SCC.443 Information System Forensic Investigation

An analysis of anti-forensic techniques, with a consideration of the future of digital forensics

Ben Goldsworthy, 32098584
b.goldsworthy@lancaster.ac.uk

Abstract

I. Introduction

The common perception of modern forensic investigation techniques owes much to TV series' like *CSI: Crime Scene Investigation*. In a standard episode, the team examine the reliably grisly crime scene *de jour* in minute detail until—aha!—the killer left half a fingerprint on the back of a wine glass in the adjoining room. Before long, the fingerprint's owner is discovered and justice duly served. Sometimes, when the writers feel the need to shake up the formula for an episode, the killer might think to wear gloves, or a hairnet, or some other anti-forensic tool. Whilst the field of digital forensics is largely unrepresented in such media, the issues facing it are the same—anybody taking part in potentially-illegal activity is going to be doing everything that they can to not make things easy for a future investigator. Obviously, this goes beyond the wearing of gloves, although that would limit the fingerprints on the keyboard.

In this essay, the various types of anti-forensic techniques are categorised and example tools of each given. Then, the three that are considered to pose the greatest threat to the forensic analyst and their work are further analysed, along with the responses to each from the digital forensics community. Finally, possible futures for forensics and anti-forensics are suggested. Whilst the use of anti-forensic tools is examined from the point-of-view of a criminal user attempting to evade law enforcement, it is worth noting that such tools can also be used to protect activists from prosecution by authoritarian regimes where such a regime is still nonetheless bound by a need to present evidence to back their charges, and to protect their connections and allies where such a regime is not.

Additionally, two notions that may be considered anti-forensic are not considered below. Firstly, legal restrictions on bodies attempting forensic analyses of devices, such as whether they may compel a suspect to provide access via passwords or biometrics (Vaas 2014). Whilst this topic can provide enough interesting material for several essays, any analysis would be too limited to individual jurisdictions for the purposes of the current overview of universally-applicable techniques. The second notion is that of physical booby traps on devices. Though these certainly pose a threat to forensic investigators (particularly first responders), they are here considered to fall within the field of physical forensics.

II. A Taxonomy of Anti-Forensic Techniques

There are a range of anti-forensic techniques available to the aspiring miscreant. They are generally united in their aim—to '[m]ake it hard for them [law enforcement] to find you and impossible for them to prove they found you' (Berinato 2007)—but differ in where along the forensic process they target to cause such disruption. Within this 'genus' of anti-forensic techniques there is as yet no universally-accepted delineation of the distinct 'species'. However, Rogers (2005) proposes the following categorisation of anti-forensic techniques into:

- data hiding;
- artefact wiping;
- trail obfuscation; and
- attacks against the forensic process/tools.

In this section, each of the four types shall be examined in further detail and examples of tools (where possible) presented.

A. Data Hiding

Our legal system operates on a presumption of innocence, wherein the prosecuting party assumes the burden of proving that the wrongdoing which they allege has actually occurred. Obviously, if such proof cannot be found, any case will collapse. Data hiding encompasses techniques that attempt to stymie the forensic investigator's attempts to find illicit material, even in the event of a device seizure or data interception. These attempts vary in complexity. One could attempt to encode sensitive text in a non-human-readable form, but it can be easily decoded if the encoding process is known (e.g. Base64). Alternatively, encryption can be used to achieve the same ends, and is only decryptable with possession of the correct key (b. 2012). Data can also be hidden in unexpected places, such as in memory (Offensive Security 2015), within hidden partitions (VeraCrypt 2017) or by using the Alternate Data Streams provided by Windows and the NTFS file system format. Finally, steganography describes the process of hiding sensitive data within innocuous data. For example, the data for an illicit image may be placed over the least significant bits (LSB) of the colour values of another, legitimate image. As the LSB has a negligible effect on the image when opened in image viewing software, the effect of overwriting the values will be unnoticeable. However, the illicit image can be extracted by anyone who knows where to look for it. Alternative, a benign text file could contain another text file in the form of whitespace characters. bartitsu59 (2018) proposes the use of archaic file formats in order to further confound efforts at discovery.

B. Artefact Wiping

In 2015, FBI agents captured the laptop of Ross Ulbricht, the alleged mastermind behind the Silk Road online cryptomarket. They first drew his attention with a 'staged lovers' tiff', whereupon an agent snatched the laptop and transferred its files over to a USB flash drive—all Ulbricht could do was watch (Bertrand 2015). Other software developers were watching too, however. Shortly thereafter, usbkill was released online (Hephaest0s 2016). usbkill runs in the background of the user's PC, checking any newly-gained or -lost USB connections against a whitelist. If the connection, or sudden lack thereof, is not pre-approved, then the software can be configured to respond with anything from a lock screen to a full wipe.

usbkill/Silk Guardian[1] is a prime example of an artefact wiping technique. Whereas data hiding aims to keep an investigator from finding the evidence they need, the aim of artefact wiping is that there shall be no such evidence *to* find. The degree of precision can vary, from a full-disk wipe to the targeted deletion only of sensitive files and directories. The techniques can also vary in effectiveness—obviously just moving your incriminating file to the Recycle Bin will only serve to save an investigator time, but 'permanent' deletion too can leave traces of the file on the hard disk, if the memory is not written over afterwards. Additionally, Rogers (2005) describes 'prophylactic' techniques that avoid producing artefacts in the first place, rather than deleting them afterwards. One example of this could be the use of an amnesiac operating system such as Tails, which is run from a Live USB or DVD and will 'leave no trace on the computer...unless you ask it explicitly' (The Tails Project 2009).

[1]Silk Guardian is a Linux Kernal Module re-implementation of usbkill, see Brune (2016)

C. Trail Obfuscation

Trail obfuscation techniques take aim at the attribution phase of forensic investigation, operating on the assumption that an investigator may gain access to incriminating data and in a form useful to them. One trail obfuscation tactic is the modification/wiping of automatically-generated system logs to hide evidence of suspect activity and its timings. Alternatively, willing and unwilling intermediates can be used in order to insulate the suspect from identification. For example, criminal activity can be performed on the account of a compromised user without their knowledge, or from behind a VPN. This can make assigning blame difficult for the investigator, as well as enabling an attacker to potentially frame other entities for their activities.

D. Attacks Against the Forensic Process/Tools

Hartley (2007) argues that attacks against the forensic process can be classed as 'antiforensics', whilst those targeting specific tools with the aim of '...directly prevent[ing] the investigator from analyzing collected evidence' represent 'counter-forensics', in the same vein as the distinction between '..antiterrorism and counter-terrorism...'. Attempts to undermine the forensic process rely on the high standards placed on evidence within the court system, and aim to introduce reasonable doubt in order to have potential evidence discounted. For example, using a tool like Timestomp (Bishop Fox 2014) can easily modify file timestamps, confounding investigator attempts to establish a reasonable timeline of events: '[f]orensic investigators poring over compromised systems where Timestomp was used often find files that were created 10 years from now, accessed two years ago and never modified.' (Berinato 2007)

Turning now to counter-forensics, the range of available forensic tools is relatively small, with the subsequent 'Vendor & Tool dependency...' leaving forensic investigators '...very vulnerable.' (Rogers 2005, p. 10) Given that the forensic process is very '...tool centric' and that '...most LE [law enforcement] and CF [computer forensic] practitioners are "tool monkeys" who don't understand what is happening under the hood' (Rogers 2005, pp. 12–13), attacks against such tools are liable to go unnoticed by the investigator. For example, the risks of an investigator not knowing that file timestamps can be modified are obvious. The dangers of non-technical investigators' over-reliance on automated tools is seconded by Hadi (2016), who writes that '[w]ithout proper understanding of the under-laying[*sic*] technology, its[*sic*] just like you're searching for a needle in a haystack!' Some attacks against forensic tools can also be malicious, with the aim of disabling the investigator's device or corrupting their investigation files—an example of such an attack would be the creation of circular directory references, which may trap a program in an infinite loop (Perklin 2012, p. 37), or the use of an XML or zip bomb (Zdzichowski et al. 2015, p. 20).

Counter-forensic techniques that can also be considered under this umbrella are those that aim simply to increase the amount of time and money a forensic investigation will cost, in the hopes that doing so may dissuades it from happening or encourage the offering of some sort of cheaper settlement deal by the investigating agency. Perklin (2012) bases his analysis of anti-forensic techniques around this increasing-cost-calculation approach and presents a number of means of quickly and easily doing so, such as NSRL scrubbing to stop an investigator from excluding non-user-created files on a system and data saturation via the use of a large number of different devices, each of which must then be investigated.

III. Three Techniques to Watch out For

In this section, the three anti-forensic techniques that appear to pose the greatest risk to a forensic investigation are selected. They are: largely-unintentional data saturation; data hiding on the cloud; and the continuing potential for exploitation of the overreliance on a limited number of automated forensic tools. The attempts by the digital forensics community to respond to each technique are then assessed.

A. Data saturation

Consider two trends in the six years since Perklin's presentation: the rise in the number of devices per person, and the decline in police funding (in the UK, at least). Even just two years on from Perklin, Waring (2014) wrote that the number of Internet-connected devices would reach '...an average of 1.7...for every person on the planet' by the end of that year. By 2020, the number is predicted to be 4.3. That was before the rise of the Internet of Things, which has doubtless inflated that number and will continue to do so. When a forensic investigator has to contend with the potential that some trace of their target's illicit activity might be found not just on their PC, personal mobile phone, tablet, work mobile or work PC, but also on their fridge, television, lightbulbs and washing machine, everybody becomes an unintentional data saturator. Meanwhile, BBC News (2018) report that '...between 2010 and 2017 there has been about a 20% cut in police funding in real terms' in the UK. These two trends drastically increase both the ease of causing, and the potential impact to an investigation of, data saturation—which, recall, aims only to make a comprehensive investigation prohibitively costly.

The only obvious solution to this problem is to increase the automation of forensic analysis, allowing one officer to perform the work of several, but this, as can be seen below, raises issues of its own. Regardless, it appears to be the response that the digital forensics community has taken, with the continuing dominance of a small number of simple-to-use automated tools.

B. Data hiding on the cloud

The price of cloud storage has dropped precipitously in recent years, and looks set to continue doing so (Kim 2015). Cloud storage can be used to exploit that most fundamental of obstacles in cyber security—transjurisdictionality. Data stored on Russian servers is likely beyond the reach of US law enforcement, and all it takes to keep it from being intercepted in transit is the use of a HTTPS connection. If a miscreant can ensure that they keep no local copies of their files, and keep their caches clear, they can easily ensure that nothing incriminating can be found on their devices in the event of a seizure. Various trail obfuscation techniques can be further employed to keep even the server's location hidden.

Solutions to this problem are unlikely to emerge from within the digital forensics community, but rather the legal and political ones mentioned in the introduction as being outside of the scope of this essay. Assuming that the suspect in question is successful in keeping their device clean, access to the cloud storage will have to rely on data-sharing and data-extradition treaties between jurisdictions. The only solution available to the forensics community, it appears, is to be able to identify recurring server addresses (when obfuscation has not succeeded in hiding them) and hope that a jury will accept the conjecture that multiple suspects all accessing the same extranational servers, perhaps coinciding with suspicious activity uncovered by (hopefully unmodified) timestamp data, is worthy of suspicion even without knowledge of the content on those servers.

C. The risks of automation

Surely, the only thing worse than having your evidence compromised is not *knowing* that you've had your evidence compromised. In 2005, Rogers highlighted EnCase, FTK and DD[*sic*] as examples of the '...tool centric' approach to digital forensics. Nowadays, EnCase technology '...has been deployed on an estimated 32 million endpoints' and claims organisations such as NASA and Deloitte amongst its customer base (Guidance Software 2017). Meanwhile, university digital forensics courses primarily teach the correct use of tools like FTK, dd and Autopsy (Lancaster University 2017). Additionally, many of these tools are closed source, keeping researchers from assessing them for potential vulnerabilities before the 'bad guys' can. This appears to be a continuing trend, with Autopsy as the only major open source exception to have emerged in the intervening years. It appears that Rogers's warning about the vulnerability of 'Vendor & Tool dependency...' have gone largely unheeded in the 13 years that followed.

IV. The Future (or Lack Thereof) for Digital Forensics

The preceding section does not make optimistic reading. The solution to the unavoidable issue of increasing data saturation, unconsciously on the part of many potential suspects, even in the absence of sustained police cuts is automation, yet automation presents another major issue to the field. Whilst it is likely, due to their continued adoption, that the most commonly-used forensic tools have proven themselves thusfar to be robust and reliable pieces of software, a central tenet of cyber security is that nothing is to be considered secure just because it has been so far. It does not seem unreasonable to imagine that EnCase (to use a random example) may one day be found to be fundamentally flawed in the same way as the past decade's worth of Intel CPUs were by the Spectre vulnerability discovery (Pott 2018). Consider that EnCase has been in use for at last a decade and a half by this point: a DNA manipulation scandal that risks affecting '...more than 10,000 cases...' in the UK only applies to four years' worth of cases (Dearden 2017).

As the issue of tool dependence have been steadfastly ignored over the last decade and a half, there seems to be no reason not to expect that to continue. In light of this, we may expect the increase of automation in order to counter the increase in devices and data needing to be analysed. Recent trends to apply 'artificial intelligence' (a.k.a. neural nets and the like) to such problems of scale can be expected to rear their heads here, leading to the implementation of algorithms that can analyse huge swathes of data and report on (or, more dystopically, act on) what they consider to be suspicious patterns. Whether these results would hold up in court is another issue, especially given recent unease over the unexplainability of 'artificial intelligence' decision-making (Goodman and Flaxman 2018). We may yet see digital forensics at the forefront of the battles over the need to understand your AI's *why* before its intelligence becomes actionable.

However, none of this solves the issue that storing data locally has increasingly become something of an eccentric throwback to a bygone age. This shift to the cloud will serve to make attempts to seize devices with useful data on increasingly unproductive, and the chances of the US and Russian leaderships arranging data extradition deals seem slim to none at present. Perhaps the time for digital forensics is coming to an end, and we shall look back fondly on the days when you could find an illicit file on a suspect's device just by carving an image of their hard drive, in much the same way people now recall chiptune music or socialising on BBSes. That such reminiscing may have to take place cowering in the shadows of our infallible judgement-bot overlords is dampens the appeal somewhat.

REFERENCES

b., david (2012). *How Terrorists Encrypt 4: "Mujahideen Secrets" Software*. URL: http://privacy-pc.com/articles/how-terrorists-encrypt-4-mujahideen-secrets-software.html.

bartitsu59 (2018). 'Breaking Standards'. In: *2600: The Hacker Quarterly* 35.1, pp. 11–12.

BBC News (2018). *Reality Check: Is police funding falling?* URL: http://www.bbc.co.uk/news/uk-43699623.

Berinato, Scott (2007). *The Rise of Anti-Forensics*. URL: https://www.csoonline.com/article/2122329/investigations-forensics/the-rise-of-anti-forensics.html.

Bertrand, Natasha (2015). *The FBI staged a lovers' fight to catch the kingpin of the web's biggest illegal drug marketplace*. URL: http://www.businessinsider.com/ross-ulbricht-will-be-sentenced-soon--heres-how-he-was-arrested-2015-5.

Bishop Fox (2014). *Metasploit Anti-Forensics Project*. URL: https://www.bishopfox.com/resources/tools/other-free-tools/mafia/#timestomp.

Brune, Nate (2016). *Silk Guardian*. URL: https://github.com/NateBrune/silk-guardian.

Dearden, Lizzie (2017). *Convictions in doubt as more than 10,000 cases could be affected by data manipulation at forensics lab*. URL: https://www.independent.co.uk/news/uk/crime/forensic-labs-data-manipulation-criminal-convictions-doubt-randox-testing-services-investigation-a8066966.html.

Goodman, Bryce and Seth Flaxman (2018). 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"'. In: *AI Magazine* 38.3. URL: https://web.archive.org/web/20110722203632/https://www.issa.org/Library/Journals/2007/August/Hartley-Current%20and%20Future%20Threats%20to%20Digital%20Forensics.pdf.

Guidance Software (2017). *Customers*. URL: https://www.guidancesoftware.com/customers.

Hadi, Ali (2016). *Anti-Forensics: Leveraging OS and File System Artifacts*. URL: https://www.ashemery.com/docs/speaks/Ali-AntiForensics.pdf.

Hartley, W. Matthew (2007). 'Current and Future Threats to Digital Forensics'. In: *ISSA Journal* 5.8. URL: https://web.archive.org/web/20110722203632/https://www.issa.org/Library/Journals/2007/August/Hartley-Current%20and%20Future%20Threats%20to%20Digital%20Forensics.pdf.

Hephaest0s (2016). *usbkill*. URL: https://github.com/hephaest0s/usbkill.

Kim, Eugene (2015). *This One Chart Shows The Vicious Price War Going On In Cloud Computing*. URL: http://uk.businessinsider.com/cloud-computing-price-war-in-one-chart-2015-1.

Lancaster University (2017). *17/18: SCC.443: Information System Forensic Investigation [1]*. URL: https://modules.lancaster.ac.uk/course/view.php?id=22182.

Offensive Security (2015). *About the Metasploit Meterpreter*. URL: https://www.offensive-security.com/metasploit-unleashed/about-meterpreter/.

Perklin, Michael (2012). *Anti-Forensics and Anti-Anti-Forensics*. URL: http://www.perklin.ca/~defcon20/perklin_antiforensics.pdf.

Pott, Trevor (2018). *You can't ignore Spectre. Look, it's pressing its nose against your screen*. URL: https://www.theregister.co.uk/2018/01/29/you_cant_ignore_the_spectre_pressing_its_nose_against_your_glass/.

Rogers, Marcus (2005). *Anti-Forensics*. URL: https://www.researchgate.net/publication/268290676_Anti-Forensics_Anti-Forensics.

The Tails Project (2009). *Tails - Privacy for anyone anywhere*. URL: https://tails.boum.org/.

Vaas, Lisa (2014). *Police can demand fingerprints but not passcodes to unlock phones, rules judge*. URL: https://nakedsecurity.sophos.com/2014/11/03/police-can-demand-fingerprints-but-not-passcodes-to-unlock-phones-rules-judge/.

VeraCrypt (2017). *Hidden Volume*. URL: https://www.veracrypt.fr/en/Hidden%20Volume.html.

Waring, Joseph (2014). *Number of devices to hit 4.3 per person by 2020 – report*. URL: https://www.mobileworldlive.com/featured-content/home-banner/connected-devices-to-hit-4-3-per-person-by-2020-report/.

Zdzichowski, Patrycjusz et al. (2015). *Anti-Forensic Study*. URL: https://ccdcoe.org/sites/default/files/multimedia/pdf/AF_with%20intro.pdf.