Computers & Operations Research 69 (2016) 1–9

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

A heuristic for maximizing investigation effectiveness of digital forensic cases involving multiple investigators

Jatinder N.D. Gupta [a,*], Ezhil Kalaimannan [b], Seong-Moo Yoo [c]

[a] College of Business Administration, University of Alabama in Huntsville, Huntsville, AL 35899, USA
[b] Department of Computer Science, University of West Florida, Pensacola, FL 32514, USA
[c] Department of Electrical & Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899, USA

ARTICLE INFO

Available online 19 November 2015

Keywords:
Digital forensics
Crime scene investigation
NP-hardness
Mixed integer programming
Heuristic solution
Computational results

ABSTRACT

Digital forensic investigation refers to the use of science and technology in the process of investigating a crime scene so as to maximize the effectiveness of proving the perpetrator has committed crime in a court of law. Evidences are considered to be the building block of any crime scene investigation (CSI) procedure including those involving cyber crimes. Selecting the right set of evidence and assigning the appropriate investigator for the selected evidence is vital in time critical forensic cases, in which results have to be finalized within a specified time deadline. Not doing this may lead to the scope creep problem, which is a significant issue in digital forensics. Therefore, major challenges with respect to digital forensic investigation are to determine the right set of evidences to be assigned to each of the available multiple investigators and allocate appropriate investigation time for the selected evidences to maximize the effectiveness of the investigation effort. A mixed integer linear programming (MILP) model is developed to analyze and solve the problem of evidence selection and resource allocation in a digital crime scene investigation. In view of the problem being NP-hard, a heuristic algorithm with polynomially bounded computational complexity is proposed to solve the problem. Results of extensive computational experiments to empirically evaluate its effectiveness to find an optimal or near-optimal solution are reported. Finally, this paper concludes with a summary of findings and some fruitful directions for future research.

1. Introduction

Digital forensic investigation is the procedure of examining a crime scene once a fraud or a crime is suspected to have been committed. The reasonable rate of overall effectiveness of digital forensics and the incident response procedures in organizations was reported to be around 55% [8]. However, the rates for marginal and very effective procedures were noted to be only around 23% and 20%, respectively. Therefore, there is an urgent need to develop effective and efficient techniques to solve the digital forensic and crime scene investigation (CSI) problems.

The first basic CSI model described by Pollitt [14] requires completion of three primary steps namely acquisition, identification, and evaluation of computer forensic investigation before admission of the evidence in court. This model further described the fact that the path taken by any digital evidence comprises media (Physical context), data (Logical context) and information (Legal context). Kerrigan [10] presented the digital investigation process as a Capability Maturity Model (DI-CMM) as a tool for analyzing an organization's digital investigation capability. This model can be applied to real-time digital investigations to improve its current capabilities and how the methodology highlights differences with an organization's subjective perception of its capabilities.

Computer Forensic Field Triage Process Model (CFFTPM) proposed by Rogers et al. [16] and Rogers and Seigfried [17] is designed to complete digital forensic investigation in a short time frame without the requirement of taking the system/media back to the laboratory for an in-depth examination or acquisition of a complete forensic image. The authors of this model considered time as an expensive commodity in some special cases where results are balanced against the time spent in the forensic investigation procedure. This model is consistent with various other models and general enough to be applied across a wide spectrum of investigations. Bulbul et al. [6] developed the Analytical Crime Scene Procedural Model (ACSPM) that primarily focuses on digital

* Corresponding author. Tel.: +1 256 824 6593; fax: +1 256 824 4339.
E-mail addresses: guptaj@uah.edu (J.N.D. Gupta), ekalaimannan@uwf.edu (E. Kalaimannan), yoos@uah.edu (S.-M. Yoo).

http://dx.doi.org/10.1016/j.cor.2015.11.003
0305-0548/© 2015 Elsevier Ltd. All rights reserved.

crime scene investigation procedures rather than focusing on whole digital investigation process and phases that end up in a court. This model clearly analyzed the relevant literature models in an analytical way in order to provide a model with thorough and secure implementation of digital forensic investigation procedures at a crime scene.

Overill et al. [13] proposed triage template pipelines to guide the investigation procedure, enabling devices and data that they contain to be examined according to a number of prioritized criteria. This approach is specifically targeted over examinations done at the laboratory and hence is significantly different from the on-site triage forensics dealt by Rogers et al. [16]. A model for handling incident analysis and digital forensic investigations in SCADA (Supervisory Control and Data Acquisition) and industrial control systems was proposed by Spyridopoulos et al. [18]. In the light of significance of SCADA for the resilience of critical infrastructures and the related incidents against them, this model focuses on analyzing the current capabilities of SCADA operations to handle security incidents from a robust cyber security and digital investigation perspective. Further, this model analyses the logging capabilities of SCADA systems and the analytical and investigative tools that help in managing the forensic readiness of the current threat requirements. A comprehensive review of major works related to shaping the process of digital forensic investigation was presented by Agarwal and Kothari [1]. The authors argue that there is a need to make digital forensic research more effective through the creation of new forensic investigation models.

Zainudin et al. [20] extended the existing models to form a focused investigatory model for online social networks (OSN) such as Facebook, Twitter and LinkedIn. Bogen and Dampier [4] presented a software engineering perspective with a core set of modeling views for a unified computer forensics modeling methodology: investigative process view, case domain view, and evidence view. The authors of this model describe investigative process as a sequence of activities relating to standard operating procedures and examiners' notes. To assist the forensic analyst in the evidence search process, Herrerias and Gomez [9] developed the log correlation model to collect, filter and correlate events coming from diverse log files in a computer.[1] Wang and Daniels [19] applied graph modeling approaches to network forensic analysis to facilitate evidence presentation, automated reasoning, and interactive hypothesis testing in an attempt to identify the attacker's non-explicit attack activities from the secondary evidence.

Bashir and Khan [3] outlined a triage principle based framework to carry out digital forensic investigations that are time consuming in nature. The authors argue that the amount of data found in a digital crime scene is constantly increasing and therefore a substantial amount of time is needed to acquire the digital evidences and perform analysis on them. For example, if the digital investigation is based on a novel attack, it takes enormous amount of time to trace the evidence followed by performing all the necessary steps from problem identification to problem resolution. Raghavan [15] presented a detailed literature survey on digital forensic research since the year 2000, in which the conceptual and practical advancements in digital forensic investigation are described. The author illustrates that the majority of digital forensic investigations are conducted manually, because the forensic tools and investigative methods currently in existence are designed to locate pieces of digital evidence; however, they do not assist in analysis or examination of collected evidences i.e., to conduct an investigation effectively.

From the brief review, it follows that the existing literature has mostly dealt with conceptual and procedural models for analyzing and improving the overall investigation procedure. No existing model applied mathematical optimization approaches to select and allocate resources in digital forensic investigations. Realizing the above weaknesses of the existing literature, for the first time, this paper proposes a Mixed Integer Linear Programming (MILP) optimization model for a sequential and parallel scenarios of evidence analysis and examination. The goal of this proposed model is to maximize the overall effectiveness in identifying the perpetrator of the crime.

The rest of the paper is organized as follows: Section 2 formulates an optimization model for the sequential digital forensic investigation with multiple investigators and proves that the problem is NP-hard at least in the strong sense. Section 3 develops a polynomially bounded heuristic algorithm to generate an optimal or near-optimal solution and discusses appropriate strategies to extend the proposed model to make it relevant to digital forensics practice. The computational results about the effectiveness of the proposed heuristic algorithm in finding an optimal or near-optimal solution are provided in Section 4. Finally, Section 5 concludes the paper with some fruitful directions for future research.

2. Sequential digital forensic investigation model

The entire process of CSI differs in the aspect of how it is being executed in terms of field work and laboratory [11]. The team of experts who work at the crime scene and in the laboratory are termed as crime scene analysts and forensic scientists respectively. Not all crime scene analysts are forensic scientists. Some crime scene analysts just work in the scene to collect the evidence and then pass it to the forensics lab. In digital forensic investigations, the analyst must still possess a good understanding of digital forensics in order to recognize the specific or probative value of various types of digital evidence acquired from the scene. Hence, practically in most of the digital forensic cases, these jobs[2] overlap with each other [5].

Evidence analysis and investigation in a laboratory can be classified into two basic scenarios: *parallel* and *sequential*. In the *parallel* scenario, as soon as an evidence is acquired at the crime scene, it is sent to the laboratory for analysis. An immediate examination and analysis of this evidence would help the team at the crime scene to determine what additional evidences need to be collected for further analysis and what evidences need to be discarded that do not fall under the scope of investigation. Also, there might be evidences that are not available until a particular time instance of the whole investigation. This process continues until the perpetrator of the crime is fully identified or the resources available (usually total available time) for the investigation are exhausted.

In the *sequential* scenario, on the other hand, the investigation in the laboratory will begin only after the acquisition of all evidences from the crime scene. Compared to the parallel scenario, the need for transportation from the crime scene to the laboratory for the sequential scenario is reduced. However, the total number of investigators needed in a parallel scenario may be more than those needed in a sequential scenario. Compared to the sequential

[1] Log files are specific files that are generated to keep a history of actions occurred on the system.

[2] In the United States, responsibility for collecting digital forensic evidence from crime scenes is often shared amongst crime scene analysts and forensic investigators/scientists.