Digital Investigation 16 (2016) S124–S133

Contents lists available at ScienceDirect

Digital Investigation

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

DFRWS 2016 Europe — Proceedings of the Third Annual DFRWS Europe

Facilitating forensic examinations of multi-user computer environments through session-to-session analysis of Internet history

David W. Gresty*, Diane Gan, George Loukas, Constantinos Ierotheou

C-SAFE Centre, Dept. of Computing and Information Systems, University of Greenwich, UK

Keywords:
Digital forensics
World wide web
Session-to-session analysis
Context analysis
Pattern of life
Internet history analysis

ABSTRACT

This paper proposes a new approach to the forensic investigation of Internet history artefacts by aggregating the history from a recovered device into sessions and comparing those sessions to other sessions to determine whether they are one-time events or form a repetitive or habitual pattern. We describe two approaches for performing the session aggregation: fixed-length sessions and variable-length sessions. We also describe an approach for identifying repetitive pattern of life behaviour and show how such patterns can be extracted and represented as binary strings. Using the Jaccard similarity coefficient, a session-to-session comparison can be performed and the sessions can be analysed to determine to what extent a particular session is similar to any other session in the Internet history, and thus is highly likely to correspond to the same user. Experiments have been conducted using two sets of test data, where multiple users have access to the same computer. By identifying patterns of Internet usage that are unique to each user, our approach exhibits a high success rate in attributing particular sessions of the Internet history to the correct user. This can provide considerable help to a forensic investigator trying to establish which user was using the computer when a web-related crime was committed.

Introduction

During the course of a digital forensics examination, the investigator has a variety of locations and artefacts to search through to find the clues to show if a device was used, misused or contains the evidence required for the purpose of the investigation. In addition to the documents, pictures, media files etc. which could be the immediate target of the investigation, devices such as computers, laptops, tablets, smart phones etc., routinely have Internet connectivity, which can often also provide a treasure trove of investigative clues to how the device was used.

The Internet history is an ordered list of artefacts that contain a date, time and Universal Resource Locator (URL) address of websites or resources that were accessed. The history artefacts show the experienced investigator the types of websites that were accessed and the times of day that the users were active; they provide information about the email and social media contacts that a user has; they can show files, movies and pictures that have been downloaded. The Internet history is where the investigator gets to see the terms submitted to search engines, the poor spelling and the languages that the users speak.

Analysis of Internet history artefacts is however time consuming and to-date an ad hoc process. The artefacts may be few in number due to the small size of the storage medium, incomplete due to normal overwriting actions or ‘private browsing’ anti-forensics or even be quite extensive.

* Corresponding author.
E-mail address: D.Gresty@Greenwich.ac.uk (D.W. Gresty).

http://dx.doi.org/10.1016/j.diin.2016.01.015

Each of these challenges do not detract from the importance of the investigative clues contained within the Internet history artefacts. However, they do dictate whether the Internet artefacts are only usable by the investigator as clues to get a sense of how the device was used, or can be presented as useful evidence in their own right at a court or tribunal.

Above all, the Internet history artefacts show a user, an actual person, interacting with the device. Such interaction, shows the mental component of an action, the *mens rea*, of the person at the keyboard. Brenner et al. (2004) highlight the R v Schofield case from the United Kingdom, where the prosecution was forced to dismiss charges for possession of unlawful pictures because Trojan Horse software was located on the defendant's computer and ultimately the analysis had not established responsibility for the creation of the unlawful pictures resting with the defendant, or indeed any actual person.

For the forensic investigator it is difficult to show the intent of the user of the system without placing the artefacts into contextual ordering. In the above case, the unlawful pictures were considered in isolation and consequently the intent of a user had not been distinguished from that of the activity of a Trojan Horse program. However, if for example in that case there were other artefacts showing search terms submitted to a web browser before the pictures were downloaded or link files after the download that showed the access of the pictures, then despite the fact that the Internet history and the link files were different artefacts to the pictures, we would see a timeline of artefacts that an analyst could contextually order. If downloaded pictures on Schofield's system were always preceded by search artefacts and followed by link files showing access this would also form a pattern, and if the pattern could be observed, occurring over a variety of times, then the existence of repetitive behaviour could be established.

One-time and repetitive patterns

We propose in this paper that activity on a computing device, and specifically within the Internet history for such a device, can be classified as either one-time events or repetitive events:

- A One-time event could be a single event such as the moment malware was executed, or a short series of events that are never repeated such as someone searching the Internet for the phrase "how to make a bomb" then proceeding to view a number of websites that are relevant to the search term, but no subsequent search or similar web page access is located within the Internet history.
- Repetitive events show habitual patterns. Therefore, to be considered a repetitive event, an event must occur and re-occur at least at one other point. Repetitive events are temporally ordered sequences or clusters of activity within a temporal proximity to each other. Sequential patterns are such that A, B and C occur, in that order, within a timeframe. Temporal Clusters are where A, B and C may occur in any order, combination or repetition within a timeframe. For example, ACBAB.

Certain crimes or investigative goals lend themselves to the identification and analysis of repetitive or habitual behaviours, such as the accessing of indecent material or 'grooming' types of offences. Wherever there is a concern about who was the operator of a device at a particular point in time, even if that particular point in time is a one-time activity, such as the sending of an inappropriate email, if the one-time event is in close temporal proximity to a repetitive pattern then an investigator may be able to demonstrate the likeliness that the user at the time of interest is the same user at a number of other times, which can refute the "it wasn't me" defence, as there is certainly the appearance of a regular user of the device operating it at that time.

Within this paper we discuss related work to the analysis of Internet history artefacts and the profiling of digital devices. We describe our approach of aggregating the Internet artefacts into groups and then show how these can be broken into components which allow the aggregate groups to be compared to each other to see if there is overlap in the membership. The experimental section of this paper describes two different problems: finding the most effective time value for the aggregation into sessions and testing the effectiveness of the session-to-session comparison. As this paper reports on an ongoing research project, our evaluation and conclusions review the encouraging results from the experiments and highlight possible techniques to improve the performance of our approach.

Related work

One of the first attempts for a tool for forensic timeline analysis was Zeitline, introduced by Buchholz and Falk in 2005 (Buchholz and Falk, 2005). Its purpose was to reconstruct artefacts and enable an investigator to create complex events, using searching and filtering to populate and analyse timelines. Different applications and different operating systems leave behind footprints of their activity. The approach by Khan and Wakeman (2006) is to determine the footprint of applications on a system based upon the typical artefacts that are created in normal usage. These features are then used to train a neural network which could be used during a forensic examination to attempt to reconstruct a timeline of events concerning when applications were used.

In 2009, the Cyber Forensic TimeLab (CFTL) tool was developed by Olsson and Boldt (2009). CFTL can parse a hard drive for known predefined artefacts to produce a histogram timeline. It does not automatically analyse the artefacts, but requires the analyst to make a visual correlation of different timelines overlaid to display clusters.

Another tool for forensic investigations, log2timeline, was reported in Gudjonsson (2010). This tool creates a super-timeline by placing all the information into a monolithic list which can then be processed. This approach was endorsed by Carbone and Bean in the review of