORCID:- 0000-0002-0685-0650

Abstract

As part of a digital forensic examination, a practitioner may identify data traces that they believe to be relevant to their inquiry and seek to interpret their meaning, forming a primary investigative hypothesis. In addition, practitioners should also consider whether any traces could mean something else. This work discusses the need for practitioners to consider 'technically possible' alternative meanings (TPAMs) as a standard component of their interpretive process. It is proposed that, when considering whether any TPAMs exist in addition to the practitioner's primary investigative hypothesis regarding a data trace, the practitioner's position may be expressed in one of six ways - '*the six categories of TPAM*', based upon the available objective support related to or present within their case, from which the TPAM is derived. These six categories are proposed in order to help a practitioner effectively communicate their reasoning for offering a TPAM in regards to any data trace found during an investigation and are defined and discussed.

Keywords: Digital Forensics; Hypothesis; Interpretation; Alternative Possible Meaning; Investigation

Introduction

The reliable interpretation and communication of investigative findings is paramount in all of forensic science, yet whilst more traditional branches of this domain have established somewhat rigorous best practices for achieving this [1 2], the field of digital forensics (DF) arguably remains underdeveloped in this area [3 4 5 6]. Conducting a forensic examination of digital data is far from a simple task given the complexity of modern systems and the large qualities of diverse data that a practitioner can be faced

Published by Taylor and Francis. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial License (CC:BY:NC 4.0).
The final published version (version of record) is available online at DOI:10.1080/00450618.2022.2071988. Please refer to any applicable publisher terms of use.

with [7] [8]. Following the searching and sifting of digital content, a practitioner may identify specific data traces which they believe to be of value to the inquiry being conducted by their wider investigatory team. This 'value' is ascertained through an interpretive evaluation of these data traces in order to establish their 'meaning'; when distilled - *what are they, why are they present* and *what do they mean?* Once addressed, the answers to these questions may form the practitioner's primary investigative hypothesis regarding any data trace. Whilst a practitioner may be tempted at this point to convey this to their client, their interpretive duties do not end here. In addition to their primary hypothesis, a practitioner must consider whether any 'technically possible' alternative meanings (TPAMs) for any given potentially relevant data trace could exist, and communicate these appropriately [14].

This work discusses what a TPAM is and why consideration of them is required as part of any DF investigation in Section 2. It is proposed that, when considering whether any TPAMs exist in addition to the practitioner's primary investigative hypothesis regarding a data trace, the practitioner's position may be expressed in one of six ways - '*the six categories of TPAM*', based upon any available objective support present within their case from which any TPAM is derived. These six categories are defined in Section 3 in order to help the practitioner to effectively communicate their reasoning for proposing a TPAM in regards to any data trace found during an investigation. Finally, the justification for identifying six categroies of TPAM is outlined and conclusions are drawn.

2 Establishing meaning

As part of a digital investigation, practitioners may identify a number of digital traces that they believe to be of relevance to the circumstances of their investigative team's inquiry. When determining a data trace's relevance, the practitioner will need to assess what they think the data trace is or could be, how they think the data trace came to be present on the system and what they think the data trace could mean. Through the use of validated tools, testing, knowledge and experience a practitioner may establish answers to all of the aforementioned questions, forming their primary investigative hypothesis of any data trace(s). Whilst this is a fundamental and challenging task, it is not the end of any practitioner's

obligations in regards to establishing the meaning of any data traces on a device, where consideration of any TPAMs is required (see Figure 1).

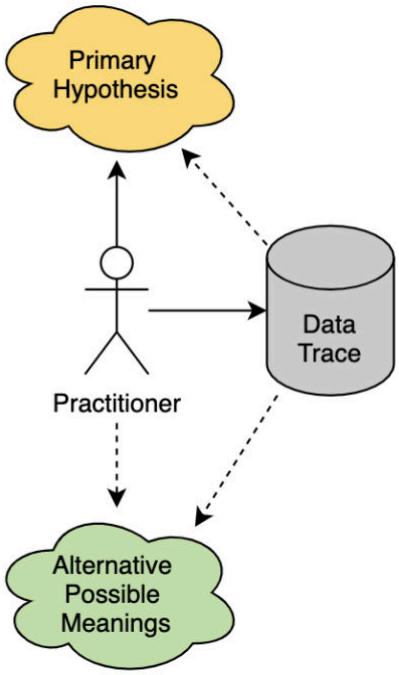

Figure 1: Forming meaning.

The consideration of TPAMs is fundamental to ensuring that the practitioner's interpretative process is conducted thoroughly [9]. Sunde and Dror [10] note that the assessment of TPAMs helps to ascertain both the strengths and weaknesses of the primary hypothesis and to reduce the risks posed by confirmation bias. Failure to consider TPAMs in relation to a data trace may result in the DF practitioner presenting only their primary interpretive hypothesis in regards to it, which may mislead or misdirect future investigative inquiries and result in misinterpretation or unsafe conclusions being reached. Fundamentally, TPAMs must be contemplated and evaluated for an examination and interpretation process to be considered comprehensive and reliable.

2.1 Forming a TPAM

In regards to any relevant data trace on a system and a practitioner's primary investigative hypothesis of it, multiple TPAMs may exist and these may not be easy to distinguish and define [11]. Take for

example a visit to a known-bad website discovered in the device's web browser history logs. A practitioner may proceed to hypothesise that the Internet history log is present due to the device's user having typed in the web address of the site and requested to be served its content. Whilst the data situated in the web browser logs may be consistent with this set of events, a practitioner should consider whether there are other possible reasons why that specific Internet history record is present, for example, could the record be the result of malicious software being active upon the device.

When considering whether TPAMs may exist for any given data trace and their primary investigative hypothesis of it, it may help the practitioner to consider two investigative concepts; '*where could the data trace have come from?*' and '*what actions could have caused the data trace?*' - source and activity level behaviours (see Cook et al's., [12] hierarchy of propositions). Where a digital trace came from aims to determine the source of the trace, where consideration of all of the places that a digital trace could have originated from must be given. In addition, an evaluation of those actions that could have caused the digital trace's presence/creation must be carried out. At which point, a practitioner may be able to define one or more viable TPAMs which must be explored through an evaluation of any and all existing additional relevant digital traces present within their examination scenario which could support any proposed TPAM. In addition, a practitioner should determine the prospect of any proposed TPAM having taken place, but having not left trace evidence behind on a device.

Once a practitioner has formed a TPAM, they may be required to quantify and express the likelihood of its occurrence and describe the methods for ascertaining this conclusion [1 2], or simply communicate whether it is viable given the current technical circumstances of the case and any surrounding available information, and why this is the case (see Figure 2).

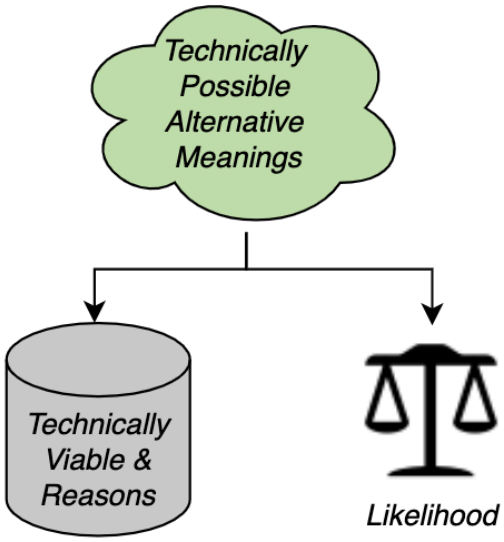

Figure 2: Two ways to describe a TPAM.

This work will not stray into the remit of quantifying the likelihood of a TPAM (see a recent discussion of the challenges and advancements in this area offered by Overill and Collie (2021)), but instead concentrates on practitioners who seek to express a TPAM as a technical possibility. In that sense, in acknowledgement of Jackson et al., [13] p.37) a TPAM may be considered 'an explanation' - its chance of occurrence is not quantified and converged. When distilled, it may seem that a question of 'technical possibility' is one that is binary; something *is* either possible, or *it is not*. Whilst this stance may not be incorrect, it does limit the ability of those seeking to rely upon the TPAM being expressed to understand why the practitioner has derived it.

It is proposed that, when considering whether any TPAMs exist in addition to the practitioner's primary investigative hypothesis regarding a data trace, the practitioner's position may be expressed in one of six ways based upon any available objective support present within their case from which the TPAM is derived. These six categories of TPAM are defined and explored in Section 3.

3 The six categories of TPAM

When considering the proposal of a TPAM in relation to their primary investigative hypothesis regarding a data trace, it is suggested that the practitioner should consider conveying their position in

one of six ways; *No other possible explanation*, a *full trace possibility*, a *partial trace possibility*, a *documented technical possibility unsupported by available traces*, a *theoretical technical possibility*, and finally, *an unknown technical possibility* (see Figure 3). These six categories are proposed both in order to help a practitioner to clarify and express the objective evidence-base from which they have derived and offered a TPAM, whilst also helping those who are seeking to rely upon the practitioner's results to understand why a TPAM has been suggested. All six categories are discussed below.

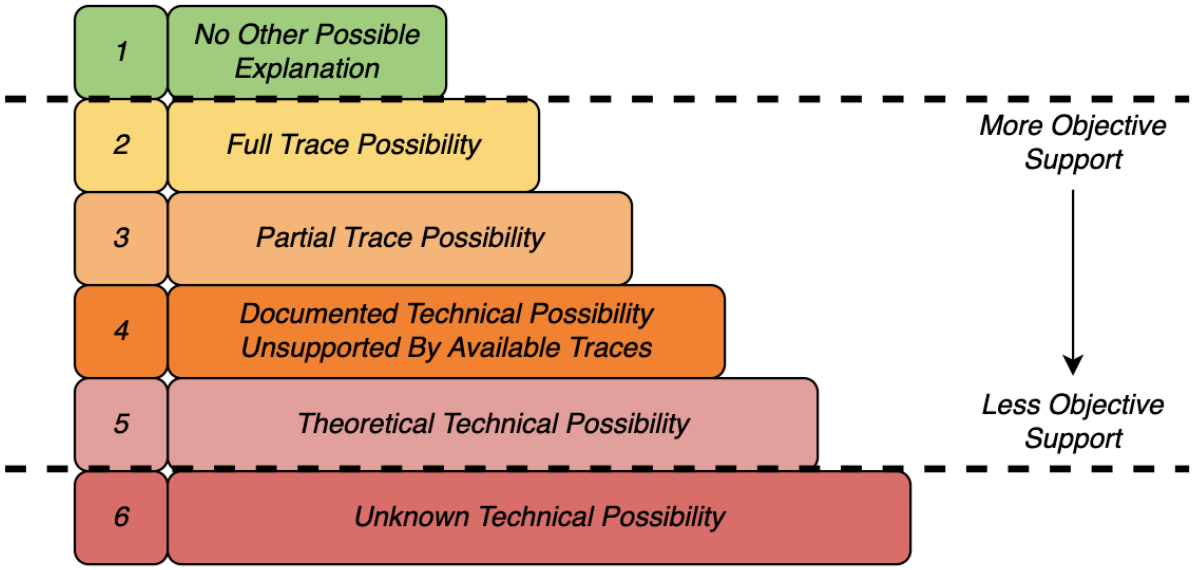

Figure 3: The six categories of TPAM

1 - No other possible explanation: In category 1, a practitioner may consider that there are no other possible explanations other than their primary hypothesis in regards to a specific relevant data trace. This category must be considered one where an absolute avoidance of doubt must exist in relation to the potential for a TPAM, where a rigorous evaluation of the case circumstances is required in order for the practitioner to arrive at this category. It is argued that it will be rare for practitioners to reach such a conclusion given the large number of potential scenarios which can often result in the creation of digital traces on a device, meaning that in many cases, multiple potential reasons will exist for a data trace's presence on a system or its meaning. Therefore it is suggested that caution be taken when seeking

to express such a category as in almost all cases it is arguably possible to derive an alternative technical explanation, regardless of how potentially unlikely it may seem.

2 - Full trace possibility: In some instances, a device/dataset may contain all of the digital traces required to validate both the practitioner's primary hypothesis in regards to a data trace, and also their TPAM. These scenarios are considered *full trace possibilities* as all of the digital traces that would be expected to be present if the TPAM had occurred, are present on the system. This means that there is all of the required technical information present on the device/dataset being examined to support both the primary hypothesis and TPAM in relation to a data trace. An example may include a scenario where an individual has claimed that malicious software has downloaded illegal material from the Internet onto their device. An investigation of the device may show the presence of malicious code which is considered capable of doing these actions and there may also be evidence of this being carried out, forming the basis of a practitioner's primary hypothesis. However, in addition, the examination may also reveal data traces in a web browser's logs showing a visit to a relevant website followed by logged download transactions and relevant illegal files residing in the downloads folder that the browser is configured to store its downloaded files. In such a case, all of the traces required to technically support both hypotheses could be present if all relevant traces for an act of a user-initiated download are also present, and the two acts are digitally indistinguishable. In essence, there are no data traces available that would support the primary hypothesis over and above the TPAM.

3- Partial trace possibility: In contrast to category 2 above, a partial trace possibility arises where a device/dataset contains some of the data traces for validating a TPAM, but not all. In that sense, a TPAM is only partially supported by the data traces identified during an examination, meaning that whilst its occurrence may still be possible, the practitioner would have expected the presence of additional data traces on the system should it have occurred. In these circumstances, the practitioner should determine why these additional data traces may not be present on a device before expressing their category of TPAM. To be clear, TPAMs categorised as partial trace possibilities still cannot be discounted as viable, yet there is less objective support for their occurrence present on the device.

An example may include the presence of a relevant picture on a mobile device, where a practitioner may generate a primary hypothesis that suggests that the picture had been acquired following access to a cloud storage account. A category 2 TPAM may arise if the suspect claims that they were sent the picture without requesting it from a contact they met online via an instant messaging client (IMC). Recovered chat logs may indicate that a picture with the same file name was sent, but it may no longer be present in the default location where the IMC stored received pictures, nor is there any definitive information available to suggest the picture in question was in fact the image sent during the conversation (for example, a hash value). The fact that the picture may have the same file name as that noted in the conversation may support a partial trace possibility, meaning that the TPAM could be viable but there are only partial system traces to support it.

4 - Documented technical possibility unsupported by available traces: In category 4, a practitioner may propose a TPAM in relation to a data trace that is a documented technical possibility, but one that is unsupported by any available data traces i.e. there is no objective evidence of the TPAM having occurred on the system. To note, this category does not preclude the TPAM from occurring, only that its occurrence may not be evidencable within the confines of the current examination scenario. In category 4, any proposed TPAM must be derived from a known, documented source that demonstrates the technical possibility for the TPAM within the given circumstances of the examination scenario. However, whilst the TPAM may be a technical possibility, it is one where there are no data traces present within the examined device/dataset that would support the occurrence of it. These circumstances may arise where there is a known and documented instance of a specific event that could occur within the current technical setup of the case being examined, and be a viable TPAM for a specific data trace, but the documented event is found to leave no digital traces on a system denoting its occurrence. Therefore it may neither be possible to 'prove' the occurrence of it, nor discount it.

An example of this category of TPAM may include the cases where the practitioner may seek to infer meaning and reliance upon specific file metadata (for example, EXIF data, timestamp information).

Such metadata may support a primary hypothesis regarding specific interactions upon a system, however, it may also be considered possible for the metadata to be manipulated via an anti-forensic third party process. Whilst the manipulation may be considered technically possible, to validate it, the practitioner may have expected to see the data traces showing relevant file/system interactions for the process responsible for it - however, these may not be present following an examination. Therefore the TPAM could technically occur, but previous occurrences of it suggest certain data traces would have been present on the suspect's device.

5 - Theoretical Technical Possibility: In category 5, a theoretical technical possibility is a descriptor for those TPAMs that a practitioner considers viable at a technical level, but there are no traces present on a system to suggest its occurrence, nor are there any documented occurrences of the TPAM known to them (for example, literature or past experiences that they can draw upon). A practitioner may derive theoretical technical TPAMs through an evaluation of the system/device under investigation and their knowledge of appropriate technology/services that are capable of being responsible for the primary data trace being investigated. Where at category 4 a TPAMs occurrence has been documented, here there are no documented occurrences available to validate it. In this category, practitioners should be mindful of developing theoretical TPAMs which may be considered wildly unrealistic and may only seek to complicate the interpretive process and mislead those seeking to engage with it and rely on any results.

An example may include a TPAM regarding communication that could have been 'spoofed'. The act of spoofing may be known to be theoretically technically possible, but no examples are known in the context of the current investigative scenario, nor are there any data traces that might be expected to be present should spoofing have occurred. However, it cannot be absolutely discounted from causing any relevant data traces on a suspect device.

6 - Unknown Technical Possibility: Practitioners may also acknowledge their position as being one where they are unable to determine whether there are any TPAMs for a specific data trace other than their primary hypothesis. This is not an acknowledgement that no TPAMs exist, only that one cannot

be identified given the current case circumstances and knowledge of the practitioner. This could be considered to be a stance adopted that is somewhat opposite of that shown in category 1 TPAMs.

3.1 Interpreting the category

It is argued that the six categories of TPAM provide support for expressing the basis for a practitioner's proposal of any TPAMs and help members of the wider investigatory team to understand the underpinning information used to form it. When interpreting the six categories, whilst 1 conveys absolute certainty and 6 acknowledges a state of 'not knowing', categories 2-5 require further discussion. TPAM categories 2-5 can be interpreted in two ways, those TPAMs that have '*in-case support*' (2 & 3) and those which have '*external support*' (4 & 5). TPAMs proposed at categories 2 and 3 are offered due to the presence of relevant data traces found during an investigation, therefore these are TPAMs considered to have *in-case support*. In contrast, at categories 4 and 5, practitioners propose TPAMs which are derived from external information, and whilst there must be the technical ability for the TPAM to occur within the examined device, there are no digital traces present that support its occurrence. The distinction between levels of support may suggest that some proposed TPAMs have a stronger likelihood than others - a position that must be approached with caution.

3.2 Strength of TPAM

For the purposes of simplicity, if we consider the DF hypothesis formation process linear (Figure 4), then the first task for the practitioner is to form their primary investigative hypothesis of any data traces. This is a process which occurs in all DF investigations as the production of investigative results forms a core component. The first challenge to address in the interpretative process here is the consideration and formation of any TPAM for any data traces. Only once a TPAM has been established/determined can the practitioner consider the possibility of quantifying the likelihood of its occurrence - *challenge two*. To be clear, the six categories of TPAM are not to be interpreted as a measure of likelihood-quantification, but they could play a supporting role in this process should the practitioner be subsequently required to carry it out.

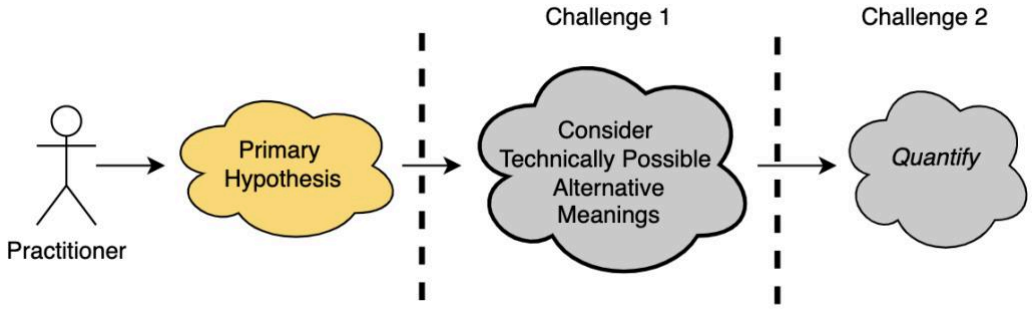

Figure 4: A simplification of the DF hypothesis formation process.

Whilst we may consider that the greater the amount of evidence available to support a particular hypothesis, the greater the chance of it having occurred, this is in fact a dangerous oversimplification in relation to digital evidence, where clarification is required. Digital systems are complex and multifaceted, offering their users an almost infinite set of functionality. In doing so, it must be appreciated that for any given data trace on a system, there may be more than one possible reason for its presence and in turn, its presence could be interpreted in multiple ways. Whilst the goal of a DF investigation is to ascertain meaning attributable to any actions deemed relevant, in a reliable way, it may not always be possible to do this as some actions do not leave a full and conclusive set of digital traces behind. Therefore it may not be possible to determine whether some digital actions have been carried out on a digital device as part of a post mortem DF examination of it. For this reason, it is difficult to base the strength of a digital hypothesis solely on the presence of available data traces that support it. Whilst the availability of relevant data traces should be considered an important part of evidencing a hypothesis, wider interpretive considerations are needed. For that reason, the proposed categories of TPAM albeit ordered, are not ordered directly by the strength of the TPAM and its likelihood of occurrence.

Instead, the ordering of the states of TPAMs should be interpreted with reference to the amount of what we consider here as '*objective support*' attributable to them. Objective support is deemed to be the amount of information present either in the original digital scenario which could be attributed to the actions relating to a TPAM should it have occurred, or information applicable to a TPAM and the

circumstances of the case. For that reason, TPAM states 2-5 are arranged by the amount of objective support they are deemed to have, and bookended by scenarios where no TPAM is believed to exist (TPAM category 1) and category 6, where it is unknown as to whether a TPAM exists.

This work, nor the six states of TPAM defined above set out how a practitioner should interpret a digital trace or establish a TPAM. The diversity of both trace-types and scenario-circumstances mean that codifying this process is a task beyond the intended remit of this work and one which will often be individual to each case scenario.

4 Final thoughts, justifications and conclusions

First, it is necessary to acknowledge the tightly defined remit for using the six categories of TPAM. These categories can only be used to supplement the communication of a TPAM for a data trace that is technically possible through categorization based upon what is considered here as *'the amount of objective information that is available that supports it*'. The six states may be used to supplement any quantification of likelihood in relation to a TPAM, but when considered in isolation, they should not be considered as a measure of the hierarchical likelihood of the TPAM. Therefore the categories should be considered as descriptors that are designed to assist in showing those seeking to rely upon the findings presented by a practitioner what underpins their practitioner's proposed TPAM. In addition, the six categories may serve as a mechanism for encouraging the practitioner to scrutinise their own underlying belief in a TPAM and what such beliefs are based upon.

Offering TPAMs quantitatively (an evaluative approach - see Champod et al., [1] for evaluative reporting guidance) and as simple technical explanations (an investigative approach - see Champod et al., [1] for investigative reporting guidance) are two separate challenges where arguably the former must build upon the latter. To that extent, before the likelihood of a TPAM can be quantified, the possibility of it occurring at a technical level must first be established, shown in Figure 4. For that reason, it is argued that support for clearly categorising and communicating TPAMs at a technical level is first required. It is hoped that by encouraging practitioners to define their position in relation to a TPAM, those seeking

to interpret this as a part of future investigative decision making ('the hypothetical client') have a clearer insight into the practitioner basis for a TPAM. In turn, the structured approach to catagorising a TPAM may encourage practitioners to increase the veracity of their evaluation of any TPAM. Finally, as a minimum, it is suggested that by formalising the types of TPAM that can exist, this may prevent practitioners from omitting to consider TPAMs in relation to any relevant data traces found during an investigation (either deliberately or accidentally through an oversight), and proceeding to present their primary hypothesis absent of any further considerations. By raising awareness of the need and types of potential TPAMs, clients of any DF investigation may be more inclined to raise such questions during any case conference held with the practitioner. Such steps are proposed in an effort to increase both the transparency of the DF interpretative process and as a measure to support quality assurance with regard to DF examination and interpretive practices.

References

1. Champod, C., Biedermann, A., Vuille, J., Willis, S. and De Kinder, J., 2016. ENFSI Guideline for evaluative reporting in forensic science: A primer for legal practitioners. Criminal Law and Justice Weekly, 180(10), pp.189-193.
2. National Institute of Forensic Science Australia New Zealand, 2017 'An introductory guide to Evaluative Reporting' Available at: https://www.anzpaa.org.au/forensic-science/our-work/projects/evaluative-reporting-project/evaluative-reporting (Accessed: March 8, 2021).
3. Overill, R.E. and Silomon, J.A., 2010, September. Digital meta-forensics: Quantifying the investigation. In Proc. 4th International Conference on Cybercrime Forensics Education & Training (CFET 2010), Canterbury, UK (September 2010).
4. Overill, R.E., 2014. Quantifying likelihood in digital forensic investigations. Journal of Harbin Institute of Technology (New Series), 21(6), pp.1-4.
5. Horsman, G., 2020. Digital evidence certainty descriptors (DECDs). Forensic Science International: Digital Investigation, 32, p.200896.
6. Overill, R.E. and Collie, J., 2021. Quantitative evaluation of the results of digital forensic investigations: a review of progress. Forensic Sciences Research, 6(1), pp.13-18.

7. Garfinkel, S.L., 2010. Digital forensics research: The next 10 years. digital investigation, 7, pp.S64-S73.
8. Caviglione, L., Wendzel, S. and Mazurczyk, W., 2017. The future of digital forensics: Challenges and the road ahead. IEEE Security & Privacy, 15(6), pp.12-17.
9. Sommer, P., 2018. Accrediting digital forensics: what are the choices?. 25 pp.116-120.
10. Sunde, N. and Dror, I.E., 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. Digital investigation, 29, pp.101-108.
11. Spafford, E., 2006, January. Some Challenges in Digital Forensics. In IFIP International Conference on Digital Forensics (pp. 3-9). Springer, Boston, MA.
12. Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A., 1998. A hierarchy of propositions: deciding which level to address in casework. Science & Justice, 38(4), pp.231-239.
13. Jackson, G., Jones, S., Booth, G., Champod, C. and Evett, I.W., 2006. The nature of forensic science opinion--a possible framework to guide thinking and practice in investigations and in court proceedings. Science & justice: journal of the Forensic Science Society, 46(1), pp.33-44.
14. Casey, E., 2020. Standardization of forming and expressing preliminary evaluative opinions on digital evidence. Forensic Science International: Digital Investigation, 32, p.200888.

https://dspace.org/

Considering ‘technically possible’ alternative meanings for data traces found during a digital forensic examination

Horsman, Graeme

2022-09
Taylor & Francis
https://doi.org/10.1080/00450618.2022.2071988
Horsman G. (2023) Considering ‘technically possible’ alternative meanings for data traces found during a digital forensic examination. Australian Journal of Forensic Sciences, Volume 55, Issue 5, September 2023, pp. 689-698

Downloaded from a DSpace repository. https://dspace.org/