Chapter 9

Conclusions and future work

. . . Chance has put in our way a most singular and whimsical problem, and its solution is its own reward. . . .

Arthur Conan Doyle

This dissertation investigated the theory and practice of event reconstruction in digital investigations. The main outcome of this work is a formalisation of event reconstruction in terms of state machine model of computation. This formalisation has been validated through the development of an event reconstruction algorithm and using it to perform sample event reconstructions.

This chapter summarises the main points from this study and concludes the work.

9.1 Problem

Digital evidence is commonly encountered in many types of criminal and civil investigations. It has been argued, however (see for example [76], and [4]), that currently widespread *ad-hoc* analysis of digital evidence is inappropriate from forensic point of view, because it is error-prone and because its findings

are hard to explain and defend in court. More rigorous methods of analysis have been called for.

To answer this call, this dissertation explored the problem of event reconstruction in digital investigations, whose aim is to determine the sequence of events that happened in a given computer system during the incident. The need for such reconstructions arises in a variety of cases ranging from investigations of technically advanced network intrusions [38] to seemingly straightforward blackmailing cases [9].

To clarify the problem, a study of digital forensics and related disciplines has been performed. The study has shown that event reconstruction is expected to be

- efficient,
- effective (i.e. reliable and precise), and
- based on a scientifically valid methodology.

The study has also shown that current practices of event reconstruction are essentially manual, and their reasoning is based on common sense and investigator's experience rather than on any scientific theory.

The idea of this project was that, to improve this situation, event reconstruction should be defined as a computer science problem and solved using methods of computer science. More specifically, the objectives of this project were

- to formalise event reconstruction in a general setting, that is, assuming nothing specific about the digital system under investigation or about the purpose of event reconstruction, and
- to show that this formalisation can be used to describe and automate selected examples of digital forensic analysis.

9.2 Solution

To formalise the event reconstruction problem, this project used the state machine model of computation. This model of computation is convenient from the forensic point of view, because

- the operation of state machine closely resembles the operation of actual computing devices, which is appealing to the fact finder;
- state machine models are widely used in practice to specify and verify computing systems.

The idea behind the solution was to model the system under investigation as a finite state machine, and to define event reconstruction as the process of finding all possible computations of the machine that agree with the evidence of the incident.

To make this idea completely formal, the evidence about the incident had to be formalised. For this purpose, the notion of *evidential statement* has been developed. Evidential statement represents the evidence as a system of *observations* about the properties and change of the system state during the incident. By doing so, it restricts possible computations of the finite state machine. The formalisation of an incident, therefore, consists of two parts: a finite state machine model of the system under investigation, and an evidential statement that represents the evidence.

A precise definition of the event reconstruction problem was then given in Section 6.2.5. Informally speaking, event reconstruction is defined as the process of (1) finding all computations of the state machine that agree with the evidential statement, and (2) identifying how parts of these computations match individual observations within evidential statement.

To show that the developed formalisation can be used to automate event reconstruction, an event reconstruction algorithm has been designed and implemented. To fulfil the second aim of this project, the implementation of

the event reconstruction algorithm was then used to perform two examples of forensic event reconstruction. The results of this practical application, as well as other results of this project are summarised in the next section.

9.3 Lessons of the project

This research has produced a number of original and innovative ideas and results, as well as encountered some problems. They are summarised below.

9.3.1 Achievements

The key achievements of this research are as follows.

- For the first time, a precise mathematical definition of event reconstruction in digital investigations has been given. Apart from providing a basis for automation of event reconstruction, it contributes to the development of digital forensics theory as a discipline.

- A generic event reconstruction algorithm has been designed and implemented. Unlike many digital forensic tools, it has solid theoretical foundation, which can be used to defend its admissibility in legal proceedings.

- As example applications of the developed formalisation, two instances of digital forensic analysis have been formalised and automated. They demonstrate why and how formal event reconstruction can be used in practical investigations.

- It has been demonstrated on a practical example, that formal approach to event reconstruction can discover weak points and hidden assumptions in informal event reconstructions.

- The formal approach to event reconstruction has been compared with existing semi-formal event reconstruction techniques. It has been shown

that formal event reconstruction is likely to have lower error rate and provide more complete event reconstruction that the existing techniques.

- A study of legal and practical aspects of forensic event reconstruction identified the deficiency of existing semi-formal techniques in the context of digital investigations.

On the personal side, this was a very interesting and delightful, albeit very long and sometimes difficult project. The author acquired knowledge and skills in many areas, which include (but are not limited to) the following.

- Legal concepts surrounding the use of digital evidence in litigation.
- The investigative process as well as the techniques for collection, examination, and analysis techniques of digital evidence.
- Existing semi-formal techniques for forensic event reconstruction.
- Formal specification of computing systems.

9.3.2 Problems encountered

In addition to many interesting and innovative results, a number of problems associated with the formal approach to event reconstruction have been discovered. They are summarised below.

- *Computational complexity of the event reconstruction algorithm.* The running time of the event reconstruction algorithm has been estimated in Chapter 7. An upper bound on the running time of the algorithm has been derived analytically. It turned out to be exponential with respect to the parameters of the evidential statement, but polynomial with respect to the parameters of the finite state machine. The exponential complexity suggests that the algorithm may not be able to handle large evidential statements with many observation sequences.

Although this is a serious limitation on the applicability of the developed algorithm, it has been demonstrated in Section 8.3.2 that even small and incomplete models can be useful in practical investigations.

- *Complexity of real world systems.* The formalisation of event reconstruction developed in this dissertation relies on state space exploration to perform event reconstruction. The state machines considered in this dissertation are very simple. For example, the ACME investigation described in Section 8.3.1 required a finite state machine with only 25 states and 75 possible transitions. The majority of investigations are likely to encounter systems, whose finite state machine models are much more complex. The experience of formal methods suggests that for many real-world systems the brute-force exploration of their exact finite state machine models is infeasible due to limitations of modern computers.

  The experience of model checking is also instructive in another respect. At the beginning of model checking era, the complexity of exact finite state machine models of almost all systems was beyond capabilities of model checking programs. However, as the model checking algorithms improved, and various model reduction techniques were developed, model checking of industrial scale systems became possible. The same kind of success may be possible with formal event reconstruction in digital investigations.

- *Formality of the approach.* Despite providing more effective event reconstruction, the formal approach developed in this dissertation requires considerable effort for formalising the incident. This, together with the need to learn formal notation is likely to be a deterrent for its use in practical applications. Nevertheless, the developed formal approach will probably find its application in cases, such as [38] where the success of the legal action depends on the comprehensiveness and reliability of event reconstruction.

9.4 Future work

The work performed in this project provides basis for future research in several areas. At least four such areas can be identified. These areas include:

- extending formalisation of event reconstruction,
- developing more efficient event reconstruction algorithm,
- investigating new ways of constructing system models, and
- developing practical applications of the results of this work.

The following sections discuss each of these areas in more detail.

9.4.1 Extending formalisation of event reconstruction

The formalisation of event reconstruction developed in this dissertation can be extended in several ways. One possible extension is to provide support for uncertain reasoning. Uncertainty is deeply ingrained in forensics. The blackmail example from Chapter 8 is a vivid demonstration of this fact. It shows that many assumptions in actual investigations are being made on the basis of the investigator's experience of what is and is not *probable* in the specific circumstances of the case. Although the complete formalisation of this uncertain knowledge is problematic (consider, for example, measuring uncertainty of an eyewitness statement), there are statistical measures that can and should be incorporated into event reconstruction process. In addition, there is a body of applied mathematics [5] that has been specifically developed for calculating the impact of known statistical properties of the world — such as probability of finding bloodstains on clothes in the general population — on the probability of investigative hypotheses.

Formalisation of event reconstruction given in this dissertation ignores the issue of uncertainty. Although, as explained in Chapter 6, this assumption has some merit, the formalisation of event reconstruction can be enhanced by

adding explicit measures of uncertainty to formalisation of event reconstruction problem. The possibility and the impact of such additions on the process of event reconstruction needs to be investigated.

Another possible extension is to enrich the expressiveness of evidential statements. For example, the introduction of variables would allow the analyst to specify statements such as "some (unknown) phenomenon X has been observed twice":

$$os = (\,(x, 1, 1), (C_T, 0, \textit{infinitum}), (x, 1, 1)\,)$$

Note, however, that any such extension may increase complexity of event reconstruction. The analysis of the impact of such extensions on the semantics and complexity of event reconstruction problem is a topic for separate research project.

9.4.2 Developing more efficient event reconstruction algorithm

The event reconstruction algorithm presented in Chapter 7 is quite inefficient despite its ability to handle examples described in Chapter 8. Development of a more efficient algorithm is an important area for future research.

One possible direction of research is to investigate more efficient representations of computation sets. It can be shown that, if a single step of backtracing was *adding* a fixed amount of elements to the representation of its input computation set (rather than *multiplying* the size of the representation a fixed number of times), the time required for SolveFOS algorithm would become polynomial in the number of backtracing steps. Symbolic representation techniques used in model checking [28] represent one possibility that should be investigated in this respect.

In addition, the applicability of model reduction techniques, such as data abstraction and partial order reduction described in Chapter 5, should be

researched in the context of digital investigations.

9.4.3 Investigating new ways of constructing system models

As noted in Chapter 5, the initial formalisation of the problem is a critical and laborious part of any formal analysis. Simplification or automation of this process is an important direction for future research.

Some of the existing approaches to the development of finite state machine models of systems have been described in Chapter 5. However, there may be other approaches, which are more suited for the purposes of digital investigations. One interesting question is whether the model of the system under investigation can be derived directly from observations of the system behaviour without consulting the source code or user manuals. The applicability of machine learning techniques should be investigated in this respect.

9.4.4 Developing practical applications

The results of this work can be applied to practical investigations in several ways. Some of these ways are discussed below.

The most straightforward application is the development of a general-purpose event reconstruction tool similar to model checkers used for systems verification. When using such a tool, the human investigator would provide a formal description of evidence and a model of the system under investigation. The tool would calculate and visualise possible incident scenarios consistent with the given formal description. Such a tool would be welcome in investigations such as [38] where success of the legal action depends on the comprehensiveness and reliability of event reconstruction.

Another possible application of the developed formalisation of event reconstruction is proving correctness of forensic analysis tools. As discussed in Chapter 3, some of the tools used in digital forensic analysis can be viewed as performing specialised form of event reconstruction. For example, the recov-

ery of deleted files can be viewed as reconstruction of events in the file system back to the moment when the given file was deleted. Such specialised event reconstruction can be defined (with respect to the file system model) by the evidential statement

$$es_x = (a_0, \ldots, a_n, ((C_T, 0, \textit{infinitum}), (x, 1, 0)))$$

where $(x, 1, 0)$ formalises the knowledge of the final state of the system, and observation sequences $a_0, \ldots, a_n$ formalise assumptions made by the designers of the analysis tool. To prove correctness of the tool one should prove that for all possible inputs x, the meaning SPR_{es_x} of the evidential statement es_x is linked to the tool's output out_x according to some well defined interpretation relation $\overset{R}{\sim}$:

$$\text{for all possible } x, \quad SPR_{es_x} \overset{R}{\sim} out_x$$

The interpretation relation $\overset{R}{\sim}$ can be that out_x is equal to some part of SPR_{es_x}, or that it can be derived from SPR_{es_x} by some function.

In addition to proving correctness of forensic *analysis* tools, the formalisation of event reconstruction can also be used to check correctness of evidence *collection*. To achieve this, the process of the evidence collection itself can be reconstructed. The results of such reconstruction can be tested to see if they contains possible scenarios that involve modification of the collected evidence.

9.5 Summary

Overall, this project was a success. This dissertation extended the theory of digital forensic science by formalising and solving event reconstruction problem using methods of computer science. The formalisation of event reconstruction has been validated through the development of an event reconstruction algorithm and using it to perform sample event reconstructions. A number of novel results as well as problems associated with the developed formalisation have

been discovered.

Several possible directions for future research have been proposed. They include: extending formalisation of event reconstruction, developing more efficient event reconstruction algorithm, investigating new ways of constructing system models, and developing practical applications of the results of this work.

Yet, every doctoral dissertation should have an end. So I rest my case!