A Framework for the Forensic Investigation of Unstructured Email Relationship Data

John Haggerty, University of Salford, UK
Alexander J. Karran, Liverpool John Moores University, UK
David J. Lamb, Liverpool John Moores University, UK
Mark J. Taylor, Liverpool John Moores University, UK

ABSTRACT

The continued reliance on email communications ensures that it remains a major source of evidence during a digital investigation. Emails comprise both structured and unstructured data. Structured data provides qualitative information to the forensics examiner and is typically viewed through existing tools. Unstructured data is more complex as it comprises information associated with social networks, such as relationships within the network, identification of key actors and power relations, and there are currently no standardised tools for its forensic analysis. This paper posits a framework for the forensic investigation of email data. In particular, it focuses on the triage and analysis of unstructured data to identify key actors and relationships within an email network. This paper demonstrates the applicability of the approach by applying relevant stages of the framework to the Enron email corpus. The paper illustrates the advantage of triaging this data to identify (and discount) actors and potential sources of further evidence. It then applies social network analysis techniques to key actors within the data set. This paper posits that visualisation of unstructured data can greatly aid the examiner in their analysis of evidence discovered during an investigation.

Keywords: Digital Forensics, Email, Framework, Social Network Analysis, Structured and Unstructured Data

INTRODUCTION

Email has replaced the letter as the main written communications medium, both in business and personal life. It is estimated that the average employee spends a quarter of their time on email-related tasks and the average number of emails sent by a corporate user per day is 43 (Orloff, 2011). Email can therefore provide the investigator with a wealth of information through both structured and unstructured data. Emails are structured in that they are formatted according to RFCs 5321 and 5322 (Klensin, 2008a, 2008b). However, they also provide unstructured information through the communications links and contacts that form...
social networks. Digital forensics software, such as EnCase (Guidance, 2011) or the Forensics Toolkit (FTK) (Access, 2011), are useful for analysing structured data, i.e., the content of the emails in their textual form. However, they do not elucidate the unstructured data, such as relationships within the email network, power relations or network bridges that may be a key concern to a forensics investigation. Social network analysis and visualisation techniques can significantly contribute to evidence discovery and collection by identifying and understanding relationships and data flow between actors. Moreover, it may be used to quickly identify key events of interest within the email social network.

A number of challenges exist to today’s digital forensics investigations involving email. As with many forensic investigations, cases routinely involve more than one computer (Richard & Roussev, 2006) and the investigator is unlikely to have access to all computers involved in the email network. Digital forensics investigatory models do not currently differentiate between email and any other data. Current work on digital investigations involving email data focus on techniques for the extraction of evidence, for example, data mining (Wei et al., 2008) or clustering algorithms (Bird et al., 2006). Recently there has been some focus on process models for investigations that involve email data. These approaches generally provide a theoretical framework or software application, which detail techniques for the visualisation and extraction of specific email artefacts or features. However, this research focuses on particular aspects of email data rather than the wider process. Finally, much of the evidence that is recovered during an investigation may not be analyzed beyond the structured data view. For example, an examiner will manually trawl through the emails relating to an activity under scrutiny to search for those relevant to the investigation. However, they rarely explore the relevant social relationships and networks that these, and other network communications such as ‘chat’ sessions, will reveal due to the lack of facility in the tools they have at their disposal. These social networks are potentially great sources of interest as they may lead the investigator to other relevant sources of evidence, actors related by events or power relationships that are relevant to an investigation.

This paper presents a novel framework for the forensic investigation of unstructured email data. This framework follows traditional digital forensics procedures but incorporates tools and techniques for the triage and analysis of emails. This is achieved by using social network analysis and data visualisation to identify relevant evidence from unstructured data. This paper is organised as follows. In the next section we discuss relevant literature in the areas of social networks and email analysis. Following this, the novel framework is posited. In order to demonstrate the applicability of the approach, relevant stages are applied to the Enron email corpus as a case study. Finally, we make our conclusions and discuss further work.

SOCIAL NETWORK ANALYSIS AND EMAIL INVESTIGATIONS

Investigations involving social networks have raised in prominence due to the reliance of users on their digital communications and the information about a suspect these data sources may yield to the forensic examiner. Computer forensics tools, such as EnCase (Guidance, 2011) and FTK (Access, 2011), are used by examiners to recreate files and data from a suspect’s computer. However, these tools are designed to analyse structured data rather than elucidate unstructured data that is present in communications. Moreover, these tools do not provide a visualization of evidence or quantify the importance of actors within a social network.

There are various representations of social networks in the digital world, of which email is just one. Other social network representations include friends on Facebook, followers of a Twitter account or a contacts list within a mobile phone, each with different levels of information that they may provide to the examiner. Our reliance on email within the business environment provides the investigator with an
An Adjustable Interpolation-based Data Hiding Algorithm Based on LSB Substitution and Histogram Shifting
Yuan-Yu Tsai, Yao-Hsien Huang, Ruo-Jhu Lin and Chi-Shiang Chan (2016).
International Journal of Digital Crime and Forensics (pp. 48-61).
www.igi-global.com/article/an-adjustable-interpolation-based-data-hiding-algorithm-based-on-lsb-substitution-and-histogram-shifting/150859?camid=4v1a