Combining fundamentals, traditions, practice, and science in a digital forensics course

Martin S Olivier
Department of Computer Science
University of Pretoria
Pretoria, South Africa
ms.olivier@olivier.ms

"Of the various facets of underresourcing, [we are] most concerned about the knowledge base. Adding more dollars and people to the enterprise might reduce case backlogs, but it will not address fundamental limitations in the capabilities of forensic science disciplines to discern valid information from crime scene evidence" [7, p.15].

ABSTRACT

Digital forensics is still in a transition period from being a *craft* practised by technical people, to a *science* that can provide scientifically justifiable answers about root causes. The problem faced by the educator is that limited time is available to teach a course that is inherently multidisciplinary (including facets of technology, science, law and reporting). The specific concern addressed by this paper is the fact that some rather detailed procedures have developed over the years around the digital forensic craft; while it is not the intention of this paper to dispute the value of those procedures, a course presented in a limited time (such as a semester) may be consumed by the craft and not provide an opportunity to consider the scientific aspects (that are indeed still in their infancy, but that are essential for the future of digital forensic science).

Note that the inherited terminology of a previous era is still important. Some of the old techniques are still valid and practical, and continue to be used. However, where the validity or practicality of older techniques are beginning to be questioned (if not rejected outright), new developments often occur against the backdrop of the older techniques — and therefore the terminology and other details related to older techniques still need to be studied in order to comprehend new developments.

The thesis of the paper (as well as that of the course the paper describes) is that, while digital forensics necessarily focuses on details, it is possible to create an environment with reduced (but still realistic) complexity. By removing details it should be, the paper contends, able to design a course that covers a greater scope of topics without sacrificing the inherent detailed-oriented nature of digital forensics.

More specifically, the paper proposes a course that employs a breadth-first approach that, as briefly as possible, explores the scope of digital forensics, but in a realistic rather than abstract context. Whereas most courses have to choose between breadth and depth, this paper seeks to find a trade-off that achieves sufficient coverage of the subject area in sufficient detail during, say, the first half of a semester. In the second half new developments — in particular in digital forensic *science* — may be explored, with the students able to contextualise such forays while appreciating the low-level complexity inherent in such a new development.

The paper considers the requirements of using a simplified environment. It concludes category of textbook, a specific product (such as the Raspberry Pi) and an acute awareness of new developments in the field can be used to establish a suitable environment.

The course described in this paper was tested in 2013 — not for the sake of research, but as a practical solution to a dilemma that educators increasingly face as the scientific part of digital forensics develops. The course yielded some unexpected benefits (and challenges). The nature, rationale, unexpected benefits and challenges are presented in this paper as a case study. Formal research is required to confirm the envisaged and informally observed benefits and solutions to challenges.

SACLA 2014, 25-26 June, Port Elizabeth, South Africa

Categories and Subject Descriptors

K [**3**]: 2—*Computer and Information Science Education*; K [**5**]: m— *Miscellaneous*; K [**7**]: m—*Miscellaneous*

Keywords

Digital Forensics, forensic Science, Raspberry Pi

1. INTRODUCTION

The theme of the 2014 SACLA conference is "ICT Education in the Cyber World". The real world and the cyber world have merged such that the two often cannot be distinguished from one another anymore. One of the examples that illustrates this convergence very clearly is seen on a daily basis in courts of law. Very often actions in the physical work leave traces in the cyber world and these traces often play a crucial role when the innocence or guilt (in the physical world) of a party has to be established. The converse may also be true, but is not explored in the current paper.

The forensics example is one where education in the physical world is lacking — see below. The move to the digital world aggravates the situation and therefore the need for education becomes even more urgent.

The most critical assessment of (mostly physical) forensic sciences was published five years ago by the National Academy of Sciences [7]. While this report focussed on deficiencies in forensic sciences in the United States, these deficiencies are, without doubt, universally present. The report devotes one of its 11 chapters to the theme of forensic education (amongst others, as part of the solution to the problems identified). In addition, the word

education occurs repeatedly on almost every page of the report. In February of 2014 the United States announced its response to the report — an extensive restructuring of its forensic landscape that will arguably reverberate throughout the world. Any weakness with some forensic approach identified anywhere in the world is bound to be discovered by a party who needs it and used (rightly or wrongly) to cast doubt on incriminating or exculpatory evidence that needs to be discredited. Clearly, in such cases, insight is required by those who collect and analyse evidence, reach conclusions (or opinions) based on such evidence and those who have to weigh the claims based on the evidence to determine the outcome of a dispute that hinges on such evidence. However, few countries have the means to respond on the scale the United States is responding.

Such limiting factors include economic realities (including limited resources at universities and insufficient career opportunities). These factors stand in the way of the introduction of full digital forensics programmes at universities in many countries. Therefore digital forensics education currently often has to be limited to one or two elective semester modules of a host subject (or degree) such as computer science. In a number of cases active research programmes in digital forensics do exist at postgraduate level, but then the emphasis is on research rather than education.

This paper explores an option that provides computer scientists with a solid educational foundation in digital forensics, given such limiting factors that apply in developing countries, such as South Africa.

The paper's challenge is to find an appropriate balance amongst the various challenges. The thesis of this paper is that a suitable brief introductory course to digital forensics should strive for a balance between the various forces at play in the discipline. It should, for example, find a balance between the *science* and the *craft* in the discipline. While the science part should obviously form the focus, much of the discipline is practised in the field by craftsmen (and a few women) — relying "on apprentice-type training and a guild-like structure" [7, p.15]. Many of the popular tools are arguably focused on the craft, rather than on a science; agencies "cannot verify the results with their equipment and in many cases rely solely on an independent validation study of other peoplesŠ equipment. This creates the issue of tools in reality never being tested" [3].

Secondly, a balance needs to be found between complexity and simplicity.

The remainder of this paper is structured as follows: Section 2 identifies the primary forensic education requirements using the National Academy approach as point of departure. The following section considers the balancing act that is required in the transition from craft to science; neither can the standards and practices that developed with the craft be summarily dismissed, nor can one afford to attend to those practices and standards in detail, because time spent in this manner on aspects that are not scientifically justifiable is time that could have been spent on the primary objective; however, digital forensic science develops as a discourse with its predecessor craft, and therefore has to be able to 'speak the language' of the craft. Moreover, aspects of the craft are still considered 'best practice' and often when scientists ignore such best practices they do so at their own peril. Section 4 considers the shift in what should be considered as fundamental knowledge possessed by computer science students. Section 5 presents the core thesis of this paper: the selection of appropriate hardware is a core enabler of the transition that this paper envisages. Section 6 discusses the simple investigative process used as the basis for the proposed course. Section 7 then zooms in on the examination phase of the process, because this is where digital forensic *science* happens in particular. Section 8 briefly mentions the final of the four phases used in the process model introduced in Section 6 — this is done primarily for the sake of completeness. Section 9 concludes the paper.

2. EDUCATIONAL REQUIREMENTS

The National Academy, as noted above, highlights some of the requirements of forensic education early in the report [p.26][7]: "Forensic science examiners need to understand the principles, practices, and contexts of scientific methodology, as well as the distinctive features of their specialty." For the sake of brevity we will assume that embedding a digital forensics module in a degree programme means that the student will be familiar /with these concepts because they ought to be encountered throughout one's studies. However, the fact that this assumption is not necessarily valid in computer science should be noted [14].

The National Academy report continues (p.26): "Ideally, training should move beyond apprentice-like transmittal of practices [which usually happen in crafts] to education based on scientifically valid principles." This is arguably *the* key concern that faces the digital forensics community. This is therefore the primary concern addressed by this paper. The transformation of digital forensics from a craft to a science started about a decade ago. Currently digital forensics contains many aspects that are arguably rituals that have indeed been transmitted from master to apprentice in years gone by. The research done during the decade done since the (scientific) emergence of the field has done much to improve its scientific basis. (See, for example, the paper by Olivier and Gruner [19] on the *scientificness* of digital forensics, but also the — still current — critique of the field by Cohen [6]).

The report emphasises this point (pp.26–27): "a trainee should acquire rigorous interdisciplinary education and training in the scientific areas that constitute the basis for the particular forensic discipline ..." In contrast, SWGDE admits that "[a]ssumptions in digital forensics have not been subject to the rigor of a scientific vetting process" [9, p.2], which poses a significant challenge for digital forensic education.

One or two modules in a degree does not turn graduates into forensic scientists; one therefore has to consider whether such modules have merit at all. The inclusion of digital forensics in, say, computer science may be justified on a number of grounds. Firstly, it at least exposes students to the notion of digital forensics as a science; such awareness should have a positive impact on legal proceedings in the country since awareness should at least in due course cast some doubt on 'rogue' digital evidence that may otherwise have been accepted by a court despite a lack of a scientific foundation. Secondly, such courses provide a starting point for those students who are intrigued by the topic and decide to obtain the necessary professional training elsewhere and/or conduct research in this area. Finally, digital forensics creates a platform for critical reflection on other areas of computer science — on the trust placed in computed results, on the need for designing systems that, when necessary, will help rather than hinder investigations and on dispelling the belief that misdeeds on computing platforms cannot be traced to the originator if the originator is just 'clever enough.' It also empowers the individual (or organisation) who is falsely accused and now aware that exculpatory evidence could have been created in the form of a variety of often surprising digital artefacts spread around the system.

Note that a number of universities and other institutions have developed curricula. In fact the ASTM published a standard containing model curricula [2]. However, in all cases the curricula assume that time and money are available, while these are precisely the issues that the current paper claims are not readily available and therefore are serious constraints to be overcome.

3. BALANCING SCIENCE AND CRAFT

Fitting a digital forensics module into one or two semesters (along with a number of other modules) presents a number of challenges.

The first of these challenges is the division of time spent on applied work and time spent on new developments in the field. Prior experience leads me to the belief that students struggle to properly comprehend simple concepts such as drive imaging (see below) if it is only referred to in passing and I postulate that this is likely to worsen in future. Section 4 elaborates on this point. On the other hand, a module that only focuses on the practical will omit much of the new insights in digital forensics. As an example, while drive imaging may still be considered as the gold standard for evidence acquisition, it is becoming less practical in real-live situations every day. In two classic pieces in *Communications of the ACM* [1] argues in favour of so-called 'dead' analysis, while [5] argues in favour of the 'opposite' technique, 'live' analysis. However, each of those arguments in favour of a given approach is based on discrediting the alternative approach, which, of course, leaves one with a conundrum (about which some remarks will be made below). In the same manner, many of the latest developments are based on earlier techniques that need to be understood: one of the obvious challenges of cloud forensics is the fact that it becomes virtually impossible to remove a drive of a (virtual) computer in the cloud to image it — something that was (almost) always standard practice for real computers that needed to be examined. Thus there is clearly a need to study the old and the new, which often (but not necessarily) implies a need to study the applied and the theoretical.

The thesis of this paper is that the key to solving such problems is the removal of unnecessary complexity to the extent possible. This is a relatively common approach in computing. The thesis that underlies Tanenbaum's Minix is that a simplified operating system provides students with the possibility to actually apply operating systems theory without being distracted by the complexity of real operating systems [23]. Over the years a number of programming languages and other tools have been used to teach students programming concepts without bogging them down in the details of 'real' languages, with (the turtle graphics part of) Logo and Alice [8] examples that readily come to mind. MIT's programmable brick (the basis of the LEGO MINDSTORMS product range) enables children (of all ages) to build a range of behaviours (for robots, ubiquitous computing, and so on) without having to deal with the complexities of manufacturing or the complexity of programming devices often used in manufacturing and similar environments).

The problem the paper addresses is one of finding an appropriate tool that provides sufficient complexity to teach digital forensics (including new developments), yet strips away as much unnecessary complexity as possible.

Many forensic processes have been proposed and it is not our intention to review them here. With simplicity in mind, the most basic of the process models is used in the remainder of the paper. It consists of the following stages:

Pre-incident

- Preparation

Post-incident

1. Collection / acquisition
2. Examination / analysis
3. Reporting

These phases naturally distribute the introductory material over four or five weeks of a 14 week semester. Assignments (described below) provided the required depth.

The National Academy report [7][pp.85–86] alludes to such phases: "In order for qualified forensic science experts to testify competently about forensic evidence, they must first find the evidence in a usable state and properly preserve it." Locating and preservation "are important matters having to do with the proper 'processing' of forensic evidence. The lawŠs greatest dilemma in its heavy reliance on forensic evidence, however, concerns the question of whether — and to what extent — there is *science* in any given 'forensic science' discipline" [emphasis in original].

During the initial phase of the course we needed a textbook that would provide coverage of the entire discipline: preparation, acquisition examination and even reporting. We settled on Sammons's *The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics* [21]. The first two thirds of the book worked really well to provide the broad coverage we required (and even some low-level details about Windows forensics). However, the final third of the book addressed security issues presented as forensics topics. The book also did not cover reporting sufficiently for our introductory purposes.

The remainder (in other words, bulk) of the course focussed on in-depth treatment of new developments in the field using the research literature as source. This included critical coverage of work that emphasises science, including work by Gladychev [11], Carrier [4], Cohen [6] and Olivier [17, 15], as well as advanced aspects taught by a colleague. However, the current paper is not intended to address that part of the course. Its primary purpose is to demonstrate how the pre-scientific craft may be taught in a relatively brief period, providing sufficient opportunity for the discussion of science in digital forensic science. In addition, as will be shown, this initial phase was designed to create some unease about existing practice by, primarily, continually raising questions of reliability in this initial phase.

4. ABSTRACTION AND THE BOUNDARY OF ASSUMED KNOWLEDGE

One of the key skills taught in computer science is that of abstraction. Parnas's [20] seminal concept of *information hiding* captures the essence of this notion in its very name. Slightly more recently the object-oriented idea of *encapsulation* reinforces it. *The cloud* is a current buzzword that originates from a similarly named earlier concept that emphasised the fact that the user did not have to be concerned about how things happen in this cloud; the user (or, more specifically, the customer) could simply use it. Many, many more examples could be listed.

In a similar vein, programming languages have changed people's concept of what a computer is. My 'native' programming language was assembly language [18], which means registers, program counters and control units are an inherent part of my notion of *computer* (which I obviously usually abstract away). Later generations used pointers and goto statements, which made the linear structure of memory obvious. The latest generation arguably thinks of a computer in terms of objects (and in some cases, visual objects. Often they still encounter computer architecture and assembly language somewhere, but it is no longer part of their core frame of reference.

In fact, early personal computers were supplied with just a programming language, which implied an understanding of almost the entire operation of the machine. Later systems expected users to configure software — again implying extensive knowledge of the system. Even as so-called shrink wrap software became popular details about servers (say, for email) needed to be specified. A current user touches a "button" in some "store" that provides "apps'. Often a username and password complete the entire installation process.

These developments are not limited to application software either. Whereas programmers in the olden days had to build

libraries and link software, a modern programmer registers a project, enters code in various blocks, and pushes a button to build the entire system.

There was a time when users could poke characters into a screen buffer to display them on a screen. Later it became possible to turn pixels on and off in a similar manner creating simple graphics. Colour graphics arrived next. A current programmer works in terms of abstract concepts, such as canvasses or display boxes, rather than screen buffers.

Where current forensic tools work in terms of bits and bytes (as well as structures built on top of those bits and bytes) such bits and bytes are foreign concepts to the current student. Where a new programmer could, in years gone by, easily write a function to extract a substring from a longer string, many current students seem to consider such a programming task as esoteric. I was shocked recently when I asked a fourth year class what happens to a value when it is shifted left by a single bit — and they did not know. On the positive side, current students are able to build impressive systems that would have been considered impossible for a student to do just a few years back.

The question is what the basic level is that one should understand to be able to fully explain (or find the definitive root cause) on a system. For the current generation of digital forensic experts that is arguably the bit and byte level and, where necessary, the assembly language level. Yet, that is not really the lowest level. Firmware lies below assembly language and underneath it all are transistors (densely packed into integrated circuits). It is possible to build tools that will, say, carve a file from a medium with the click of a button — and without really understanding the mechanism used (and, in particular, understanding the limitations of such a mechanism). A case can be made that the bit and byte level is indeed *the* fundamental level to understand. This paper does not attempt to make that case; however, it uses that assumption as a premise. Given the state of student knowledge, this implies that attention should be spent on such low level concepts.

On the hardware level similar abstracting developments occurred: In past ages we regularly removed a disc pack from a disc drive to mount another one that was required. Threading a tape drive was a spectacular exercise — especially those that were semi-automated. The central processing unit (CPU) was an impressive box full of electronics.

When mini-computers came along many of the boxes were replaced by integrated circuits and, consequently, many of the wires disappeared. Personal computers typically reduced the number of boxes to one — and a rather small one at that. In many (perhaps most) cases today when someone buys a computer it consists of one aluminium (or plastic) unit that cannot be opened without causing permanent damage. Discs are replaced by solid state devices (SSDs). If such a device is not hidden in a permanently sealed case, an SSD may still look like a disc drive looked; however, it may be just a bank of integrated circuits.

Many will argue that it is necessary for forensic scientists to be familiar with hardware irrespective of its age, since it is impossible to predict what hardware a criminal will use. However, a fundamental difference is that bits and bytes are still used — whether the user knows it or not. Obsolete hardware is often discarded and play no role in current computing. On odd occasions one may work on a case where such odd hardware does play some role. However, one's understanding of everyday computing does not depend on such odd cases.

To make this point more concrete consider the "birth date" of the various bus technologies used to connect these discs to the computer. The details (and phrases represented by the acronyms are not important)[1]:

- First 8" floppy discs: around 1970
- MFM floppy discs: 1978
- MFM hard discs/ST-506: 1980
- SCSI: 1981
- IDE (later renamed to PATA): 1986
- FireWire/IEEE 1394: 1994
- SATA: 2003

In comparison, most of my students in my forensics class were born around 1991; they were about 12 when SATA debuted; their primary computer is usually a laptop. In fact, most of them have never touched a Molex connector.

A typical current forensic write blocker will contain SATA and PATA (and possibly other) interfaces. I argue that this is one of the complexities one can ignore for the sake of coverage of more important aspects. The student who does eventually work in a forensic context will readily acquire knowledge about old technologies in an apprenticeship fashion.

5. SELECTING HARDWARE

In order to make matters as convincing as possible, the primary hardware selected for the course should be (or resemble) a computer. Thankfully, such devices are common and in daily use. A smart phone is just one common example of relatively simple devices that are, in fact, computers. However, such phones are, in many ways, simpler (or less complex) than 'real computers'. As a start, their memory (or 'discs') has a very limited capacity compared to typical current computers. At the time of writing this, a 'real' computer often has disc capacity in the order of a terabyte. In contrast, smart phones are usually limited to about 16 to 64 GB of storage (but, in many cases, even less). The processors used in cell phones — even though they are often dual or quad core processors — are simpler and slower than those used by their bigger cousins. This not only leads to potentially simpler software (where 'bloat' cannot be tolerated), but also a thriving market for 'computers on a chip': integrated circuits that implement all (or almost all) the required functionality on a single chip. A number of ARM processors and the Intel Atom are the best known examples in these categories.

Unfortunately, cellular phones and their cousins (such as tablets) are notoriously hard to access on a low level. A whole branch of digital forensics — mobile device forensics — has emerged that primarily deals with the problems of accessing the memory on such devices. The current norm is to use a product such as XRY or Cellebrite — products that cost thousands of dollars (and often with some associated recurring cost). This defeats the objective of simplicity (since a complex piece of equipment used in the process either has to be 'understood' or dismissed as a black box, which flies in the face of a course intended to make the process comprehensible).

The common availability of these low-cost processors also lead to the development of a range of non-phone devices such as "TV sticks" that turn (almost) any television into a smart television. Personal experimentation has demonstrated to me that it is relatively easy to root such devices and replace the system with one of the simpler Linux distributions, turning the stick into a 'real' computer. However, this process sometimes does lead to problems; the quality control over these devices does not seem to be as stringent as one could hope for and I have bought some duds. (Some were easy to repair, such as a loose connection in the power supply, but still caused a significant degree of frustration.) In addition, the internal make-up of these systems often

[1]Dates obtained from Wikipedia; use with circumspection

change without notice, and a recipe that works for one device 'bricks' (or 'kills') the very next seemingly identical device.

One of the most interesting projects to emerge from this system on a chip trend is the Raspberry Pi. The fact that it traces its roots back to the BBC Micro adds some credibility. Briefly, for those unfamiliar with computing in the ancient early 1980s, the BBC wanted to produce a series on computing to familiarise the public with this new phenomenon called a personal computer. To make matters practical they wanted the public to be able to follow the program (and experiment further) using a similar computer to the one used in the series. The BBC commissioned a Cambridge-based company, Acorn Computers, to produce a machine that would ultimately become the BBC Micro. However, a problem arose: Intel was unable to provide Acorn with processors, so the engineers at Acorn decided to develop their own processors, to be called the Acorn RISC Machine, with the word *RISC* denoting its reduced instruction set architecture. These Acorn RISC Machine — or ARM — processors were developed and the BBC Micro was successfully produced. Production of ARM processors continued and, in an ironical twist of events, Intel was eventually forced to design an equally simple chip — the Intel Atom — to compete in the new market of low-power devices.

The first fact that makes the Raspberry Pi an interesting choice for a digital forensics course is that it is being built as a computer with a variant of Linux as its primary operating system. Out of the box (literally[2]) it boots to a Unix-like environment complete with a GUI. Some tools are replaced by lightweight alternatives, but the real tool is typically available for installation. A notable example is its included browser, Midori (which one can replace within minutes with (almost?) any of the standard (Linux) browsers, such as Chrome, Firefox or Opera).

The (henceforth) Pi is relatively affordable. At the time of writing, the falling exchange rate of the South African Rand against the British Pound is adversely affecting the cost of these units, but a basic Pi sells for about R450 and a basic case costs about R100 (although some very interesting cases are available for those who want to splurge). In its basic configuration it has HDMI and composite connectors for video output, a standard 3.5mm socket for sound, a micro-USB connector for power, two USB sockets for peripherals and a slot for an SD card. It also has some low-level connectors for those who want to interface stuff directly to the system board. The usual version is supplied with 512kB of RAM (which cannot be expanded).

Note that an online emulator is available and that students do not have to obtain a physical unit to participate in practical assignments. However, the physical unit does allow for rather useful demonstrations in class.

One of the first notable benefits for using a Pi for forensics training is the absence of any on-board long-term storage facility. It uses an SD card (to be inserted into the SD slot) as its sole disc. To get a 'clean' machine one simply has to write a copy of the operating system to a clean(ed) SD card and, when booted from this card, the system has no memory of any earlier existence.

Figure 1 depicts a typical (staged) crime scene around a Pi. What this illustrates is how easy it is to stage a crime scene for illustrative purposes. It takes minutes to add crime scene markers. In the figure the Pi was marked "F". There is a spare disc on the scene (the SD card marked "B"). In the example a disc adapter is also present ("E"), which suggests that 'non-standard' discs could have been used. Such a 'non-standard disc' (for the Pi) is indeed present in the form of a microSD card ("H"). Had this card not been marked it could easily have been overlooked during evidence collection. Note that a Bluetooth adapter ("D") is also present, raising some interesting questions. However, the point now is the ease with which a scene can be staged and steps to process it justified. After this, physical evidence can be collected, bagged and tagged using, say, ordinary bank (coin) plastic bags and a little bit of imagination (and a lot of emphasis on the chain of evidence). As noted the process takes a few minutes, but provides a strong visual context from which to proceed to the (for us) more important issue of logical acquisition.

To physically acquire the evidence on the disc installed in the Pi it is quite possible to pull the plug and remove the SD card — within a few seconds. It can be demonstrated in class and again illustrates the process very clearly without any unnecessary fiddling with screws and cables.

The SD card as a disc to be used for forensic purposes has another inherent benefit: to protect the contents of the removed disc one simply flips the write-protect switch on the SD card. This obviates the need for a write-protector which usually makes provision for all those 'old' cables — things that are necessary for a real investigation, but at this stage arguably introduce only introduce complexity for a matter where a simple flip of a switch is adequate. (As an aside, if the course includes practical demonstrations for assessment, the question of whether that switched is flipped at the correct moment forms part of the assessment.)

6. THE LOGICAL PROCESS

The standard ('craft', which does not diminish its scientific merit) recipe indicates that the disc should now be imaged. While many forensic tools will do this with the push of a button, this is one of the places where I am in favour of using 'old' tools, because I hope that they bring the message across more clearly. And, for what needs to be done to image the disc nothing beats the simplicity (and simultaneous complexity and danger) of the `dd` command. Some work is required to explain to the students that the discs are mounted on a Unix computer (and, in particular, that the individual partitions of a partitioned disc may be mounted individually) — again without getting lost in the complexity of system design. However, inserting and removing another SD card convinces them that the SD card appears on the system as, say `/dev/sdc`; some serious explanation that the card sits at `/dev/sdc` and its partitions at, say, `/dev/sdc1`, `/dev/sdc2`, and so on. If nothing else, the notion that an SD card may be partitioned is news to most students. The distinction between the device (say `/dev/sdc` and the mount point (say `/mnt/card`) is an important distinction — one that the forensic scientist needs to understand. Hopefully the message is simple enough in this reduced context devoid of other devices (such as optical drives on my 'forensic workstation') that the number of distractions are minimised.

The "forensic workstation" I use is an old, very basic, laptop loaded with a suitable version of Lunix (although other operating systems could have been used). The use of the forensic workstation illustrates how forensic work is typically conducted in a clean environment (rather than on the 'suspect' Pi). It has the additional benefit that its disc and processor are somewhat faster. One may speed some of the subsequent processes up by using a fast SD card — if supported by the forensic workstation.

The standard recipe now proceeds as follows. First one calculates the hashes of the original medium. The MD5 hash is typically calculated by entering

```
md5 /dev/sdc
```

or

```
md5sum /dev/sdc
```

on the forensic workstation. The SHA1 hash is similarly computed:

[2]The basic Raspberry Pi is just a circuit board with connectors. It runs fine in this configuration; however, most users will probably want to build it into a box to prevent short-circuits and other mishaps.

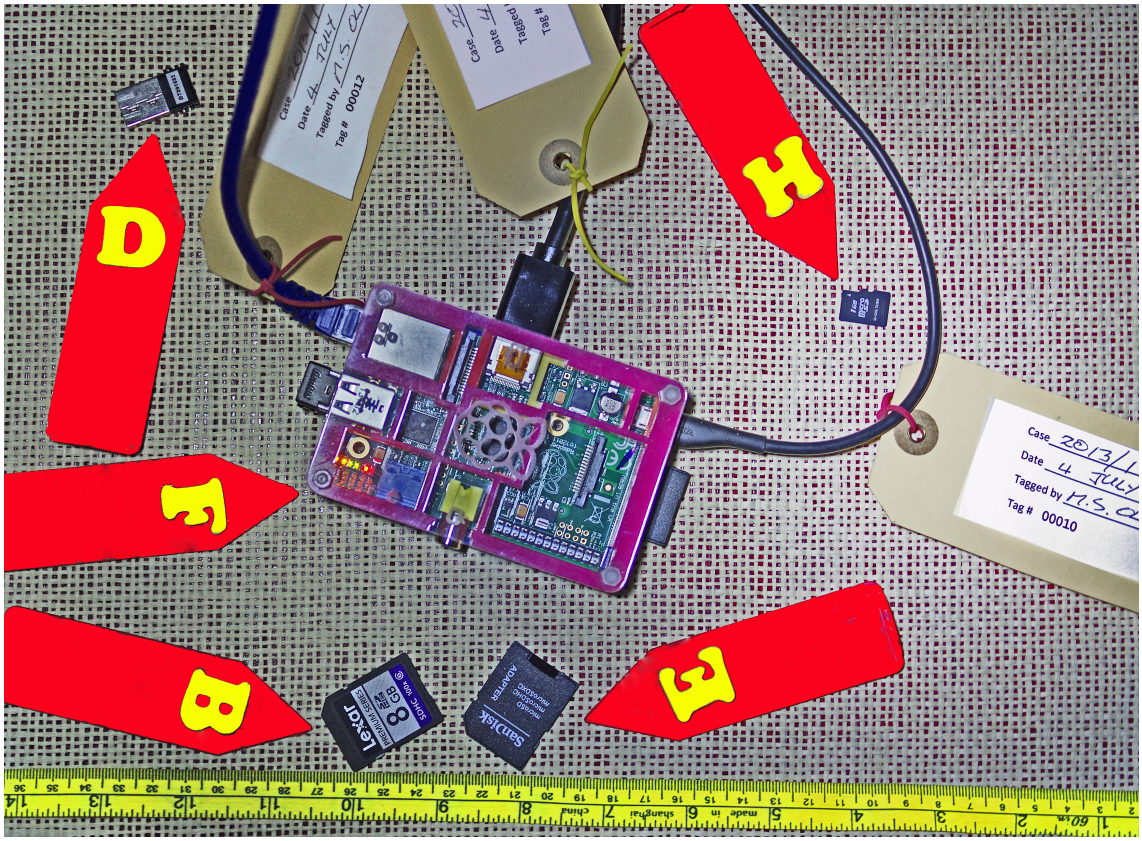

Figure 1: A typical "crime scene" around the Pi

```
shasum /dev/sdc
```

One may then proceed to copy the evidence:

```
dd if=/dev/sdc of=evidence.bin
```

Next, one can verify the correctness of the copy by calculating the hashes of the copy:

```
md5 evidence.bin
shasum evidence.bin
```

If everything was done correctly, the hashes will match and the students (hopefully) be impressed.

It should be noted that the process described above is notoriously slow and therefore hardly ever demonstrated in class. However, the Pi runs quite comfortably from a 2GB SD card. (Such cards are increasingly hard to find, but old cameras is one viable source.) To image such a small card takes perhaps ten or fifteen minutes and can therefore be comfortably demonstrated in class. One just has to be prepared to 'pad' this period of inactivity by continuing the lecture with another relevant aspect of forensics. Note that those few minutes do feel like hours for the lecturer who is anxious that some typo may be the cause of the silence — or that one is busy destroying the hard disc of the workstation; one option is to open a second console and verify — also for the sake of the class — that the (in this example) `evidence.bin` file is indeed growing.

At this stage, the evidence file is supplied to the students. (The exact distribution mechanism has to be determined, because not all students can afford a 2GB download from the Internet.)

Checking the hashes, and using `dd` to dump the file to their own blank SD cards is usually fun — and booting the Pi (or the emulator) from this dump is a revelation to many students; they have read the book, but they are far removed from bits and bytes in their everyday life. Note that they do not have to use 2GB SD cards, which — as noted — are becoming scarce. However, using a bigger card (say an 8GB card) holds another surprise: After dumping the image to the card, the card suddenly 'becomes' a 2GB card. It is a good exercise for them to think about the fate of the remaining 6GB — as well as about ways of recovering that lost space. Why does reformatting the card not work? As noted earlier, there is a wealth of opportunities to learn about the operation of computers in such a course about aspects of computing not related to forensics as such.

The preceding discussion dealt with the acquisition of evidence — in a realistic, but simplified manner. The acquisition process described is part of a "traditional" recipe that arguably has been inherited from the 'craft' days of digital forensics. However, it is scientifically sound in the sense that it preserves stored evidence (but may destroy more transient evidence). In the space of less than an hour the students have witnessed isolation of the computer, logical and physical acquisition, preservation of the integrity of the evidence (using write-blocking over the short term, bagging and tagging of the physical evidence and the calculation and verification of hashes of the logical content) and restoration of the image to a clean disc to convince them that the copying process worked. To make matters more concrete gloves are worn during the entire process (to emphasise) that latent physical evidence may be present and that the digital acquisition should not contaminate such physical evidence). Before anything is touched pictures are taken (before and after cables have been tagged, but not yet unplugged). All of this (gloves, write blocking, hash calculation) is intended to reinforce the importance of preventing contamination (or, in other words, preserving integrity). It also conveys a clear message about doc-

umentation of the process (photographs, tagging, taking notes). And the fact that the Pi boots from a clean (read: newly purchased and opened in class) SD card after the image acquired from the "suspect" SD has been written to it, is a convincing demonstration that the imaging process actually works.

After this rapid demonstration students should have enough insight to understand (but probably not yet answer) questions such as how one would acquire evidence from the cloud. They should also appreciate some questions about possible volatile evidence that may have been lost by pulling the plug. They have a concrete reference model in their minds against which challenges, alternative strategies and *science* can be discussed. With evidence bags, tags and other material all on hand it is also a simple matter to understand the basic notion of forensic readiness.

In addition, the student should at this point have absorbed some of the values and virtues of the craft. The output of the process id an image of a disc that needs to be examined. This is where the primary transition has to occur: that disc image needs to be examined with scientific values and virtues in mind. The quest is no longer for integrity; it is now for truth.

7. EXAMINATION

Examination (or analysis — depending on the terminology one chooses to use) is intended to make sense of forensic evidence. Despite this simple goal, the question whether current practice is indeed scientific is hotly debated — see, for example, [22] and [6]. Elsewhere [16] that a proper scientific foundation is still lacking. However, introducing some esoteric theory at this stage of a course introduces several problems — not least of which is the fact that such a course would leave students with knowledge that would not be recognised by digital forensics practitioners and effectively excludes students from the digital forensics discourse. Hence a slightly more devious approach is necessary. The strategy used was to present the students with 'classical' cases to examine. These cases were again intentionally simple — to convey the basic skill, rather than challenge the students with unnecessary details. However, in each case the students were expected to explain how certain they were that there result was corrects and — more importantly — why they expressed the confidence they did. This is consistent with the NAS report which notes that a fundamental part of forensic competency tests measures the candidate's ability to "assessment of the uncertainty of the results based on scientific understanding of the theoretical principles of the method and practical experience" [7, p.114]. Justification is generally hard. However,in forensic science "the goal is to construct explanations ('theories') of phenomena that are consistent with broad scientific principles, such as the laws of thermodynamics or of natural selection" [7][p.112]. In fact, reliability and known error rates are repeating themes throughout the National Academy report. Generally the students tend to solve cases successfully, but are unable to provide a scientific indication of the correctness of their results (even when they "know" the results are correct. This justifies a paradigm shift because current theories are clearly unable to address this fundamental issue.

The cases are provided in the appendix. Below the logic for including some of them is explored.

Once the case has been made (during the Pi demonstration) that the SD card is just another disc it is no longer necessary to use the full Pi system for every case. In case 2 I used the SD card as a "data disc" and stored pictures of objects on campus with GPS coordinates in the picture's EXIF metadata. A simple algorithm determined which picture each student had to find. In many cases the object in the picture occurred more than once in the given location and I tempted students to identify the exact object based on properties of that specific object. However, most resisted the urge to do so — hopefully based on the lecture about which facts a digital forensics investigator should consider within his or her field of speciality.

In another case the disc was again used as a data disc, but the initial sectors of the disc were overwritten with zeroes. The intention was that the students should carve the recoverable images. The zeroes destroyed the system area of the disc, as well as the header of the first picture. Similar to the other cases the idea was not to create an almost impossible task. The pictures were therefore positioned on the disc such that each file was stored in contiguous sectors, with no gap between them. Students were expected to find the file signatures (or start and end markers) and use a tool such as `dd` to extract the bytes between these markers (both inclusive) and store them in a file with a `.jpg` extension. Unfortunately I did not specify this clearly and some students used automated extraction tools (including well-known forensic tools). Each file included two embedded thumbnails of the main picture. An interesting outcome was that those students who used automated recovery tools were able to recover the three of the four initial images, with — depending on the tool — one or two thumbnails for each of those three images. The students who used the preferred (manual) approach were additionally able to recover one of the thumbnails of the damaged image. This provided an excellent opportunity to discuss the extent a professional digital forensic scientist should rely on the use of tools — and what one is able to testify to in court (using tools and/or a more manual approach).

The differences in their ability to recover images turned out to be an excellent opportunity to raise the point of reliability and error rates again. Was there an error in the reports of those who could not find the final image? Is it possible to talk about *error rates* in this case? How certain were those who found images that those images were actually on the disc? Were they able to quantify their certainty of the presence of the images? Although not initially planned as such, the assignment served as a very useful introduction to error rates as explored by Gladyshev [11], Carrier [4], Cohen [6] and Olivier [17].

Other assignments included recovery of the browsing history of a Pi user (who tried to hide this history by deleting it, as well as deleting the `bash` history) and an assignment that required the exploration (and recovery) of a database that behaved strangely after someone fiddled with the database's metadata.

It should be emphasised that not all practical assignments were equally hard, nor did they count a similar number of marks. The initial assignments were intended to introduce students to the concepts of digital forensics, use the appropriate terminology and get used to using system structures on a low level. The final assignment in this list was based on more recent database forensics work and required a deeper insight into the work. It is worth noting that it was possible to install PostgreSQL on the 2GB card along with the Pi operating system; PostgreSQL provided an ideal vehicle to explore database forensics because of its rich metadata structure.

Note that some areas in digital forensics do regularly deal with error rates. However, they tend to focus on specific application areas, such as author identification of a document [12] or anomaly detection in logs [10]. On the one hand this is quite natural since an examination is inherently coupled to the suspected crime (or nature of a non-criminal incident) that is being investigated. However, this does not provide a good basis to study *scientific* digital forensic examination in general.

Also note that, historically, most examinations consisted of the location of documents of interest, where such documents could then be examined further by other specialists. The term *documents* here is intended to be interpreted broadly to include images, emails, financial statements, textual documents and any

other form of useful information. In many cases documents of interest are contraband, where the mere possession of such documents may be illegal. In other cases the documents may be assembled to form a narrative of events. This may, for example, be emails sent by individuals that show intent to commit a certain action. Another example is log files that may explain why a system crashed — whether it was due to human actions or not. Key factors in the identification and extraction of such documents include authenticity and origin. An email in someone's outbox may, for example, have been planted by a cracker without the knowledge of the owner of that email. Similarly, a criminal may have cracked an individual's computer and use it to store contraband without the user's knowledge. On top of these challenges, the evidence may have been partially destructed — either because it was deleted some time ago, or by a perpetrator who (expecting the investigation) actively attempted to destruct incriminating evidence. The primary digital forensic tools are therefore aimed at solving such questions. These issues are and remain an important part of digital forensic analysis.

However, computer systems are more and more being integrated into various ecosystems that continually synchronise devices with other devices. In many senses this is not new. Operating systems have for a long time provided the ability to synchronise time with external servers. Email between a server and client were continually "synchronised" in the sense that new incoming emails were downloaded and newly composed emails were sent. HTTP provides the option to regularly reload pages from a server in order to synchronise the browser copy with the server copy. With the introduction of automatic updating of operating systems and application software, as well as a plethora of apps that facilitate distributing a user's operations over a variety of platforms, it makes it harder and harder to reach conclusive findings based on mere presence of incriminating evidence found on a digital device. I submit that it becomes necessary to indicate causality to deem digital evidence incriminating (or exculpatory, for that matter). This would necessitate a proper scientific approach.

8. REPORTING

As noted, the focus of this course was on forensic examination, which only left a limited time to teach the genre in which expert witness reports should be written. The guidelines provided by the Medical Protection Society [13] were used. Ideally more time should be spent on such reporting, because reporting skills in many computing courses are not adequately addressed in general. Where a new writing in a new genre is expected one can only hope to achieve a limited amount of success in the one week set aside for it. We can only hope that is, at least, created some awareness of the requirements of forensic reporting.

9. CONCLUSION

This paper presented an approach that transfers the craftsmanship of digital forensics through hands-on experiential learning in a very short time. The experiential learning not only provides a concrete picture in students' minds about the concepts used as the foundation of deeper study, but also provides an opportunity for a deeper discussion of natural 'problems' that arose out of the solution of the problems.

A course was presented using this approach in 2013. The intention was not to present it as a research exercise, but simply as a new approach to teach the relevant material. Anecdotal evidence suggested that many students enjoyed the problem-solving approach. The assignments introduced more opportunities to discuss higher-level (scientific) problems, such as correctness and certainty of findings (or, conversely, the impossibility to reach conclusions about certain aspects of a case) than expected.

Most of the work discussed in this paper was presented over about a four week period at the start of a semester. The "higher-order" problems (such as correctness of results) only became acute during the more advanced later part of the course — for example when digital forensic *science* was discussed — amongst others based on our earlier paper [16].

The database forensics exercise was the only one that required a deeper theoretical basis (provided somewhat later in the course) [15].

It is hoped that the experiential learning will be extended in future to include more of the advanced topics discussed later in the course. In addition, one of the anonymous reviewers suggested the inclusion of engineering as another category of knowledge that should be considered. This is indeed an excellent suggestion that will be included in future work.

10. REFERENCES

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

[2] ASTM. Standard guide for education and training in computer forensics. Standard E2678 - 09, ASTM Subcommittee: E30.12, 2009.

[3] J. Beckett and J. Slay. Digital forensics: Validation and verification in a dynamic work environment. In *Proceedings of the 40th Hawaii International Conference on System Sciences*, pages S12–S22, Hawaii, 2007.

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

[5] B. D. Carrier. Risks of live digital forensic analysis. *Communications of the ACM*, 49(2):56–61, 2006.

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

[7] N. R. C. Committee on Identifying the Needs of the Forensic Sciences Community. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, 2009.

[8] W. P. Dann, S. Cooper, and R. Pausch. *Learning to Program with Alice*. Prentice Hall, 3rd edition, 2011.

[9] J. Darnell. Swgde chair's letter to john paul jones ii, executive secretary, rdt&e iwg in respose to the rdt&e iwg letter to swgde. Available from the DWGDE document collection on swgde.org, Jan. 2012.

[10] B. K. L. Fei, J. H. P. Eloff, M. S. Olivier, and H. S. Venter. The use of self-organising maps for anomalous behaviour detection in a digital investigation. *Forensic Science International*, 162(1–3):33–37, 2006.

[11] P. Gladyshev. *Formalising Event Reconstruction in Digital Investigations*. PhD thesis, University College Dublin, 2004.

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

[13] MPS. A guide to writing expert reports. MPS 1089, Medical Protection Society, Apr. 2013.

[14] M. S. Olivier. *Information Technology Research — A Practical Guide for Computer Science and Informatics*. Van Schaik, Pretoria, South Africa, 3rd edition, 2009.

[15] M. S. Olivier. On metadata context in database forensics. *Digital Investigation*, 5(3–4):115–123, 2009.

[16] M. S. Olivier. On complex crimes and digital forensics. In *Intricate/Sec at ISSA*, pages 139–149, June 2012. In press.

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

[18] M. S. Olivier. Catching the bug: pupils and punched cards in South Africa in the late 1970s. In A. Tatnall and B. Davey, editors, *Reflections on the History of Computers in Education — Early Use of Computers and Teaching about Computing in Schools*, volume 424 of *IFIP Advances in Information and Communication Technology*. Springer, 2014. In press.

[19] M. S. Olivier and S. Gruner. On the scientific maturity of digital forensics research. In G. Peterson and S. Shenoi, editors, *Advances in Digital Forensics IX*, IFIP Advances in Information and Communication Technology — Advances in Digital Forensics, pages 33–49. Springer, 2013.

[20] D. L. Parnas. On the criteria to be used in decomposing systems into modules. *Commun. ACM*, 15(12):1053–1058, Dec. 1972.

[21] J. Sammons. *The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics*. Syngress, 2012.

[22] SWGDE. Digital forensics as a forensic science discipline, version: 1.0. Technical report, Scientific Working Group on Digital Evidence, Feb. 2014.

[23] A. S. Tanenbaum and A. S. Woodhull. *Operating Systems Design and Implementation*. Prentice Hall, 3rd edition, 2006.

APPENDIX

A. CASE 1

Download the file called 'evidence1.zip'. It contains an image of a 2GB SD card and the md5 and sha1 digests of the image. The (disc) image purportedly contains a (visual) image of a suspicious character. Your assignment is to find any picture(s) on the disc image and do the following report (on a single sheet of paper — no cover page or addenda allowed).

1. Say how you *know* that you worked with an exact copy of the original evidence.
2. Provide a passport-sized/thumbnail version of any picture(s) you find (or say that there are none).
3. Explain how you were able to extract the picture(s), if any.
4. Say why you think your answer for point (3) is forensically justifiable.
5. Could you find any digital evidence that the SD card was used on a Windows computer?

Note that regarding point (5) it is dangerous to say that some hypothesis *is* true; it is much better to say that certain artefacts found are consistent/inconsistent with the card being used on a Windows computer. Don't answer questions that have (not yet) been asked. For example, information about the camera used to take the picture will become relevant in Assignment 3. Assignment 1 counts 5 marks out of the 100 practical marks. It's due two weeks after the evidence has been made available.

B. CASE 2

The disc (image) supplied in the files section is believed to contain some images of evidentiary value. However, when the card is placed in a Windows computer (write-protected, of course) Windows says the card needs to be formatted, which is probably not a good idea. They (law enforcement — not Windows) approach you to help. Again your answer should fit on one side of an A4 sheet of paper; marking stops at the end that first page... Please use a 12pt font.

1. Provide (small) copies of any pictures on the disc, or say that there are none. (2+2)
2. Say how you managed to retrieve the pictures or establish that there are none. (2+2)
3. The judge asks how certain you are of your results. If you were working in DNA forensics, you would probably answer something like 99.973%. How do you answer in this case? Look at Casey's certainty scale (but it will probably not help). The judge says: "Please answer the question." What do you say? Begin your answer with "My Lord, ..." (2)

This mini-prac counts 10 marks towards the 100 marks set aside for practical assignments. Again the evidence file will expand to a 2GB file when unzipped.

C. CASE 3 — IN THE FOOTSTEPS OF MR S TALKER

For this case files have been obtained from a computer's hard disc, so there is no need for you to recover them from an image. (See the evidence section of this site.) There is reason to believe that each file identifies some artifact or other object. Your task is to identify the object. To lighten your burden you only have to identify one object. Use the final digit of your student number to determine which file is assigned to you. If your number is XXXXXXX3 it means file3.??? contains the evidence you should process.

1. We will use the following unorthodox approach to "prove" that you managed to identify the object: Provide a picture of you pointing towards the object. Composition and other aesthetic photographic qualities do not count. A wide shot that shows something of the location/general area plus a closeup of just your pointing finger and whatever it points to will be ideal. (1)
2. Explain how you went about identifying the object. Use about 3 to 8 sentences. (3)
3. (a) In forensics one should distinguish between class and individual characteristics. All rounds fired by, say, a Luger (any Luger), may share some identification marks (known as striation marks or striations). However, for various reasons, John X Doe's Luger may exhibit some particular striations; they are almost like a fingerprint: if we find one of those rounds we can say it was fired by John X Doe's Luger. Question: Is the artefact/object you are pointing at in the picture in (1) a member of the class, or the individual object identified by the evidence. (1)
   (b) How certain are you of your answer in (3a)? Use Casey's certainty scale or any other approach you can justify as a digital forensic scientist. Your answer will be a single word, phrase or number (and identify the scale). (1)
   (c) Justify your answer at (3b). Use about 5–10 sentences. (3)
   (d) What other digital evidence would, if you could obtain it, enable to increase your confidence? Why will it help?

Although the marks for the subquestions add up to more than 10, this case enables you to earn up to 10 marks for the practical/assignment portion of this course. Keep in mind that this case requires some fieldwork, and your ability to complete it may depend on the weather. Plan ahead. Rain and snow will not be viable excuses. Your entire answer should again fit on one side of an A4 sheet of paper. Use a 12pt font.

D. THE BIG CASE 4 — DID JANE VISIT THOSE HORRIBLE SITES?

The file evidence4.zip is available in the evidence section. Based on a proper chain of custody (like earlier assignments) we know that the evidence has not been tampered with between the point of acquisition/seizure and reaching your hands. The image is a copy of a hard disc (a 2GB SD "hard disc") seized with Jane Y Doe's computer. You are tasked with establishing whether the disc indicates that the computer from which it was removed, had been used to access www.cs.up.ac.za and/or google.*. Is it possible that the user "accidentally" accessed those sites (if those sites were accessed), or is there any reason to believe the visits were intentional?

Answer your question in the form of a proper forensics / expert witness report. Issues such as tool validity, protocols followed and confidence levels are implicitly part of such a report.

You are limited to three pages (12pt font). However, you are allowed to add appendices in this case; such appendices will, however, not necessarily be marked.

This case counts a whopping 25 marks (25%) of your assignment mark.

E. CASE 5

Case 5 was based on work a colleague presented to the class and is therefore not included in this paper.

F. CASE 6 — THE FINAL ASSIGNMENT

For case 6 you have to examine the hard disc drive seized from a certain Ms Kim Mér. It contains a database with 500 000 entries in each of a number of tables of what look like credit card details. [For the sake of compactness it only really contains 5, but you should provide a solution that would have worked if there really were 500 000 records.]

The analysts that have already looked at it observed these entries in a PostgreSQL database. You are given an exact copy of the seized disc (which you may verify as usual by hash comparison). The case investigators think they will have a much better case if they can show that the disc contains "complete" credit card details, rather than fragments that may be (or may not be) in these records. Can you help them? If you can, it may be worth up to 25 marks (out of the 100 practical marks) for your semester mark. Note that the earlier examiners have already seen that some columns seem to indicate credit card data (such as a column named csv). Don't even speculate about this. Also note that since you are getting an "original" image of the disc there is no need to worry about what the other examiners may (or may not) have done with the evidence. See what you can do, do it, provide the results and state how "certain" you are — in this case: how much weight should the court attach to your evidence. Formulate your answer as a proper forensic report using up to three A4 12pt pages.