Advancing Digital Forensics

Katrin Franke, Erik Hjelmås, Stephen D. Wolthusen

Advancing Digital Forensics

Katrin Franke[1], **Erik Hjelmås**[1], **and Stephen D. Wolthusen**[1,2]

[1] Norwegian Information Security Laboratory, Department of Computer Science, Gjøvik University College, Norway, {katrin.franke | erikh}@hig.no
[2] Information Security Group, Department of Mathematics, Royal Holloway, University of London, United Kingdom, stephen.wolthusen@rhul.ac.uk

Abstract: The diversity of computing and communication systems used as well as the sheer volume of data processed in all aspects of personal, government, and commercial activities poses considerable challenges to law enforcement and particularly compliance officers. While commercial tools exist for a number of common problems, this is, however, not always sufficient in many more complex cases. Moreover, investigators only familiar with such tools may not be aware of limits in scope and accuracy, potentially resulting in missing evidence or placing unwarranted confidence in it. Moreover, not only is it critical to have an in-depth understanding of the underlying operating principles of the systems that are analyzed, there will also at times be a need to go beyond capabilities of existing tool sets, the enabling knowledge, concepts, and analytical skills for which we argue is currently not offered in a concise higher education context but rather tends to be acquired in an ad-hoc manner.

We therefore propose elements of a curriculum for the M.Sc. and particularly the Ph.D. level which provide the necessary rigorous theoretical foundations and perspectives in mathematics, computer science, and engineering combined with a background in forensic sciences which enable both a sound appreciation of existing techniques and the development of new forensic evidence collection and analysis methods. We argue that these abilities are crucial in developing a more rigorous discipline of digital forensics which will both be able to address new challenges posed by evolving information systems and also to satisfy the stringency expected from it given its increasing importance in a broad range of application areas.

Keywords: Digital Forensics, Curriculum Development

1. Introduction

Digital forensics (also referred to at times as computer forensics) encompasses approaches and techniques for gathering and analyzing traces of human and computer-generated activity in such a way that it is suitable in a court of law. The objective of digital forensics is hence to perform a structured investigation into past and ongoing occurrences of data processing and transmission whilst maintaining a documented chain of evidence, which can be reproduced unambiguously and validated by competent third parties. Challenges to such investigation, in addition to legal issues which are beyond the scope of this paper, are practitioner-driven approach currently pursued [1].

A number of programs on digital and computer forensics exist both at the B.Sc. and M.Sc. levels along with a large number of modules integrated in information security and general computer science. The former include offerings by the Universities of Bedfordshire, Bradford, Middlesex, Strathclyde, Teesside and Westminster, UK as well as the John Jay College of Criminal Justice at the City University of New York, Sam Houston State University and the University of Central Florida in the U.S., as well as concentration areas embedded in computer science (e.g. [2]) or forensic sciences in case of George Washington University, Marshall, Purdue, and Stevenson University, the Universities of New Haven and Rhode Island in the U.S. and the University of East London in the UK and the University of Western Sydney in Australia; details (albeit with a U.S. focus) can be found in a recent survey by Taylor *et al.* [3] as well as earlier work by Yasinsac *et al.* [4] and Gottschalk *et al.* [5] with examples of undergraduate programs being described e.g. by Bem and Huebner [6].

Specific offerings at the Ph.D. level are, however, more limited, and although advanced research in the area is not limited to the above-mentioned institutions and a number of specialized publication outlets such as the IFIP 11.9 conferences and SADFE (Systematic Approaches to Digital Forensic Engineering), DFRW (Digital Forensics Research), and Computational Forensics (IWCF) workshops along with publications such as the IEEE Transactions on Information Forensics and Security, the International Journal of Digital Evidence and related research in a number of other outlets, it appears that it is often the application of research to forensics that is acting as the determinant rather than the subject matter itself. However, it is instructive to note that in a recent proposal of a Ph.D. curriculum for digital forensics, Cohen and Johnson stated expert knowledge in the application area as the teleology for higher education at this level rather than research itself [7].

This paper aims to raise three questions with regard to research-oriented higher education in digital forensics, which we think should be the rule rather than the exception at the Ph.D. and M.Sc. levels as opposed to the more vocationally oriented undergraduate and certificate-based offerings. First, we consider it necessary to delineate the scope of digital forensics; here, we concentrate on technical aspects as is suitable for research. Secondly, based on the preceding analysis we identify the topics and areas which we consider unique to digital forensics or at least sufficiently specialized to warrant inclusion in a forensics research curriculum according to the preceding criteria. Finally, we argue that one of higher education's and particularly research's roles in digital forensics should be on the enabling or the development of forensics-friendly mechanisms and systems as this holds the promise of considerable medium- and long-term benefits. The remainder of the paper therefore addresses the issue of delineation in section 2 followed by a discussion of subjects and topics we consider sufficiently unique to digital forensics in section 3 followed by our arguments for research on systematically enabling forensics in section 4 before a brief summary and conclusions in section 5.

2. Delineating Forensics

While the impetus for digital forensics is originating with legal requirements, we focus here primarily on areas amenable to scientific and mathematical inquiry as this is more likely to be beneficial for the types of investigations and research found in M.Sc. and Ph.D. work. Digital forensics does have an intersection with what may be called conventional forensics in that the underlying physics of information processing devices rather than the logical abstractions formed by computer science and engineering may determine whether evidence can be collected and, if so, how reliable the data captured is to be considered. Beyond these foundations, however, engineering issues will dominate the accessibility for most types of storage (e.g. magnetic, optical, solid-state, and in some cases also considering volatile storage sub-types).

As noted in [7], it is hardly possible to provide students of digital forensics with a solid grounding in all such foundational aspects, and it will be incumbent on students wishing to pursue research in this area to acquire the specialized knowledge and skills from physics and electrical or computer engineering required. When using the abstraction provided as a metric, the issue of extracting, classifying, and visualizing the patterns resulting from the data lies at the other end of the spectrum from device physics. One is, however, confronted with a similarly large area of research as in the case of the physical sciences, and most research involving these areas is more likely to be applied or derivative in nature, although there is clearly considerable room for using domain-specific knowledge

to enhance general techniques and approaches. One key characteristic of both individual data items and particularly of any hypotheses and chains of evidence is verifiability, which should become more significant as the field matures from relying on individual expert opinion to objective standards. This requires a rigor and change of emphasis compared to the typical approaches found in information security where the existence of a certain false-positive rate is accepted based on the assumption that analysts will discard such indicators in subsequent steps (e.g. in case of anomaly-based pattern matching and classification used in intrusion detection). Given both the potential adverse consequences of such a false positive match for an individual falsely accused and the likely impact that the detection of a false positive has on the credibility of the forensic mechanism and the expert giving evidence, this clearly provides an impetus for research into verifiable approaches or, if that is not feasible in a given area, ones for which error characteristics can be determined rigorously. This implies not only a sound understanding of statistics and probability as well as of formal models of causality in applying forensic techniques as discussed in section 3, but also imposes constraints on the gathering and particularly on the processing of evidence in such a way that any probabilistic aspects and errors are well understood. This specific aspect of digital forensics is made particularly relevant by the volumes of data which may need to be subjected to analysis and reconstruction both in compliance processes and in discovery or court cases.

Moreover, the combination of potentially large volumes of data to be considered and the need to present the resulting evidence, hypotheses, and chains of reasoning to non-experts in such a way that challenges can still be met rigorously also presents a number of challenges, beginning with requiring an understanding of the human perceptual system and particularly of the limitations of intuitive reasoning that may bias the perception of evidence by non-experts. In addition — as will also be discussed in section 3 — the pervasive character of information processing systems in all aspects of life also has implications for the potential sources of data and evidence. While traditional information security is often limited to easily accessible and commonly used environments, forensics must consider a large number of unconventional devices such as embedded systems as data sources. Many such embedded systems will be quite limited in their scope and ability, potentially requiring the fusion of a number of data sources to obtain the evidence or accuracy desired. This, however, requires that the individual sources and evidential data be interlinked, which will often be possible only in conjunction with explicit models of an overall system or the external (physical) environment, e.g. in case of vehicular systems. Such models are not necessarily part of either the curriculum or research agenda in information security or, beyond this, computer science and applied mathematics. and hence require a solid grounding in the physical sciences beyond the immediate needs of gathering the evidence from digital systems themselves mentioned above. In delineating digital forensics one must also consider the more general case of *computational forensics* or forensic information technology [8], which is more general in nature and employs

computational approaches in support of forensic investigations such as hypothesis generation and validation. However, given that in this case the bounds of the field are determined more by the application area rather than inherent in the field itself, we consider only the immediate intersection with digital forensics proper as relevant for curriculum development.

3. Unique Subjects in Forensics and Limitations

The preceding section has raised the issues of which subject areas are to be part of a curriculum for graduate and postgraduate studies in digital forensics, which are not studied in sufficient depth in computer science and (applied) mathematics curricula [9, 10]. In the following, we therefore describe both *supporting* curricular elements and those of more immediate significance to applications and research, driven in part by the results from the CISSE 2008 report by Nance et al. [1]. As noted by several authors including [7], the most universal supporting modules are indubitably statistics and probability theory, which are also a key element in general forensic science [11]. However, given the requirements outlined above, both practitioners and researchers in digital forensics will typically require a more solid grounding in the creation of models of causality and the limitations of inference models based on incomplete and uncertain information [12]. Further, more generally applicable courses and modules will typically encompass the areas of pattern classification, recognition, and matching as well as machine learning and visualization. All of these, together with the often highly optimized algorithms and data structures used will, however, require considerable background in these areas of computer science. Beyond this, however, digital forensics proper requires familiarity with several aspects of information systems, which are not commonly taught in computer science and engineering or even information security programs. Even in case of conventional computer systems and network systems, the need to cover broad concepts typically results in only an abstract coverage of the principles of operation of digital systems, operating systems, and the interaction of components ranging from storage to peripheral and network subsystems. Moreover, the same desire for abstraction often results in oversimplified models that, while applicable at some point, have long since been superseded; students of digital forensics must, however, typically be familiar with specific implementation characteristics and thus have at least a conceptual framework for studying these rapidly changing models and systems.

Moreover, as noted in section 2, this also extends to information processing and communication systems found in embedded systems that are likely to gain increasing importance as sources of evidence, which not only present different operating environments and constraints (e.g. real-time as well as computational and memory) but also a diversity of capabilities ranging from radio-frequency identification tags via sensors to the rich capabilities of smart phones and

vehicular systems. Moreover, such environments also interact with sensors and the physical environment, requiring further consideration. Beyond these topics, however, a review of research activities in digital forensics does not allow the conclusive specification of a set syllabus for courses or even an entire degree program; while areas such as host and network forensics including malicious software mechanisms to be used both for exfiltration of forensic data and as attack mechanisms to be discovered are uncontroversial, neither the depth of coverage nor the systems to be covered are defined clearly. Similarly, while a background in cryptology and particularly cryptanalysis along with ancillary areas such as steganography and steganalysis are highly desirable, they will necessarily be limited in scope. We therefore find it inevitable to structure modules and curricula in such a way that foundational courses described above together with surveys of these topics are augmented by directed individual studies in support of students' research activities.

Finally, another area specific to digital forensics — although similar issues also arise in a more general information security context — is clearly the legal domain. However, this is problematic particularly for highly international programs in that despite recent efforts at harmonization e.g. within the European Union and the EEA, legal systems as well as procedures for gathering, processing, and presenting evidence are substantially different. As with the areas discussed above, it may therefore be more efficient to provide an overview of the legal frameworks in multiple countries, leaving specialization to subsequent individual study rather than focusing exclusively on a single one; this rationale is particularly supported by the observation that not only is research in digital forensics an inherently international endeavor, but also that even practitioners are more than likely to be confronted with multinational environments whether in criminal proceedings or particularly in compliance or discovery procedures.

4. Enabling Forensics

Much as research on intrusion detection and prevention suffers from the limited scope, volume, and trustworthiness of sensor data, digital forensics is often confronted with sources of evidential data that are barely fit for purpose, incomplete, and of questionable reliability [13]. One area of research that holds considerable promise is therefore the development of new or retrofitted systems providing reliable records suitable for use as evidence, and in many cases, these requirements are paralleled by those for auditing found in other areas. However, while auditing is mostly concerned with linking events to authenticated entities for attribution, this is insufficient for forensics purposes as it is frequently not an individual event but rather a sequence of events, potentially originating with multiple event sources, not all of which are attributable or at least have differentiated degrees of confidence in attribution. This not only requires research on the derivation of appropriate metrics and their efficient incorporation into

evidential data but also further consideration on preservation mechanisms for such data in the presence of tampering and compromise on one hand and, moreover, approaches to linking individual items into a more comprehensive and coherent whole, often also based on a distributed system lacking a common time base. Thus, in addition to the subjects noted in section 3, research in this area will typically require familiarity with the relevant abstractions from mathematics and computer science such as for distributed algorithms and cryptographic primitives, e.g. for secure multiparty computation, frequently already found in more general information security curricula, but applied to the rather different models of correctness, trust, and reliability than that more commonly found in theoretical computer science and particularly in cryptography.

5. Conclusions

Digital forensics is enjoying considerable popularity as a subject of studies particularly at the undergraduate level owing in no small part to positive employment prospects together with positive connotation derived from media exposure. At the graduate and postgraduate levels, however, the emphasis of degree programs tends to still favor the application of forensics over the generation of new knowledge, also reflecting that the research agenda is still largely driven by practitioners. While the breadth of the subject area is clearly daunting, we strongly suggest that digital forensics professionals will, as in other fields of inquiry, not just require the ability to apply knowledge but to critically challenge concepts, approaches, and also evidence while at the same time being able to obtain, derive, and analyze digital forensic evidence in novel and cogent ways. We consider conducting research (either under guidance in case of M.Sc. dissertations, or largely independently in case of doctoral studies) to be both a proven pathway as well as a necessity given the — steadily growing as technology and its applications move on — number of unsolved research problems. In this paper we have therefore outlined what may be considered a core area of digital forensics, its interrelationship with information security and the broader context of applying computational methods to forensic science and how these can, at the graduate and postgraduate levels be best supported using both course modules and alternative forms of study with the former of necessity being devoted mainly to theoretical underpinnings from mathematics as well as computer science and engineering. While some areas overlap with other specializations in the field such as general information security, there is still a considerable area of specialization, which must be covered. Moreover, we also assume that the focus of a graduate or postgraduate program in the field will inherently focus on the understanding of existing and development of new forensic techniques and approaches, requiring familiarization with tools and their application through other means. Given the international nature of the programs considered here, moreover, only limited attention is paid to legal considerations that are specific to a particular country or legal tradition; we readily acknowledge that this trade-off clearly requires further

specialization in most cases. Ongoing developments and further research will concentrate on the identification of research-driven curriculum development and the trade-offs associated with offering a broader spectrum of specialized elective modules compared to combining a compulsory core area with guided individual specialization at both the M.Sc. and Ph.D. levels.

Acknowledgments: The author would like to thank J. Austen and A. Tomlinson for valuable discussions and comments.

References

Reference Style (Numerical Order) as Cited in Sections 1 and 2 as [Number]

1. Nance, K., Hay, B., Bishop, M.: Digital Forensics: Defining a Research Agenda. In: Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS '09), Waikoloa, HI, USA, IEEE Press (January 2009) 1–6 (also published as a CISSE report)
2. Figg, W., Zhou, Z.: A Computer Forensics Minor Curriculum Proposal. Journal of Computing Sciences in Colleges 22(4) (April 2007) 32–38
3. Taylor, C., Endicott-Popovsky, B., Phillips, A.: Forensics Education: Assessment and Measures of Excellence. In: Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE 2007), Seattle, WA, USA, IEEE Press (April 2007) 155–165
4. Yasinsac, A., Erbacher, R.F., Marks, D.G., Pollitt, M.G.: Computer Forensics Education. IEEE Security & Privacy 1(4) (August 2003) 15–23
5. Gottschalk, L., Liu, J., Dathan, B., Fitzgerald, S., Stein, M.: Computer Forensics Programs in Higher Education: A Preliminary Study. In: Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, St. Louis, MO, USA, ACM Press (February 2005) 147–151
6. Bem, D., Huebner, E.: Computer Forensics Workshop for Undergraduate Students. In: Proceedings of the Tenth Conference on Australasian Computing Education, Wollongong, New South Wales, Australia, Australian Computer Society (January 2008) 29–33
7. Cohen, F.B., Johnson, T.A.: A Ph.D. Curriculum for Digital Forensics. In: Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS '09), Waikoloa, HI, USA, IEEE Press (January 2009) 1–8
8. Srihari, S.N., Franke, K.: Computational Forensics: An Overview. In Srihari, S.N., Franke, K., eds.: Proceedings of Computational Forensics, Second International Workshop (IWCF 2008). Volume 5158 of Lecture Notes in Computer Science. Washington D.C., USA, Springer-Verlag (August 2008) 1–10
9. Foster, K.R., Huber, P.W.: Judging Science: Scientific Knowledge and the Federal Courts. MIT Press, Cambridge, MA, USA (1997)
10. Aitken, C.G.G., Taroni, F.: Statistics and the Evaluation of Evidence for Forensic Scientists. 2nd edn. John Wiley & Sons, New York, NY, USA (2004)
11. Taroni, F., Aitken, C., Garbolino, P., Biedermann, A.: Bayesian Networks and Probabilistic Inference in Forensic Science. John Wiley & Sons, New York, NY, USA (2006)
12. Pearl, J.: Causality: Models, Reasoning, and Inference. 2nd edn. Cambridge University Press, Cambridge, UK (2009)
13. Deane, W.: System Event Monitoring as a Security Control. Master's thesis, Royal Holloway, University of London, Egham, Surrey, UK (September 2008)