Application of Backward Chaining Method to Computer Forensic

Najib Saylani

Hofstra University, Hempstead New York

najib.saylani@hofstra.edu

Abstract:

This paper proposes the exploration of the use of Backward Chaining as one of the many methods utilized in the automated approach to events' analysis in the area of computer forensics. In addition, steps to be taken are outlined in developing an expert system that implements such a method. The new method compliments other methods that traditionally focus on searching files and performing pattern matching. Specifically, this approach is explored to complement other advanced tools in automated data analysis.

Keywords: Computer forensics, Backward and Forward Chaining, Expert Systems

Introduction:

Computer forensics is the process of mining for digital evidence related to an on-going investigation. Traditionally, this process is initiated after a crime is committed and its focus is on extracting data from different sources of the IT infrastructure (i.e.; text, graphics, video, various media files and others). The process can be generalized to the one of pattern matching. A computer forensic specialist would try to match a known pattern to one represented by some of the stored data. All matches become part of a list of evidence to be used in a court of law and for prosecuting the offender. Initially, the process was static in nature. The media to be mined must first be confiscated and forensics performed on it later. As the IT infrastructure developed to a more complex environment consisting of networked computers and networks that distribute processes over LAN, MAN, and WAN; conducting forensics work became more difficult. Current IT environment are known to be complex dynamic systems in need of real time monitoring. The method that will be discussed in this paper complements the one presented in [4] where Cellular Automata where explored. Advanced tools were developed to overcome the difficulties associated with this new environment. These tools help, not only on mining for matching patterns, but also to recreate events associated with the creation of such patterns. Today, there are some advanced tools that help in conducting live forensic investigations. As discussed in [2], security breaches are getting more complex and sophisticated and this would require not only collaboration between experts but in our opinion advanced tools and even non-traditional ones should be explored. This is a set of tasks that are initiated while the whole IT system is in operation.

The challenges facing researchers and practitioners alike are the lack of forensic readiness [2]. In many cases it is not enough to recover evidence. It is more important to find the offender, locate the intruder, and more importantly secure the infrastructure by minimizing, or if possible, eliminating vulnerabilities. A focal point in this paper is on the collection of evidence, and linking this evidence to the hypothesis. Also, the focus is on specific cases where the offender, either known or anonymous, is found to be trying to conceal traces of intrusions of data manipulations through various attempts (i.e. deleting, moving, relocating evidence, changing temporal stamps, etc).

The following conditions must be met for this approach to succeed:

- Partial evidence found which indicates malicious and criminal activities (i.e.; knowledge acquisitions)
- A hypothesis formed related to the offending process/entity.

We propose the development of an expert system that would support the hypothesis through the Backward Chaining method.

The Case for Backward Chaining:

In the context of computer forensics the Backward Chaining method consists of the following characteristics: [3]

- An event or a set of events that are representative of some type of compromise or breach of the IT infrastructure; or witnessed evidence of criminal activities initiated from within or outside of the organization. These activities are either finalized or ongoing. In this case, forensic processes start and the task of analyzing these events is initiated.
- Given that evidence of compromises or criminal activities is found in the present; it is necessary to trace back all causes that are the precursors of such evidence.
- A list of all consequences is to be created based on what really constitutes the pertinent elements of the evidence. Time constraints are very important to the investigation and a process of "normalization" or elimination of unnecessary evidence is to be achieved. A lookup for all antecedent events is initiated.
- Traditionally, the process of computer forensics is data driven, which is similar to the Forward Chaining approach. A forensic expert would collect any set of data part or stored patterns that match patterns of found evidence, a known set of patterns representing intrusion compromises, and/or criminal activities. In the Backward Chaining approach, the process is goal driven through top-down reasoning.
- In this case, proof of compromising activities or evidence of breaches is already witnessed. The expert would work backward using the found hypothesis about what really happened or is happening to find facts supporting the formed hypothesis.
- The consequences (i.e. evidence) of all antecedent events (causes) should be carefully analyzed in order to efficiently determine the rules by which a search for these antecedents can be initiated.
- All previously listed characteristics, especially the latter, will help in the creation of an explanation facility that outlines the reasoning used by the forensic expert; or the steps that would be undertaken by an expert system.

One of the goals in this paper is to outline the development of an expert system that implements the backward chain method for computer forensics.

Outline of the development of the system:

Prior to outlining the process of the development of such a system, let us revisit the Backward Chaining method. We should note along the way that the expertise needed for this method is to be extracted from a human forensic expert. The set of rules by which an expert would infer from a set of evidences is a combination of conjunctions and disjunctions. There is no single hypothesis to support; there is only a series of low to high level hypotheses. Questions are to be well formed and a set of what-if analyses are to be raised.

Visual illustration of the Backward Chain

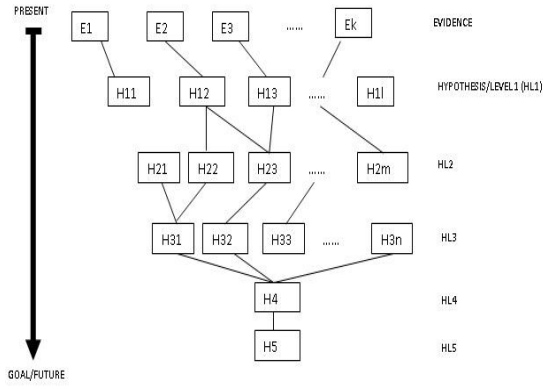

figure 1

Figure 1 illustrates a typical view of a given situation.

- Forensics start with a set of evidences (E*1*, E*2*, E*3*, …, E*k*): facts known when patterns of illegal activities or other compromises or breaches are detected. These patterns can be

the output of various monitoring tools and manual search for data that is judged by experts to be a part of sought patterns.

- A set of hypotheses is formed (H*ij* where *i* is the level *j* is the number of the hypothesis: the number of the hypotheses depends on the complexity of the problem at hand and the number of objects involved (i.e.; employees plus computers) events associated with the use of the IT infrastructure (i.e.; login, program execution, remote access, storage process, etc.). As figure 1 shows hypotheses are formed in time and belong to different levels. The goal in this process is to infer a final conclusion (H*5*). Every hypothesis or evidence can be re-evaluated based on new input from other sources and other monitoring and forensic tools.
- The process of evidence creation and hypothesis inferences can either be a static one, in the case of non-live analysis; or dynamic when the analysis is conducted live. Both have advantages and disadvantages. [1]. An example in the case of non-live analysis disadvantage is that if breaches are not caught early enough, then finding the offenders may be difficult if they already accomplished whatever they planned, and had enough time to eradicate some valuable evidence. In the case of live analysis, an example of a disadvantage is the distribution that the process might cause to ongoing normal operations of the IT infrastructure.

Regardless of these problems creating such ES should follow proven ES development procedures. We are adopting approaches outlined in [3] and adopting them to our special case of using the Backward Chaining method for application in computer forensics. A known method of software engineering in the case of commercial application is the SDLC (System Design Life Cycle). In our case we will call it the ESLC (Expert System Life Cycle).

The ESLC is actually a process model that represents the steps a designer/developer would undertake while considering the order of execution of each step, the time allocated to each step and the resources needed to complete each step. In [3] it is labeled as a meta-methodology, although, traditional methodologies of designing and developing ES are still applicable. The two most important elements of ES are:

- The knowledge base: a collection of information about historical cases of computer forensics. In our case, a dynamic set of knowledge that gets updated as the investigation goes on. It is very important to emphasize here that looking back at past successful cases of computer forensics would definitely help. This is especially true if they reflect scenarios that closely match the ones being presently tackled, from a practitioner point of view, that past knowledge represents just experience. There is the potential for facing scenarios that are novel and do not show all features of past scenarios. Also, evidence is found as the investigation progresses and this newly discovered evidence becomes part of the knowledge base.
- The inference engine: a collection of inference rules adapted to different scenarios. These rules exist because they were extracted from human forensics experts. These rules are themselves subject to update, deletion, and upgrade. Other rules may be added under the supervision of human experts.

The latter two important aspects of this system, especially real time modification of the knowledge base during an ongoing investigation, represent a case of what is called an “incremental model”. In general, and as depicted in [Giarattano], the following is a visual representation of such model, which is called the “spiral model”. The reason for this representation is that the process of adding, removing, or updating functionalities in the system is a never ending process that “macro-

spiral" over the life of the system, or "micro-spiral" during the period of investigation:

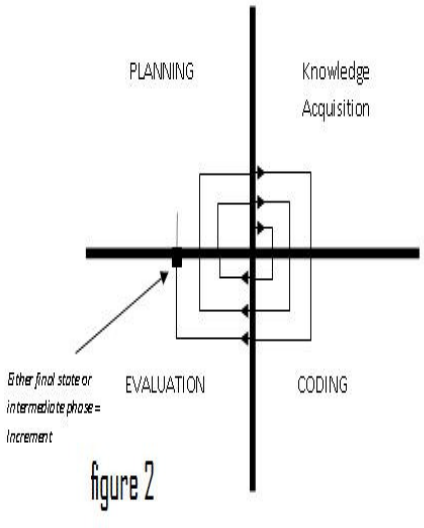

figure 2

Important steps in the ESLC:

- Planning: the feasibility of the system is determined by the fact that human forensic experts use the same approach in the backward chain method when analyzing a given case. This approach is a candidate for building an ES.
- Knowledge Acquisition: this step can be divided into two parts:
  --- "Batch mode" acquisition when historical data about past forensic cases is stored.
  --- "Real time" acquisition during ongoing investigations when current information is added to the knowledge base.
- Coding: a choice of a programming environment must be selected. Prolog seems to be the right language for this purpose. An interface to the ES must be developed, but more importantly, rules imported through the human expert are coded.
- Evaluation of the system: The ES state must be tested and all rules must be evaluated. At this level, as with similar systems development and in some cases at the end of each step, a re-evaluation of every state is conducted. There is always the possibility of revisiting the whole circuit many times until all components of the system are working to satisfaction (see figure 2).

The methodologies, as related to the Backward Chaining approach, are not difficult to implement. All inferences are based on very simple logical operations.

Conclusion:

The Backward Chaining method is an attractive approach to use in the area of computer forensics. The steps that make up this method closely mimic those that a human forensic expert would undertake. Intuitively, the Backward Chaining method in the case of live diagnosis of collected evidence and application of inference rules can be easily converted to simulate a forward chaining method. This latter option makes it easy to validate all implemented rules; either those that are part of the initial inference engine or those constructed as the investigation goes on. The last part in this paper briefly outlines the main characteristics and steps of an ESLC. It highlights the easiness by which one can develop such a system, as the logic that supports it is more accessible. Finally, a system based on the Backward Chaining method would complement existing forensics tools by monopolizing a specific area of computer forensics, thus freeing more time for the human expert or making the tasks of computer forensics more efficient.

References:

1-Brian, D. C., Risks of Live Digital Forensic Analysis, CACM, v.49, n.2, p. 56-61, (Feb 2008).

2-Casey, E., Investigating Sophisticated Security Breaches, CACM, v.49, n.2, p. 48-54, (Feb 2008).

3-Giarratano, J., Riley G., Expert Systems: Principles and Programming, PWS Publishing Company, 1998.

4-Saylani, N., **A** Novel Approach to Modeling a Highly Dynamic System for IS Security, Proceedings of the NEDSI 2008