Digital Investigation 25 (2018) 105–113

Contents lists available at ScienceDirect

Digital Investigation

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

I didn't see that! An examination of internet browser cache behaviour following website visits

Graeme Horsman [1]

Faculty of Computer Science, The David Goldman Informatics Centre, St Peter's Way, SR6 0DD, Sunderland, United Kingdom

ARTICLE INFO

Article history:
Received 15 May 2017
Received in revised form
21 January 2018
Accepted 25 February 2018
Available online 2 March 2018

Keywords:
Digital forensics
Cache
Internet
Web browser
Cache forensics
Images depicting child sexual abuse

ABSTRACT

By default, all major web browsing applications cache visited website content to the local disk to improve browser efficiency and enhance user experience. As a result of this action, the cache provides a window of opportunity for the digital forensic practitioner to establish the nature of the content which was hosted on the websites which had been visited. Cache content is often evidential during cases surrounding Indecent Images of Children (IIoC) where it is often assumed that cached IIoC is a record of the content viewed by a defendant via their browser. However, this may not always be the case. This article investigates web browser cache behaviour in an attempt to identify whether it is possible to definitively establish what quantity of cached content was viewable by a user following a visit to a website. Both the Mozilla Firefox and Google Chrome browser caches are analysed following visits to 10 test websites in order to quantify cache behaviour. Results indicate that the volume of locally cached content differs between both web browsers and websites visited, with instances of images cached which would not have been viewable by the user upon landing on a website. Further, the number of cached images appears to be effected by how much of a website a user scrolls through.

Introduction

The cache functionality of an Internet browser application is a well documented and discussed concept in the field of digital forensics (see The Chromium Projects n.d.; Habben, 2015; Ritchie, 2012). Its job is to enhance a user's web-browsing experience by downloading and storing a local version of website artefacts to provide increased efficiency in the re-rendering of a website on future visits (Howard, 2004). The cache can offer an insight into the browsing habits of a user, where although Internet history records may document the locations a user has visited online, the cache can reveal the content hosted on these webpages. Cached content can provide a vital source of evidence in many investigative scenarios and most notably, in investigations surrounding the possession, distribution and creation of Indecent Images of Children (IIoC) (see for example United States v. Tucker, 305 F.3d 1193 (10th Cir. 2002)).

Whilst at first glance, analysis of the cache may seem straightforward (in terms of understanding the structure of its stored data), questions regarding its functionality are raised, particularly in relation to the volume of data which is cached and when caching occurs. To provide context to cache related investigatory issues, a discussion of regulatory concerns surrounding IIoC and the web browser cache is offered. In cases where IIoC are found in a defendant's browser cache, cases often revolve around a defendant's knowledge of the cache in order to attribute some form of culpability over this content (Marin, 2008). In English law, liability for possession may ensue if a user knows of the cache (i.e. knows of its existence on their digital device), subject to legal tests of possession (see Atkins v DPP and Goodland v DPP, 2000 2 Cr. App. R. 248) and statutory defences (see Criminal Justice Act 1988 (CJA88), Section 160 (2)). Of particular interest (if knowledge of the cache is established) is the CJA88 Section 160 (2) (b), where a defendant may rely on a statutory defence if they can prove that they “had not himself seen the photograph [or pseudo-photograph] and did not know, nor had any cause to suspect, it to be indecent”. In this situation, the requirement to have ‘not seen a photograph’ provides an area for exploration given that the cache is an automatic function, storing the content of visited websites. This defence requires establishing what a user has viewed on their screen, a task that during a *post mortem* investigation can only be established through analysis of cached data. Yet there is currently limited research analysing the functionality of web browser caches in terms of how much of a

E-mail address: g.horsman@tees.ac.uk.

[1] Present address: Teesside University, School of Science, Engineering & Design, Campus Heart, Southfield Rd, Middlesbrough, TS1 3BX, United Kingdom.

https://doi.org/10.1016/j.diin.2018.02.006
1742-2876/© 2018 Elsevier Ltd. All rights reserved.

visited website is cached, and crucially in this context, whether it is possible to establish which (or if) content is cached without a user ever physically seeing it on their screen. Establishing with accuracy which cached files were viewable on screen and which were not, may support the application of the defence under CJA88 Section 160 (2) (b) (and equivalent international law regarding a defence involving sight) with a greater degree of reliability.

This article provides a discussion of the functionality of the Mozilla Firefox and Google Chrome Internet browsing applications, not from an information-parsing standpoint, but from a behavioural context. The digital footprint left behind in the cache of each browser is examined and correlated against standard user browsing behaviour (both on landing and following a page scroll) in order to establish whether through cached-content, it is possible to identify which parts of a website were visually present on-screen and arguably viewed by a user.

The cache

Although the structure of the cache differs between browsing applications (see discussions by Altheide and Carvey, 2011), its overarching functionality remains the same; to improve browsing performance. Cache setups are configurable by the user or in some cases can be turned off (with performance detriments), however, by default, all mainstream browsing applications dedicate a region of local storage media for the caching of website artefacts which can include text, media, application and site structural content. As cache content is utilised in the rebuilding of websites by a browser upon a re-visit by the user, cache content can also support the offline rebuilding of webpage content during forensic investigations (see tools such as NetAnalysis (Digital Detective, 2017) and IEF (Magnet Forensics, 2017)). However, Casey (2009) expresses the need for caution when undertaking such processes due to the potential for unreliable results due to the high turnover of files in the cache where multiple artefacts maybe similarly named and lead to inaccurately rebuilt pages. Even without cached page rebuilding, it may be possible to correlate the creation time and date of individual cached artefacts against Internet history records to identify websites which were visited and of evidential value. This is often a process involved in IIoC investigations where the Internet now often provides a main source of this material (Horsman, 2016).

IIoC and the cache

IIoC found in the Internet browser has been the subject to legal debate where arguments are offered both in terms of an offence of possession and that of making (Marin, 2008). The difficulty lies with the fact that the function of a web-browser cache is automated by design, which subsequently allows imagery hosted on browsed websites to be collected and stored.

The function of the cache is legitimate, but assigning culpability for its content poses issues. To determine whether a defendant is guilty of an offence of possession in regards to IIoC stored in their browser cache, a question of what constitutes a person having ‘possession’ of the cache's content is crucial. In English law, a possession offence is offered under Section 160 CJA88 where possession involves both a physical and mental element (CPS, 2017). To be in possession of cached images a defendant must have custody and control of the images (be able to retrieve/access them) and knowledge of the images following Atkins v DPP and Goodland v DPP, 2000 *2 Cr. App. R. 248*, where a defendant's knowledge of the existence of the browser cache must be established. To try and simplify, a defendant cannot be in possession of an IIoC if they do not know about its presence on their system, and following *R v Porter [2006] EWCA Crim 560*, to have custody and control over an image, a defendant must be able to access that image. Where both knowledge and, ‘custody and control’ are established, a defendant is deemed to have possession of an image. In this instance, a defendant may seek to rely on one of the three statutory defences under the CJA88 Section 160 (2) if they can prove (on the balance of probabilities) that they had a legitimate reason for possessing an image, that they had not seen the image or suspected it to be indecent, or finally, that the image was sent without any prior request and it was not kept an unreasonable amount of time (Wall, 2017). To circumvent the difficulties associated with establishing possession, particularly involving the cache, where there is evidence of a deliberate intentional act (see *R v Bowden [2000]* 1 Cr. App. R. 438), such as searching for IIoC online, a charge of ‘making’ may be attributed.

Under normal browsing circumstances consideration as to what a user of a web browser has actually physically seen on their screen is of little evidential value. Yet in cases of IIoC, the cache is assumed to be a record of what a defendant has viewed leading to potential liability, as noted above. Cached IIoC provide an insight into the severity of the offence committed (see CPS (2017) for categorisation and sentencing guidance), but limited consideration is given as to whether these images have actually been physically seen by a defendant. Arguably this stems from a lack of complete understanding, not at a technical, but functional level of the web browser cache. Although the technical cache structure is relatively well documented (see The Chromium Projects n.d.; Habben, 2015; Ritchie, 2012), often by those involved in forensic analysis, there is limited research available demonstrating the impact on the cache caused by standard user browsing actions. To place this in context, focus is drawn to the following quote by McBath, (2012), p.389.

> “the first time a user visits a website two simultaneous processes occur: (1) the computer opens the website and shows it on the screen, and (2) the computer creates a copy of all the data on that website and stores it in the cache. Thus, an image will not be stored in the cache unless the website from which it came was, at one time, on the computer screen …. Images found in the cache are simply evidence of the prior possession that the defendant had when the images were on his screen” (McBath, 2012, p.389).

This statement raises the following three generalisations regarding the cache which are arguably in need of further investigation.

1. “The computer creates a copy of all the data on that website and stores it in the cache”.
2. “An image will not be stored in the cache unless the website from which it came, was at one time, on the computer screen”.
3. “Images found in the cache are simply evidence of the prior possession that the defendant had when the images were on their screen” (McBath, 2012, p.389).

IIoC are a product used for sexual stimulation, which is arguably achieved when the imagery is physically viewed, an act often condemned (see dissenting comments in United States v. Goff, 501 F.3d 250, 258 (3d Cir. 2007) and McBath (2012)). Yet it may not be accurate to assume that cache content has always been visible to the user on their screen. Website structures vary greatly in shape and size and it remains a distinct possibility that users can visit a website and not physically witness all content hosted upon it without a thorough inspection. In addition, it is necessary to differentiate between a user who mistakenly visits a site and one who examines all content visually, where section Cache behaviour provides an analysis of the behaviour of the cache.