M. S. Olivier, "On complex crimes and digital forensics,"
Technishe Berichte 63, Hasso-Plattner-Instituts für Softwaresystemtechnik an der Universität Potsdam, 2013

ON COMPLEX CRIMES AND DIGITAL FORENSICS

MARTIN S OLIVIER

ABSTRACT. Science provides the basis for truth claims in forensics. Very little research has been done to explore the scientific basis of digital forensics. The work that has been done vary widely in what they propose; in most cases it is unclear how the philosophical remarks about such forensic science apply to digital forensics practice, or that the practical suggestions are a sufficient basis to claim that practice based on them is scientific.

This paper provides an initial exploration of the potential of decision problems from the field of algorithmics to form this scientific basis. There is no doubt that decision problems operate in the scientific domain and decision problems look similar to hypotheses to be of immediate practical use.

The paper suggests that, if decision problems are used in this manner, it is clear that current digital forensics have only scratched the surface of what is possible. Probabilistic complexity classes, for example, offer interesting possibilities for performing complex tests in relatively short times, with known error rates.

Using decision problems as a demarcation criterion makes it possible to distinguish between digital forensic science (or simply *digital forensics*) and digital forensic craft, that should rather be called *digital investigative technique* or some other suitable term that does not imply that its use leads to scientific truths.

1. INTRODUCTION

Forensics entails the use of science to determine matters of fact where such facts are required to settle disputes (for example, in courts of law) or to determine the root cause of an event of interest. Forensics employs the notion that scientific knowledge is *true* and hence a good basis to settle such disputes and/or determine causes. Digital forensics is that branch of forensics that studies evidence that exists is digital form.

In order to make such truth claims forensics has to be 'scientific'. In some cases this is emphasised by using the term *forensic science*, which in this paper will be deemed to be synonymous with the term *forensics*. The notion of *science* (as well as the notion of *truth*) has been the subject of deep philosophical reflection over centuries; so much has been said that a paper that ultimately intends to deal with a small fraction of forensic science cannot hope to do justice to.

The obvious question then is what is the nature of *digital forensic science* or, with the same meaning, the science that underlies digital forensics? Cohen [5] is the only author who has provided a coherent answer to this question by describing an *information physics* — 'natural laws' that apply to information and can be used as the basis for more complex truth claims. However, it is not yet clear that it is possible to always relate the behaviour of a complex system to truths about bits and related matters — see, for example, Hofstadter's argument [13] that a complex system may be more than the mere sum of its parts and may exhibit characteristics that are not present in the parts.

Key words and phrases. Digital forensics, forensic science, decision problems.

The author would like to thank Stefan Gruner for providing some valuable insights after reading an earlier draft of this paper and Candice le Sueur for proofreading an almost final version of the paper. The author obviously remains responsible for any errors or inaccuracies that remain.

A recent newspaper story [15] provides some insight on what may go wrong if we rely on digital forensics that cannot be trusted — it may negatively affect innocent people. However, simply discarding digital forensics because of a lack of trust turns the cyberworld into a safe haven for criminals who can exploit others without fear of being caught. Clearly a digital forensics is required that maximises the chances that the guilt of the guilty can be proven, and that will ideally never implicate an innocent party. If these requirements are met the inhabitants of cyberspace can proceed with trust even in those cases where the proactive security mechanisms fail. Note that this problem is not only present in digital forensics; other branches of forensics have also failed because they used junk science or pseudoscience [9, 18]. Regarding digital forensics, Caloyannides [3] boldly declares that "It is important for judges and juries to be highly sceptical of any claims by prosecution that digital 'evidence' proves anything at all."

This paper will examine the suitability of *algorithmics* or *algorithmic complexity theory* to form the basis of digital forensics. The justification of positing algorithmics as this basis is deferred to later in the paper when required underlying issues have been discussed.

From the outset it is important to note that the paper distinguishes between expert testimony and forensics. In many jurisdictions forensic evidence can only be introduced in a court case by means of expert testimony. However, not all expert testimony is based on forensics. Consider, for example, the medical doctor who testifies as an expert about the current standard of care for some ailment. This testimony will be partly based on medical training (including continuing education), partly on professional observation of what colleagues do, partly by standards that may have been published by national and international bodies and partly by local conditions (such as affordability of various treatment options). Clearly such testimony from an expert may be invaluable in a case where it is required. However, such evidence will not be classified as scientific evidence. In particular is this witness not basing evidence on forensic science.

The remainder of the paper is structured as follows. The next section reviews some chacateristics of science, forensic science and expert testimony to provide context for the exploration of digital forensic science that follows. Section 3 inititates this exploration by discussing two simple (and common) scenarios at length. Section 4 uses these scenarios, the notion of decision problems and expectations about digital forensic science from the literature to begin to develop a theory of digital forensics that can claim to be scientific. Section 5 briefly mentions some competing theories. Section 6 concludes the paper.

2. On science, forensic science and expert testimony

As noted earlier the intention of the current paper is not to explore the notion of science in depth. In the philosophy of science the following three landmarks are most important for the purposes of this paper. Firstly, in the period before the Second World War a group known as the Vienna Circle developed the notion of logical positivism. According to them the only meaningful judgements were the tautologies from mathematics and logic, and *verifiable* empirical claims from science. Everything else was nonsense. The second landmark is Popper's demarcation criterion for science: *falsifiability*. Only theories that can be falsified should be regarded as science. To be more specific, Popper foresees series of theories, where, when one theory is falsified, it is replaced by another theory that has greater explanatory power. Finally, Kuhn [16] describes (rather than defines) science as an endeavour where during periods of *normal* science, scientists solve puzzles using the paradigm then prevalent. Once an existing theory becomes unsustainable, it is replaced by a new theory (again with greater explanatory power) during what he calls a *scientific revolution*.

Clearly much has to be added to this admittedly superficial descriptions of science to make them useful for forensic purposes. A theory that can be falsified but has not been tested at all may qualify as science, but not as grounds for the conviction of an alleged criminal. Similarly, the mere fact that a scientist has followed the appropriate paradigm may not ensure the reliability of the results. Rather than looking at the philosophy of science for deeper understanding, we turn our attention to the law.

Expert testimony in courts have a long history. In 1782 a civil engineer and scientist testified in the Wells Harbour case in the UK. Rather than just surmising from current observations what caused the silting up of the harbour he claimed that it was "necessary to shew the natural causes by which the port of Wells has been formed" [19, p.150]. This was extraordinary since Mr Smeaton was testifying about something he did not observe, but derived from laws of nature [10]. He also did not derive those laws or even tested them. Normally such testimony would have been classified as hearsay, or even irrelevant to the specific case being heard. The opposing side did indeed attempt to get his evidence excluded. However, Lord Mansfield who was presiding over the trial wrote "I cannot believe that when the question is, whether a defect arises from a natural or an artificial cause, the opinions of men of science are not to be received [. . .] The cause of the decay of the harbour is also a matter of science, and still more so whether the removal of the bank can be beneficial. On this such men as Mr. Smeaton alone can judge. Therefore we are of the opinion that his judgement, formed on facts, was very proper evidence." This is often cited as the first use of *science* (or forensic science) in a court of law.

Of course the use of science enabled more informed judgments to be made, but over time much pseudoscience developed where claims were made based on some set of theories that was not scientific at all. A relatively modern example of a challenge that faces courts is the use of a polygraph to obtain evidence. The validity of such evidence is the topic of much debate; many reject polygraphy as pseudoscience, while others consider it to be very reliable. Many government agencies, for example, consider polygraphy as useful [6]. Even where such evidence is not accepted, a suspect who volunteers for such a test scores some credibility points.

It is therefore important that the court acts as a gatekeeper to only allow 'valid' or 'true' science to be accepted as scientific evidence. Of course it should not be necessary to qualify science using words such as *valid* or *true* because science itself implies those characteristics.

The best known 'modern' test for admissibility of expert testimony is the Daubert standard used in the USA. In this case the court decided, amongst others, that [20]

> *Faced with a proffer of expert scientific testimony under Rule 702, the trial judge, pursuant to Rule 104(a), must make a preliminary assessment of whether the testimony's underlying reasoning or methodology is scientifically valid and properly can be applied to the facts at issue. Many considerations will bear on the inquiry, including whether the theory or technique in question can be (and has been) tested, whether it has been subjected to peer review and publication, its known or potential error rate and the existence and maintenance of standards controlling its operation, and whether it has attracted widespread acceptance within a relevant scientific community. The inquiry is a flexible one, and its focus must be solely on principles and methodology, not on the conclusions that they generate.*

While this standard has been slightly revised by other courts many of the phrases of this judgment reverberates in the minds of those who are trying to establish a scientific foundation for some forensic discipline. The key phrases include *tested*, *peer review and publication*, *widespread acceptance* and, perhaps the most challenging of all (and not present in some later formulations of the standard, but still frequently highlighted) the *known or potential error rate* of the theory. Note that it is easy to critique Daubert — the minority judgment that forms part of the decision cited above [20] is a good starting point for such critique. However, some mechanism is required to keep pseudoscience out of courts, and Daubert is arguably the most prominent current standard used for this purpose.

Clearly then, for digital forensics to become (or remain) trusted, self reflection is necessary in the light of standards such as Daubert.

3. Forensic craft and science

Let us consider just two typical scenarios encountered in 'digital cases' and distinguish between the craft and science involved.[1]

The first — and apparently most prevalent — example is one where it is necessary to show that some data is present on (or absent from) some medium. The nature of the content to find may vary. In the simplest case it may be some byte sequence, such as some credit card number (say 1234-5678-9012-3456) or a specific MP3 file. In a more complex case it may be an (any) email sent between two specific parties or a (any) JPEG image depicting certain content. For ease of reference we will refer to the criteria used for searching as the *search pattern* even though the criteria may not be a typical *pattern* — such as when the criteria specify a certain file type. To find the search pattern a number of subtasks need to be completed. Firstly the disc (or other media) content needs to be acquired in a forensically sound manner. Secondly, the search pattern needs to be located. Thirdly, it has to be demonstrated that the search pattern does occur on the media (and it may be necessary to indicate what the full details are, for example the content of an email that has been found based on sender and recipient). Or it may be useful to indicate that the search pattern does not exist on the media (or, a weaker claim, indicate that the content was not found on the media).

The first of these three steps is typically not a scientific activity. On physical scenes crime scene investigators (CSIs) or first responders or some other group — rather than forensics scientists — collect (or bag and tag) the evidence. Contamination of evidence is one of the main concerns and therefore the collectors use specific collection (or acquisition) protocols to the letter. Legal issues (such as authority to collect evidence and questions about jurisdiction) may also play a role, but once again set protocols is followed or a legal expert (rather than a forensic scientist) is consulted.

Some scientific questions may arise. If, for example, it is (or becomes) clear that the container used to collect, say, some chemical or biological material reacts to its contents, it implies that such a container may contaminate such evidence. It then becomes a question of science to find (or develop) a container for which it can be scientifically shown that it will not contaminate the evidence. Similarly, if physical evidence may degrade over time it may be necessary to develop a container that restricts such degradation by, for example, maintaining the proper temperature or by preserving the evidence in some appropriate preservative compound. These are clearly questions for science. However, in some cases

[1] Note that many branches of computing combine craft and science and that a need to distinguish between craft and science — or even between art, craft, technics, engineering and science — becomes necessary. See the paper by Gruner [11] about the nature of this discourse in the software engineering discipline as an example.

the scientific knowledge, such as the temperature at which evidence will degrade, is already known and it becomes purely a question of engineering to construct a container that (for our example) maintains the appropriate temperature. There may be a question about the type of science involved in this step. The fact that some biological material does not degrade below a certain temperature may be a question of 'pure' biological science, rather than forensic science. However, we do not explore this possible distinction between 'pure' and forensic science further in the current paper.

We claim that, for the current scenario where some search pattern is to be located, acquisition is in principle very similar to physical acquisition. We have known for many years how to image disc drives. We know that we ought to use write blockers to prevent contamination or, if write blockers are not available, to use an operating system that allows the disc to be mounted as a read-only device. In the latter case we may know that it is best to boot that operating system from a read-only medium and to then bag and tag the medium as evidence in case any questions are later raised about the reliability of the operating system regarding not writing to media that are mounted in a read-only mode. We know that we have to calculate some hash (such as MD5 or SHA1) for the content of the original media as well as our evidentiary copies to demonstrate the integrity of our copy. Note that very little of this process is based on science; most of it is a matter of common sense. Where science does play a role (for example in the integrity claims supported by message digests), that science is also widely used outside the realm of forensic science. Note again that this corresponds with physical forensics. The CSI who collects DNA from a suspect by brushing a swab inside the suspect's mouth to collect some saliva is typically not a scientist with a university degree in science.[2] Similarly, the officers who collects fingerprints from a crime scene or even the officers who spray Luminol to detect spilled blood are not usually scientists. Note that this does not mean that they may be unqualified or inexperienced — their requirements and experience are just not as scientists and they are not expected to derive scientific truths.

A couple of remarks are in order about the digital forensic acquisition process described in the previous paragraph. Rather than imaging the device the CSI may simply seize the media and send it to the forensic laboratory to be imaged (and then analysed). However, the fact that imaging may occur in the laboratory does not make it a scientific process *per se*. The process of imaging described above has become known as "dead analysis", with many known shortcomings (such as its impact on business continuity for the entity being investigated). An alternative is so-called "live analysis" [1], which will not be explored further in the current paper. However, live analysis (also) desperately needs answers for the questions raised in this paper.

This concludes our discussion of the first step in the scenario described above. In summary, collection or acquisition in this scenario is primarily a technical activity (or craft), rather than a scientific activity.

Although we have distinguished between steps 2 (searching) and 3 (demonstrating presence or absence) above, the distinction does become blurred in many cases. We will, however, for the time being, use this distinction for the sake of exposition. What we do know

[2] As an example, to be formally recognised as a professional, candidate or certified natural scientist in South Africa a person has to meet the requirements specified by Act 27 of 2003 (Natural Scientific Professions Act, 2003); this act establishes the South African Council for Natural Scientific Professions which is responsible for registration of scientists who meet the prescribed requirements. In general a four year degree followed by three years of professional experience — or a higher degree followed by a shorter period of professional experience — is required to register as a professional natural scientist.

at this point is that if digital forensics is a science, the science has to be part of step 2 or 3, since it was not present in step 1.

As noted, the second step of the given scenario entails searching for and finding (or not finding) the search pattern. Again we may use physical forensics as a point of departure. Consider an apparent murder case where the (possible) murder weapon needs to be found. It is possible that a knife is still stuck in the victim's body, in which case finding it is trivial and the process of finding it will not be considered forensic science. In a somewhat harder case the investigating officers may find a knife that they think may be the murder weapon in the suspect's home. Much now depends on the characteristics of the knife: Is it bloody? Is it similar to a set of knives from the victim's home and one such knife is missing from the victim's set? Does it have fragments of cloth stuck to it that correspond to the clothes the victim was wearing? Remember that we are assuming that there is some reason why the investigators think that this may be the weapon. If the knife is bloody, matches the set in the victim's house and contains 'obvious' fragments of cloth, the search may be over before the forensics have begun. Forensics will only come into play at step 3, where it needs to be proven that the found knife matches the victim's wound (and/or whatever other matches that may add scientific weight to the claim that the knife was indeed the murder weapon). However, if there are not such a multitude of indicators that the identified knife is the correct one forensics may begin to play a role much earlier in the search. For example, it may be determined from the wound that a serrated or smooth knife was used; it may be possible to determine the length (and perhaps other measurements of the blade); paint or other traces from the knife may determine its colour, and so on. The investigators can now proceed with a (non-forensic) search based on what they have learnt from the forensic scientists about the weapon they are looking for. Finally, if the murder weapon was some poison, searching for (traces of) it may be a pure forensic exercise.

Our digital forensic scenario requires us to find some specified data on the disc image. Let us now make a sacrilegious claim: In general *any* tool may be used to search for the data. Of course the tool needs to be suitable for performing the search: if we are looking for a type of file (rather than exact text) we need a tool that is able to search for such files. To illustrate, suppose the investigator copies the image to a hard disc of a computer and then boots the computer from this disc. Suppose the investigator opens the email application and uses its search fields to find the email messages between the two parties that are of interest. Or suppose we are indeed looking for some pattern; suppose the investigator uses `grep` or some other pattern matching program to find the required files. And, in any of these cases, suppose the investigators find what they were looking for. Is there any reason to object because 'non-forensic' or 'untrusted' tools have been used? I claim that there ought to be no objection. The claim is based on the assumption that we will during step 3 prove that the search pattern does indeed exist on the medium. Whatever methods we used to locate it are irrelevant.

Objections to these claims may come from multiple sources. Firstly, the notion of using untrusted tools in a regular forensic laboratory is unthinkable. Who knows what such tools may do to the evidence and in what way they may contaminate the evidence. But in the digital world we have the luxury of working with copies of evidence. Even if we destroy a copy we can just make a new copy from our master copy, check the message digests and no harm has been done (besides our time that may have been wasted).

Another objection may be that the non-forensic tool we are using may be 'biased' in some way; for some peculiar reason it may find incriminating evidence, but miss the exculpatory evidence. Say, for example, A emails a ransom note to B and five minutes later emails a note that it was an April fool joke. This behaviour may still be illegal, but these messages may be interpreted very differently depending on whether both or only the first message are discovered. This objection clearly has merit in some cases. However, the sad reality is that in many (possibly most) current digital forensic investigations this makes little difference: In so many current investigations investigators are looking for contraband; if the suspect is guilty hundreds (or more) of examples are typically found. Exculpatory evidence (if it can exist) will have a very different form from what is being searched for. Say the disc contains many illegal (or unlicensed) MP3 files. Then it does not matter whether we find all of them; yet another MP3 will not serve as exculpatory evidence. Exculpatory evidence may exist in the form of a letter from a copyright holder granting permission to the suspect to copy their MP3 files without licences for, say, research purposes. This letter will typically be produced by the other party to explain the presence of the files. However, our claim that *any* tool may be used is dangerous when it is necessary to find *all* occurrences of the data of interest, or if the want to conclude that the data does not occur on the media at all. For such cases we need a tool we can trust; however, even for such cases there is no reason to use non-forensic tools if using them holds some benefit — such as the ability to find at least some occurrences faster than the trusted tool.

The final objection against the use of *any* tool to be considered may come from those who infer that our untested tool may go outside the boundaries of what we are legally allowed to access. This certainly is not the intention. The proper analogy to use when using these non-forensic tools is not the physical forensic laboratory, but the police officer who searches a room for evidence. This can only be done once an appropriate warrant has been issued and then the search has to be confined to the limits set out in the warrant. If this officer wants to use a flashlight to look into a dark corner of the room, it is ridiculous to require that it has to be a forensically sound flashlight. If the officer wants to read a label on a box that may be accessed and needs reading glasses, there is no need to ensure that they are forensically sound glasses. But when the officer looks into a cupboard that is beyond the limits of the warrant, evidence obtained will be inadmissable (in addition to punitive measures against the officer that should result). So, when using arbitrary tools to search data it is necessary to ensure that the limits of the warrant are respected. In many cases tools (such as `grep`) are simple enough to restrict to search within limits. Alternatively, the ‘forbidden’ areas of the disc may be redacted or the allowed areas may be copied to a clean disk. Either option, if executed correctly, will avoid any possible problems.

One practical consequence is this: If the investigator gives a copy of the (redacted) evidence to his or her sysadmin who is a Unix toolset, `bash` and scripting guru with the request to use his or her ingenuity to find the search pattern, whatever is found ought to be admissible. (This of course assumes that the sysadmin is authorised to access the evidence.) Note again that what the sysadmin does is not science — irrespective of how brilliant the search strategy may be.

We have spent an inordinate amount of space to the simple issue of searching for specific data in some data set. However, my sense is that most current forensic investigations occur in this space and that many who are looking for the *science* in digital forensic science are looking for it in this space. To illustrate the first point just consider the types of investigations that fit in this category. It includes searching for contraband, deleted logs, entries in the registry that indicate (former) presence of a specific program or device, credit card numbers, IP addresses, events in logs, events or modified files within some time period, fragments of known files and many more. Science may play a role in optimising the search strategies. However, the forensic investigation does not pose any specific requirements.

Therefore it seems inappropriate to consider ‘forensic searching’ as a relevant problem area for this scenario. Some search algorithms may hold certain benefits for forensics (and quite possibly other fields); for example strategies that yield initial results early in a specific search domain may be beneficial.

However, the requirements change when it is necessary to know that the search pattern does not occur on the disc at all or to find all instances of the search pattern. Similarly, issues arise when there is only sufficient time available to search a fraction of the available data. These cases are revisited after step 3 of the scenario has been considered.

Step 3 entails proving that the search pattern exists or (equivalently) revealing the details of the found search pattern (by, for example, revealing the credit card number found if a pattern conforming to a credit card number was used as search pattern). In its first form the requirement is clearly that a decision problem has to be answered: Does the given search pattern occur on the disc? Decision problems are well known from the field of algorithmics [12] (or computational complexity). And, from that same field we know the second formulation above is computationally equivalent to the decision problem. And thus we find ourselves with a problem for which a solid theoretical framework exists and can be answered in a scientific manner. In the scenario under discussion the question about the presence of the search pattern may be answered positively in an incontrovertible manner by simply pointing to where the data occurs on the image. Formulated in its current form the problem is tractable and answerable in absolute terms. The error rate is 0%. The forensic scientist can answer this question with absolute scientific certainty in the witness box.

If scenario 1 deals with the possession of contraband, finding contraband on the disc allows the prosecution to introduce the disc image as evidence. If contraband has not been found it is possible to simply not introduce that disc as evidence. However, the defence is potentially faced with a bigger challenge: they want to prove that no contraband occurs on the disc (or any other disc either). Suppose that the message digests of files containing contraband are known. Then they simply have to compute the message digests of all the files on the disc and show that none of those digests corresponds with any of the contraband digests. This is again clearly a decision problem. However, this pushes the ‘burden of proof’ to step 2 of the scenario and there is no step 3 where one can simply point to the fact that nothing has been found. Ideally the defence needs to know that their search algorithm is correct and that the search problem itself is tractable. The issue of correctness may again be addressed from the perspective of algorithmics where the algorithm is formally proven correct (and where the accuracy of the algorithm is therefore 100%).[3] A less desirable alternative is where trust develops in a certain search tool where opposing parties use it (and other tools) over many years and nobody finds any contraband missed by the other party. However, this only becomes scientific at the point where one can move from mere induction (‘it has worked thus far and will therefore probably work in the next case as well’) to where one may express one’s confidence in the tool in scientific terms. Formal testing of the tool seems useful in this regard.

In addition to correctness the defence in our example ideally wants the search problem to be tractable (or even if it is tractable in general, they want the answer to be available before it is needed — for testimony in court, for example). As indicated by Cohen [5]

[3] Note that correctness of the algorithm does not ensure correctness of the program; a simple option is to use multiple independent tools in parallel with the (probably valid) assumption that these independent programs will not contain coding errors that let them all fail in the same manner. However, this brings us back to an assumption, rather than a scientific fact. A deeper review of the field of software correctness is required than what can justifiably be provided here to be certain that the tool is correct.

the field of computational complexity may provide us with the answer to the dilemma of whether it is even worth starting the computation. However, there may be another alternative available: a probabilistic algorithm may provide an answer that is correct with a given certainty. Executing the probabilistic algorithm repeatedly increases the certainty (or finds a counterexample). If the problem is intractable, probabilistic algorithms may provide us with a scientific answer with a quantifiable error rate. Even if the problem is tractable but requires more time than is available, it may be possible to use a much simpler probabilistic algorithm and run it repeatedly. This will again yield a scientifically valid answer with a quantified probability of being incorrect.

This, at long last, brings us to the end of scenario 1 that set out to locate or prove the absence of some data on a disc image in a scientific manner. We now turn our attention to just one other scenario that illustrates a different case where the craft may be turned into science.

Scenario 2 deals with file carving. File systems organise files in blocks, sectors, clusters or some other units (henceforth just referred to as *blocks*). Files typically consist of multiple blocks that are linked together using metadata. If these links are destroyed the file is effectively lost even though the file contents may physically still be present on the disc. The links may be lost because of an attack or some accident. It is, for example, possible that a user deletes a file because it is no longer deemed necessary. Deleting the file typically deletes the links, but not the block contents. If it turns out that the information in such a deleted or lost file is important the question arises whether the blocks can be reassembled into the initial file. Such reassembly is known as *file carving*. Note that while the blocks are unlinked some of them may be reused for other files; therefore it is sometimes at best possible to carve a partial file.

Obviously file carving requires deep knowledge of the details of file systems. The carver needs to know how the metadata links blocks together in the specific file system as well as the other minutiae of the file system. In addition the carver needs to know the details of the file formats of the files being carved to recognise neighbouring blocks. In essence the carver is solving a jigsaw puzzle that has many extraneous pieces and where a few required pieces may be missing.

Now suppose that the carved file is used as evidence in a court case. Does the carver have scientific grounds to claim that the file has been reconstructed correctly? An intuitive answer may be that the mere fact that, say, a JPEG file that has been reconstructed from blocks scattered over a disc now successfully opens in an image viewer is sufficient evidence that reconstruction was done correctly. It seems just too improbable that a file with an incorrect block somewhere will still 'work'. But suppose the disc contained several versions of a given file with only minor differences between the versions. Is it not then possible that the reconstructed file may contain blocks from different versions forming a carved file that never existed in that exact form? And can it be guaranteed that there are no other situations where a combination of inappropriate blocks may seem like a valid file?

A somewhat different approach is to ask what can be said about the reconstructed file that is scientifically true (and hence truly forensic science). One example is the question whether the reconstructed file conforms to the expected format. File formats are often specified using some formal notation, such as a grammar. If not, it is in many cases possible to create a grammar-based specification from whatever specification exists — possibly even from reverse engineering an authoritative piece of software that creates such files. The notion of syntax checking is well understood from the field of compiler construction. The question whether the reconstructed file is syntactically correct is therefore one example of a question that may be phrased as a decision problem and answered in a scientific manner. Many other properties may be checked in this manner. If certain values in a file are expected to have some relationship to one another this may be verified. The time stamps in a log file, for example, are supposed to be ordered according to time. In some (rare) cases it may be possible to show that no other blocks on the disc can possibly be part of a file of the given type. It may be possible to show that the blocks in the carved file are arranged on disc in a manner consistent with the block allocation strategy used by the operating system. It may be possible to allocate all blocks on the disc to files that are all syntactically, semantically and positionally correct. It may be possible to test all permutations of blocks (possibly after filtering those out that cannot possibly form part of the given file type) and show that the reconstructed file is the only permutation starting from some block that yields a syntactically valid file. Based on these scientific facts the expert may then offer a professional *opinion* about the correctness of the reconstructed file. The opinion may take into account the complexity of the format, the consistency of the reconstructed file with other available information and other attributes of the file the professional may deem relevant. It is important to distinguish between science and opinion though. Different forensic scientists should arrive at the same scientific answers to questions that can be answered by forensic science. If their opinions differ, so be it. They are opinions after all, and should have less evidentiary weight than scientific facts. However, note that if it can be shown that an opinion is inconsistent with facts, that opinion is refuted.

Note that this second scenario conveniently ignored the fact that many real programs do not faithfully implement file format standards. It is therefore possible that an original file may not pass the syntax check — and if such a file is reconstructed correctly it should fail the syntax check. However, this may possibly be addressed by not only using *de jure* specifications, but also *de facto* specifications. We ignore this issue in the remainder of the paper.

To conclude note that this distinction between fact and opinion is also present in traditional (physical) forensic science. The DNA scientist cannot ‘place’ a person at a crime scene. The scientist can state as a scientific fact that, say, a hair and some saliva come from the same donor. The additional (non-scientific) information that the hair was found at the crime scene and the saliva sample was obtained from the suspect (as well as some convincing argument that there is no other logical explanation for the suspect’s hair to be at the crime scene) is required to be certain that the suspect was indeed at the crime scene.

4. DIGITAL FORENSIC SCIENCE

The two scenarios discussed earlier in this paper show that decision problems may indeed provide the scientific basis for digital forensics for some cases; in such cases decision problems may be used to distinguish between forensic *science* and expert *opinion*. Those two scenarios are insufficient to claim that decision problems can be used as the underlying theoretical base of all of digital forensic science. However, it is a strategy that seems worth exploring. As Garfinkel [8] points out, locating incriminating information (such as contraband) in large datasets was the original challenge for digital forensics and the field needs to urgently cast its net wider to remain relevant. A digital forensic science based on decision problems (and the accompanying algorithmics or complexity theory) provides much scope for forensics to develop beyond its current state. Garfinkel’s identification of the original challenge of digital forensics coincides with scenario 1 provided earlier in this paper. Much of digital forensics was originated by finding ways of solving crimes (or finding digital evidence) that may be useful to address such crimes. If we decide that decision problems underly digital forensics it also becomes possible to develop digital forensic science from the top down by determining what can and what cannot be proven by viewing the extensive body of knowledge about tractable and intractable problems from this new perspective.[4]

An initial argument that decision problems should form the basis of digital forensic science may read as follows. Many (or most) digital forensic investigators will be comfortable with characterising the examination process as a set of hypotheses that are tested and then rejected or not rejected[5] The work by Carrier [4] is a seminal text that frames digital forensics using hypothesis testing. The idea of using decision problems and determining the answers they yield (or concluding that they cannot be answered) appear rather similar to hypothesis testing. Hypothesis testing, however, typically assumes natural variation and testing a hypothesis is about determining whether minute differences between an observation and an ideal value may be ascribed to this natural variation. Digital data, on the other hand, being discrete, does not display such natural variation. The millions of statements produced by a bank on a monthly basis are not a little wrong each month because of natural variation. If the statement is not exactly correct it is because something is amiss. Natural variation may be introduced in a digital system because of external physical influences. The time that data needs to traverse a network is one such example; this may result in a natural variation between times recorded in a log at the transmitter and times recorded for the same messages at the receiver. However, it is not clear that these differences are indeed *natural*. The digital realm is one that is inherently artificial. Users can influence congestion on the network and hence the differences in times. In fact, such times are often affected by multiple natural and artificial causes that make it impractical to measure a given characteristic and associate it with scientific accuracy with some specific cause or condition. Decision problems may therefore fit digital data better than hypothesis testing would for forensic purposes.

Note that decision problems, just like hypotheses, do not prescribe *how* an examination should be conducted, but clearly delineates *what* may be offered as evidence. An 'accepted' hypothesis makes a truth claim — as does a decision problem that has been decided.

The remainder of this section reviews the (well known) classes into which decision problems fall [12]. The intention is twofold. Firstly, it shows how much of the field remains unexplored from a forensic perspective and therefore indicates a direction into which future forensic research may grow. It also shows how error rates naturally become an issue when decision problems become more complex. This potentially lends some credibility for a forensic investigator who claims 100% accuracy for a result based on a simple decision problem (relative to a whole field of varying complexity where error rates are no longer zero).

In general decision problems fall in one of four categories: they are decidable in polynomial time, probabilistically decidable in polynomial time, intractable or undecidable. The second category in this list gives us our first glimpse of what error rates may mean in the context of decidable digital forensics. We return to this topic below.

When the question of interest is polynomially decidable there is no inherent need to quantify error rates. However, even polynomial time algorithms may sometimes be too

[4] Note that Garfinkel's plea for an extension of digital forensics refers primarily to the extension of technology used for digital forensics — that is, to digital forensic craft rather than digital forensic science.

[5] Note that the mere use of hypotesis testing (outside a body of theory) would not be sufficient to make an activity scientific. For more details see [2, Chapter 5].

'expensive': to search a petabyte of information in $O(n)$ at 1 megabyte per second will take just over 30 years. A probabilistic algorithm that does not sample every byte of the petabyte and that yields a result that is reliable enough but terminates within some reasonable time will be preferable over the absolutely correct $O(n)$ algorithm. In general, given the large data sets that digital forensics often has to deal with, it may be necessary to approximate the algorithm with an even faster one (one that, for example, only uses a fraction of the n inputs) if the results of the probabilistic algorithm are correct enough — that is, if the error rate can be quantified and it is deemed small enough to sufficiently substantiate the claim that it supports.

Probabilistic algorithms (also known as randomised algorithms) are algorithms that use a random number to determine their behaviour. The type of probabilistic algorithm alluded to above is a *Monte Carlo algorithm* — one that always terminates in polynomial time and produces an answer with a known error rate. Monte Carlo algorithms may be true-biased, false-biased, or unbiased. When a true-biased algorithm returns *true* the answer to the problem is indeed *true*, which is often written as *yes*. When it returns *false* (or *no*), however, it may be wrong with some known (small) probability. The converse is true for false-biased algorithms. Unbiased algorithms may yield incorrect results (with some small probability) when they return either *true* or *false*. Monte Carlo algorithms are deigned such that the random number determines the execution of the algorithms, such that one execution of the algorithm is independent from the next and the probability of error from two executions of the algorithm are then the product of the probability of error during a single execution. To reach a particular level of certainty it is necessary to repeat the execution of the algorithm a sufficient number of times so that the combined error is small enough. Note that, if a true-biased algorithm returns *yes* during any execution, the final answer is *true*. It is only when it repeatedly returns *no* that the answer is *no* with a probability of error e^n, where e is the probability of error for a single execution and n is the number of executions. The same applies to false-biased algorithms, except that a *no* result is certain and a *yes* result is reached with a margin of error. We do not consider two-sided errors further in the current paper.

The class of problems that are solvable by probabilistic algorithms are known as the *bounded-error probabilistic polynomial* ($\mathbf{BPP}$) class of problems. Let $\mathbf{P}$ (as usual) denote the class of problems that are solvable in polynomial time. Then we contend that probabilistic algorithms are indicated for any problems in $\mathbf{BPP} - \mathbf{P}$. (Note that is is possible — and many indeed conjecture — that $\mathbf{BPP} = \mathbf{P}$.) As noted, Probabilistic algorithms may also be useful for problem in $\mathbf{P}$, where available time simply does not allow execution of an exhaustive algorithm, even though it may be tractable.

5. Alternative perspectives on error rates and digital forensic science

As noted earlier, others have proposed strategies to deal with accuracy (or known error rates) of digital evidence. Cohen [5], for example, notes that the error rates of CPUs are known and suggests that this may be used to quantify the accuracy of digital evidence. However, such random CPU errors do not necessarily translate to specific error rates in digital evidence. In many cases data is, for example, subjected to error checks (such as integrity checks in databases, digitally signed messages, ordinary parity checks for memory, and so on). Some errors may cause a program to crash, rather than produce incorrect results. Yet other errors may be inconsequential — such as when the colour of a single pixel on a screen is somewhat wrong. The fact is that such errors are extremely rare and of the few that occur, many will have no impact on evidence that is collected. If it does affect the evidence it is possible that it may affect it in such a way that it is obvious that something is wrong. Once all of this is taken into account it is easy to see that errors may occur, but that this will occur so rarely that it is safe to ignore the possibility. However, with all these factor impacting on the error rates it becomes impossible to quantify the known error rates of our forensic techniques.

An earlier approach to describe (rather than quantify) error rates is Casey's certainty scale. It, for example, postulates that an event that has been logged in two independent logs may be accepted as fact with more certainty than an event only logged in one log (but this certainty will still be very low if the two logs are not properly secured). While this makes sense, it is not a scientific truth. An event logged in a number of highly secure, independent logs may lead to a high level of certainty that it really occurred. But it is possible that the administrators of all those systems colluded and entered a fake entry in all the logs. In contrast, an event logged in a single, unreliable log may indeed have occurred. The higher degree of certainty is based — at least in part — on the assumption that a group of trusted individuals associated with independent systems will very rarely collude. While this is probably true, the average digital forensic scientist is not qualified to testify about human nature — and questions of human nature should arguably not be part of the domain of digital forensics. In any case, it seems unlikely that even a social scientist will be able to accurately estimate this probability. This does not mean that Casey's certainty scale is useless; it does mean that the certainty scale may be unsuitable to derive scientific facts. It may be very useful for an expert to express an opinion once the scientific facts have been determined.

Finally, Garfinkel et al [7] emphasise the ability to independently verify test results as the hallmark of science and encourage the development of standardised corpora that may be used for independent testing (and provide some such corpora).

6. CONCLUSION

This paper identified a possible basis to ensure that digital forensics is indeed scientific, namely decision problems from the field of algorithmics. It illustrated that decision problems may indeed be useful for some investigative problems. Decision problems also help to talk about facets of science such as *truth* and *error rates*. It provides a possible explanation for why it is currently hard to talk about such issues, because current research has only scratched the surface of this domain (once such research is rephrased in terms of decision problems).

Decision problems may be helpful to guide the construction of digital forensic tools that can be certified as reliable.

Much remains to be done. Many other investigative scenarios need to be considered to determine whether decision problems form an appropriate solution, or whether there are better options to obtain scientifically valid evidence for such scenarios. Decision problems also potentially delimit the scope of digital forensics and delineation is often a source of contention. Do authorship attribution [14] and source camera identification [17], for example, still form part of digital forensics or are they really about human and physical attributes that just happen to be represented in a digital format, but may just as well have been presented in a non-digital manner? If the proposal contained in this paper is accepted as a viable option by the digital forensics community only time will provide definitive answers to these latter questions.

REFERENCES

[1] Frank Adelstein. Live forensics: Diagnosing your system without killing it first. *Communications of the ACM*, 49(2):63–66, February 2006.

[2] Mario Bunge. *Philosophy of Science: From Problem to Theory*, volume 1. Transaction Publishers, 1998.

[3] Michael A. Caloyannides. *Digital "Evidence" is Often Evidence of Nothing*, pages 334–339. IGI, 2006.

[4] Brian D. Carrier. *A Hypothesis-based Approach to Digital Forensic Investigations*. PhD thesis, Purdue University, 2006.

[5] Fred Cohen. *Digital Forensic Evidence Examination*. Fred Cohen & Associates, 3rd edition, 2012.

[6] John J. Furedy. The North American polygraph and psychophysiology: Disinterested, uninterested, and interested perspectives. *International Journal of Psychophysiology*, 21:97–105, 1966.

[7] Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. Bringing science to digital forensics with standardized forensic corpora. *Digital Investigation*, 6:S2–S11, 2009.

[8] Simson L. Garfinkel. Digital forensics research: The next 10 years. *Digital Investigation*, 7(Supplement):S64–S73, August 2010.

[9] Paul C Giannelli. Wrongful convictions and forensic science: The need to regulate crime labs. Working Paper 08-02, Case Western Reserve University, 2007.

[10] Tal Golan. *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America*. Harvard University Press, 2007.

[11] Stefan Gruner. Software engineering between technics and science — recent discussions about the foundations and the scientificness of a rising discipline. *Journal for General Philosophy of Science*, 41:237–260, 2010.

[12] David Harel. *Algorithmics: The Spirit of Computing*. Pearson Education, 2nd edition, 1992.

[13] Douglas R. Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Harvester Press, 1979.

[14] Patrick Juola. Authorship attribution. In *Foundations and Trends in Information Retrieval*, volume 1, pages 233–334, 2006.

[15] Roger Koppl and Monique M. Ferraro. Digital devices and miscarriages of justice. *The Dayly Caller*, 2012. Online: `http://dailycaller.com/2012/06/15/digital-devices-and-miscarriages-of-justice/`.

[16] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 3rd edition, 1996.

[17] Martin S Olivier. Using sensor dirt for toolmark analysis of digital photographs. In Indrajit Ray and Sujeet Shenoi, editors, *Advances in Digital Forensics IV*, pages 193–206. Springer, 2008.

[18] Michael J Saks and Davis L Faigman. Failed forensics: How forensic science lost its way and how it may yet find it. *Annual Review of Law and Social Science*, 4:149–171, 2008.

[19] John Smeaton. *Reports of the late John Smeaton, F.R.S., made on various occasions, in the course of his employment as a civil engineer*, volume II. M. Taylor, 2nd edition, 1837.

[20] U.S. Supreme Court . Daubert v. Merrell Dow Pharmaceuticals, inc., 509 U.S. 579 (1993). Technical Report 92–102, Certiorari to the United Sstates Court of Appeals for the Ninth Ccircuit, 1993.

ICSA RESEARCH GROUP, COMPUTER SCIENCE,
UNIVERSITY OF PRETORIA, SOUTH AFRICA
E-mail address: `ms.olivier@olivier.ms`
URL: `http://mo.co.za`