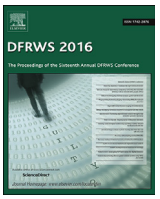
Digital Investigation 18 (2016) S66–S75

Contents lists available at ScienceDirect

Digital Investigation

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



DFRWS USA 2016 — Proceedings of the 16th Annual USA Digital Forensics Research Conference

Anti-forensics: Furthering digital forensic science through a new extended, granular taxonomy

Kevin Conlan*, Ibrahim Baggili, Frank Breitinger

Cyber Forensics Research & Education Group, Tagliatela College of Engineering, ECECS, University of New Haven, 300 Boston Post Rd., West Haven, CT, 06516, United States

Keywords:
Anti-forensics
Computer crime
Digital forensics
Categorical data set
Anti-digital forensics
Anti-forensics taxonomy
Formalizing digital forensics

A B S T R A C T

Anti-forensic tools, techniques and methods are becoming a formidable obstacle for the digital forensic community. Thus, new research initiatives and strategies must be formulated to address this growing problem. In this work we first collect and categorize 308 anti-digital forensic tools to survey the field. We then devise an extended anti-forensic taxonomy to the one proposed by Rogers (2006) in order to create a more comprehensive taxonomy and facilitate linguistic standardization. Our work also takes into consideration anti-forensic activity which utilizes tools that were not originally designed for anti-forensic purposes, but can still be used with malicious intent. This category was labeled as *Possible indications of anti-forensic activity*, as certain software, scenarios, and digital artifacts could indicate anti-forensic activity on a system. We also publicly share our data sets, which includes categorical data on 308 collected anti-forensic tools, as well as 2780 unique hash values related to the installation files of 191 publicly available anti-forensic tools. As part of our analysis, the collected hash set was ran against the National Institute of Standards and Technology's 2016 National Software Reference Library, and only 423 matches were found out of the 2780 hashes. Our findings indicate a need for future endeavors in creating and maintaining exhaustive anti-forensic hash data sets.

Introduction

Digital forensics has garnered much attention over the past ten to fifteen years despite being a relatively nascent scientific field. This attention is due to the sheer amount of data being generated by modern computer systems, which has become an essential source of digital evidence (Geiger, 2005). Scientifically valid and lawful forensic investigation of this digital evidence seeks to uncover and discern its meaning where the evidence must be both reliable, accurate and complete (Harris, 2006).

There has not been, however, as much attention, especially in the form of academic research, towards what can be deemed as “anti-digital forensics”, “anti-forensics”, or “counter-forensics” (Baggili et al., 2012). Although, one could argue that some research such as cryptography may be regarded as anti-forensic, but has not been labeled as such in the literature, thus skewing our perception of the amount of anti-forensic research being conducted. Nonetheless, anti-forensics generally means: attempts to compromise the availability or usefulness of evidence during the forensics process.

Of relevance to anti-forensics is the most recent high profile case of the Federal Bureau of Investigation (FBI) vs. Apple. The FBI had to bypass anti-forensic techniques to acquire the iPhone 5C owned by the San Bernardino County, California government issued to its employee, Syed Rizwan Farook, one of the shooters involved in the December 2015 San Bernardino attack that killed 14 people and injured 22.

* Corresponding author.
E-mail addresses: kconl1@unh.newhaven.edu (K. Conlan), IBaggili@newhaven.edu (I. Baggili), FBreitinger@newhaven.edu (F. Breitinger).
URL: http://www.unhcfreg.com/, http://www.unhcfreg.com/, http://www.FBreitinger.de/

http://dx.doi.org/10.1016/j.diin.2016.04.006

The attackers died, but the iPhone 5C was recovered. It was locked with a four-digit password hindering the forensic acquisition process due to built-in anti-forensic techniques that enforce encryption and auto-wiping the device after multiple unsuccessful password attempts. The legal case where the FBI attempted to order Apple to help with gaining access to data on the device is complex and beyond this paper's scope. Eventually, evidence from iPhone 5C was acquired due to a zero day exploit. The magnitude of this case clearly exemplifies the need of both researchers and practitioners to gain a more comprehensive multidisciplinary understanding of the impact of anti-forensics on the digital forensic community at large.[1]

As shown in the FBI vs. Apple case, anti-forensics makes investigations of digital media more difficult and time-consuming, and thus, more expensive. Users can use anti-forensic tools and techniques to remove, alter, disrupt, or otherwise interfere with evidence of criminal activities on digital systems, similar to how criminals would remove evidence from crime scenes in the physical realm.

Over time, as put forth by Harris (2006), there has been an increase in criminals using anti-forensic methods to counter the forensics process, as well as interfere with the evidence itself. This is evident by the thriving market for software tools that can be considered “anti-forensic” in nature; Geiger (2005) highlighted that there are in fact entire software packages that are designed for anti-digital forensic purposes. These tools can help circumvent widely adopted digital forensic tools and techniques. Moreover, Garfinkel (2007) emphasized that tools that can evade forensic processes are widely accessible.

With that said, Baggili et al. (2012) pointed out the lack of academic research pertaining specifically to anti-forensics, as compared to the more traditional research on digital forensics. In their work they illustrated that only 2% of their data set of 500 digital forensic research papers focused on anti-forensics. This is likely due to the fact that much anti-forensic innovation occurs outside of academic literature.

This problem means that the digital forensics community, including the academic sector, must start considering and formulating mitigation strategies towards the growing problem of anti-digital forensics (Thuen, 2007). Our work aims to address this developing obstacle through an original research initiative by understanding the existing tools and techniques, augmented with devising a taxonomy that embodies anti-digital forensic tools and techniques – thus aiding the systematization of the knowledge in this domain.

The contribution of our work is to provide the following resources to the scientific community:

- A categorical data set on 308 collected tools which can be considered anti-forensic in nature, including important variables for each of the tools, such as anti-forensic capability, developing party, country of origin, etc, (shown in Section 3.2). This data set is publicly available.[2]
- An extended version of the Rogers (2006) anti-computer forensic taxonomy, making it more robust and definitive. This extension includes deeper, more granular specifications within the existing classifications set forth (shown in Fig. 1). Taxonomies are useful for scientific fields; they provide structured classifications within a domain.
- The calculated hash values of 2780 unique installation related files derived from the categorized anti-forensic tools, and an analysis of their presence in the newest 2016 National Software Reference Library (NSRL).[3]

The rest of the paper is organized as follows. In Section Related work we discuss related literature. We then present our methodology in Section Methodology. After that, we present our results and discuss them in Section Results and discussion, and we present our proposed extended anti-digital forensic taxonomy in Fig. 1. Lastly, we present the limitations of our work in Section Limitations and conclude in Section Conclusion and future work.

Related work

Defining anti-digital forensics

As previously mentioned, anti-digital forensics is of growing interest to cybercriminal investigators and academics. However, there is still a lack of a formal definition for the term, as pointed out by Harris (2006). Without a standardized definition that can be agreed upon, practitioners and scientists will be inclined to counter anti-digital forensics with their own definitions, based on their own experiences, which will vary.

With the rise of cybercrime, as well as the amount of software which can be used to interfere with forensic investigations, practitioners must be able to identify the same anti-forensic activity that others have encountered in past experiences. A formal definition of anti-digital forensics, as well as shared terms that are pertinent to anti-digital forensics can facilitate the sharing of knowledge, and allow for better mitigation strategies. It would therefore be appropriate to first highlight how past work defined anti-digital forensics.

Table 1 displays previous definitions of anti-forensics. From these definitions, it can be seen that over time, a majority of the definitions emphasize that anti-forensics can be identified by *any* attempts to alter, disrupt, negate, or in any way interfere with scientifically valid forensic investigations. We note that our adopted definition also encompasses techniques and tools that might have not been intended to be anti-forensic, such as Virtual Private Networks (VPNs) and built-in privacy enhancing technologies such as encryption. We adopt this definition in our research for the creation of our extended taxonomy of anti-digital forensics.

Approaching the problem of anti-digital forensics

In addressing anti-digital forensics, it would be appropriate to become familiar with previous attempts that

[1] https://en.wikipedia.org/wiki/FBI-Apple_encryption_dispute (last accessed 2016-02-10).

[2] Data can be downloaded from http://www.unhcfreg.com by going to Data & Tools section.

[3] http://www.nsrl.nist.gov/ (last accessed 2016-02-10).

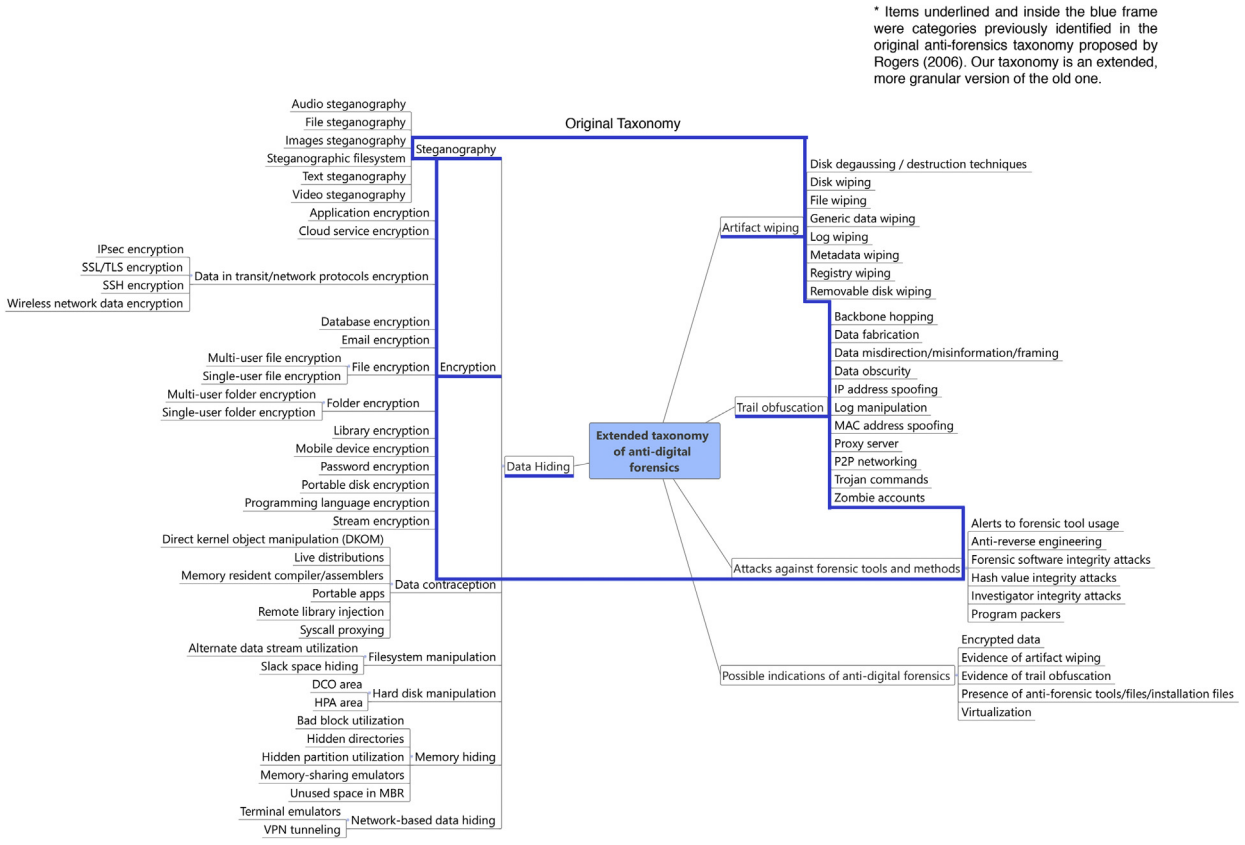


Fig. 1. Extended taxonomy.

Table 1
Previous definitions of anti-forensics.

Author(s)	Definition of anti-forensics provided
Shirani (2002)	Hiding a system intrusion attempt
Peron and Legary (2005)	Attempt to limit the identification, collection, collation and validation of electronic data
Grugq (2005)	Attempting to limit the quantity and quality of forensic evidence
Foster and Liu (2005)	Breaking tools or avoiding detection
Rogers (2006)	Attempts to negatively affect the existence, amount, and/or quality of evidence from a crime scene, or make the examination of evidence difficult or impossible to conduct
Liu and Brown (2006)	Application of the scientific method to digital media in order to invalidate factual information for judicial review
Harris (2006)	Any attempts to compromise the availability or usefulness of evidence to the forensics process
Kessler (2007)	Set of tools, methods, and processes that hinder forensic analysis
Garfinkel (2007)	A growing collection of tools and techniques that frustrate forensic tools, investigations and investigators
Berinato (2007)	An approach to criminal hacking that can be summed like this: make it hard for them to find you and impossible for them to prove they found you
Sremack and Antonov (2007)	The practice of thwarting a proper forensic investigation
Dahbur and Mohammad (2012)	Scientific methods are used to simply frustrate forensics efforts at all forensics stages
Albano et al. (2011)	Methods undertaken in order to thwart the digital investigation process conducted by legitimate forensic investigators
Stamm et al. (2012)	Disguising manipulation fingerprints or falsifying device specific fingerprints inadvertently introduced when a digital file is formed

address the domain as a whole. Several works exist that attempt to define the sub-domain of anti-digital forensics, and point towards future methods that can address the growing problem.

Thuen (2007) conducted a survey that explored the forensic side of computer security, with a specific focus on the methods and ramifications of anti-forensic techniques. The work concluded that a familiarity with counter-forensic techniques at the various stages of a digital forensic investigation is required for practitioners. Pajek and Pimenidis (2009) explored the problem of anti-forensics at various stages of a computer forensic investigation, from both a

theoretical and practical standpoint. Broad descriptions of many anti-forensic tools and their methods was provided by Kessler (2007), using the traditional anti-forensics taxonomy (i.e., data hiding, artifact wiping, trail obfuscation, and attacks on forensic tools themselves).

As put forth by Harris (2006), there are no general or contemporary frameworks with which to analyze and gauge the anti-forensics situation. Wundram et al. (2013) classified and presented both established and newer attacks on digital forensic tools; an argument is made in the work that digital forensic tool testing must develop in parallel with anti-forensic developments. The same view was shared by Anobah et al. (2014) in terms of testing if mobile forensic tools continue to operate in the presence of anti-forensic techniques.

A theoretical framework to define counter-forensics was introduced by Böhme and Kirchner (2013), and is extended to include forensic analysis and authentication requirements. A terminology was then created through a technical survey of counter-forensics against image forensics with a focus on trace suppression and authentication interference; examples and brief evaluations were provided, along with a discussion of relations to other domains in multimedia security.

Sremack and Antonov (2007) highlighted the issue that if anti-forensics succeeds, evidence will fail the Daubert standard.[4] To resolve this challenge, the authors suggested a classification of anti-forensic threats similar to related fields (e.g., digital security). A robust taxonomy was thus created, with the goal of accounting for all types of investigations (i.e., internal, civil, criminal) and threats (i.e., threats to digital evidence, threats to legal process/admissibility). They acknowledge that their taxonomy has limitations in scope, and specify that they would like to expand their taxonomy in future work.

Highlighted by Brand (2007) was the growing sophistication of anti-forensic techniques used by malicious software (malware). The work discussed the exhaustive list of anti-forensic techniques utilized by malware (e.g., obfuscation, anti-disassembly, encrypted and compressed data, data destruction, anti-debugging, etc.). The work also addressed the fact that automated detection and classification work is progressing in the field, which includes statistical structures such as assembly instructions, system calls, system dependence graphs, and classification through machine learning.

Attempts at anti-digital forensic classification and characterization

In this section we review previous attempts at the classification, identification, and characterization of anti-digital forensic tools. These works were conducted with the purpose of creating structured resources for the digital forensic community.

While there are a few general groupings of anti-forensic methods that aid in the analysis of anti-digital forensics (Harris, 2006), there is yet to be identifiable groupings of anti-digital forensic software based on an extensive analysis of widely available tools. As noted, there have been previous proposals with regards to anti-forensics; Peron and Legary (2005) divide anti-forensic techniques into four categories: an adversary can destroy, hide, manipulate or prevent the creation of evidence. Other categories were proposed by Rogers (2006): data hiding, artifact wiping, trail obfuscation and attacks against both the forensic process and forensic tools. This taxonomy is widely adopted in digital forensics research. While these past attempts are useful in classifying and categorizing anti-forensic methods, to the best of our knowledge, there is no current research that fully accounts for the categorization of anti-forensic software. In the few cases that do exist, they often focus only on segments of anti-forensics (Harris, 2006), or only on a select number of tools.

Garfinkel (2007) presented a survey of contemporary anti-forensic tools (e.g., Timestomp, Transmogrify, etc.).[5] Discussed were approaches for direct attacks against forensic tools focusing especially on the exploitation of software bugs within forensic tools themselves (e.g., data validation failures, denial of service attacks, fragile heuristics, etc.). Also evaluated was the effectiveness of anti-forensic tools against traditional forensic tools. Lastly, strategies towards anti-forensic detection and countermeasures were outlined.

The challenges of anti-forensics were also highlighted by Dahbur and Mohammad (2012). They provided a classification of anti-forensic mechanisms, tools, and techniques, and evaluated their effectiveness. Challenges of countermeasures against anti-forensics, along with a set of recommendations for future research were also discussed.

Addressed by Smith (2007) was the issue that disk-avoiding anti-forensic tools are growing in use, and needed to be described and categorized. This work built on existing categories to classify anti-forensic methods, and provided information to facilitate the understanding of contemporary trends in anti-forensics.

Lastly, Blunden (2009) examined approaches that an investigator may use towards persistent rootkits (e.g. defense in depth, static analysis of an unknown executable, etc.), as well as identified the anti-forensic possibilities that rootkits may employ.

Attempts at the detection and indication of anti-digital forensic tool usage

A theoretical approach to digital investigations, in which the investigation process is at all times aware of the possibility of anti-forensic attacks was presented by Rekhis and Boudriga (2012). The work created an investigated system scenario, the deployed security solution, and a library of anti-forensic attacks that are used against the system, with the resulting evidence being collected. Then, an inference system was proposed to mitigate anti-forensic attacks. Potential scenarios were then generated from the anti-

[4] https://en.wikipedia.org/wiki/Daubert_standard (last accessed 2016-02-10).

[5] http://www.metasploit.com/ (last accessed 2016-02-10).

forensic traces to provide models of anti-forensic actions that can occur during digital investigations.

Other past work by Geiger (2005) focused on the analysis of six anti-forensic tools (Window Washer, CyberScrub Professional, SecureClean, Evidence Eliminator, and Acronis Privacy Expert). The analysis, which consisted of observation of the tools' performance by examining the disk images with Forensic Tool Kit (FTK) concluded that significant shortfalls were discovered in each anti-forensic tool examined (e.g., incomplete wiping of unallocated space), that could permit the recovery of evidentiary data. In addition, it was pointed out that each tool had a "fingerprint" (i.e., operational signature) that can be used to identify the application's usage.

A forensic tool capable of extracting and analyzing forensic data from the file system of a created "honeypot" (i.e., a system intended to be used as a testing ground for anti-forensic tools), was introduced by Fairbanks et al. (2007). TimeKeeper, the journal-monitoring prototype developed was able to detect anti-forensic attempts, and capture previously unavailable forensic information. This information can be used to facilitate system recovery, research attack techniques, provide insight into attacker motives, and criminal investigations.

An active project titled Indicators of Anti-Forensics (IOAF) serves as an effort towards automatic anti-forensic trace detection using signature based methods was presented by James (2014). The developed program uses parsing modules to extract file metadata and registry key information from an investigated system. Predefined signatures previously stored in a database are queried for each extracted object, to see if the object consistently corresponds to a signature.

It should be clear that the idea of automatic detection and indication of anti-digital forensic activity is well worth further research and initiatives. Our work aims to build upon the endeavors of these previous works while forming a comprehensive taxonomy to improve the digital forensics body of knowledge.

Methodology

We used a methodical approach in order to create the data set of anti-forensic tools. The methodology we executed included the following overarching steps:

Data set creation: An extensive web search was conducted for software considered to be anti-forensic and the software was identified and downloaded. More details on this step are provided in Section Preliminary research and collection of anti-digital forensic tools. A data set of anti-forensic tools was then created from the downloaded applications.

Data set organization: In this step, we assigned descriptive and identifying variables (e.g. anti-forensic category, subcategories, developer, first release, etc.) that pertained to each tool.

Data set analysis: We analyzed the collected and organized data from the prior step to produce summary information.

Hashing: As many of the installation (e.g. .exe, .zip., .msi) files of the programs that were accessible ($n = 191$) were then hashed so that we could store each anti-forensic tool's computed hash value. Additionally, after some of the .zip files were unzipped, more of the resulting files ($n = 2780$ unique hashes) were also hashed.

Data set comparison with NSRL: The data set of hashes were then searched for in the NSRL database.

Extended taxonomy creation: Based on previous work, as well as the thorough categorization of each of the downloaded anti-forensic tools, our proposed extended taxonomy was conceived, which is shown in Fig. 1.

Preliminary research and collection of anti-digital forensic tools

The tools identified in this work were collected through extensive and exhaustive web searches as well as the searching of open-source repositories.

Keywords used in these web searches included "Best anti-digital forensics tools", "Most popular anti-digital forensic tools", "Popular anti-digital forensic software", "Best anti-computer forensic tools" etc. "Best", "Most Popular", and similar keywords were implemented heavily to systematically find a sample of anti-digital forensic tools that would best reflect the tools used most often in anti-digital forensic activities.

Categorical data set of 308 anti-forensic tools

For each tool we determined the following identifying parameters (in some cases, this data was simply unavailable):

I Anti-digital forensic tool name
II Anti-forensic category I
III Anti-forensic subcategory II
IV Anti-forensic subcategory III
V Web address
VI Operating system/platform
VII Country of origin
VIII Developer
IX Organizational type
X Date of first release
XI Latest version
XII License
XIII Currently Maintained? (yes/no)
XIV Programming language of development

The created categorical data set, which contains all of these documented variables, can be downloaded from http://www.unhcfreg.com under *Data & Tools*. We hope that members of the digital forensic community would benefit from this data set.

Extended taxonomy of anti-digital forensics

We designed an extended version of the widely accepted taxonomy of anti-forensics, previously devised by

Rogers (2006). Our goal was to provide a more specified, granular organized body of knowledge that could better characterize the growing domain of anti-digital forensics. Below is the original taxonomy of anti-forensics:

Original *anti-forensics taxonomy* proposed by Rogers (2006).

- Data hiding
  - Encryption
  - Steganography
  - Other forms of data hiding
- Artifact wiping
  - Disk cleaning utilities
  - File wiping
  - Disk degaussing/destruction techniques
- Trail obfuscation
- Attacks against computer forensic tools and processes

While the original taxonomy of anti-forensics provided the pillar to our extended taxonomy, expanding it to a more granular level was a necessary step due to the growing expansion and complexity of the domain. An extended taxonomy of the domain must thus reflect this trend and provide an evolving structure of anti-forensics.

Creating a taxonomy is as much about the tools and techniques as it is about the users. In digital forensics, practitioners will always face varying challenges, and thus this taxonomy was designed to capture as many possible situations a potential practitioner may encounter. The taxonomy was developed as a granular, wide-scoped classification scheme, as opposed to a controlled vocabulary or hierarchical structure design. Breadth and depth were both of interest in the design process. We are open to future modifications and alterations.

Hash value generation and comparison to NSRL

The MD5, SHA1, SHA-256, CRC32, SHA-512 and SHA-384 hash values of 2780 unique installation-related files (Windows .exe files preferred, in cases of multiple platforms) of the anti-digital forensic tools were generated using the HashMyFiles,[6] a utility that allows a user to generate the aforementioned hashes of multiple files on a system. These hash values can be downloaded from http://www.unhcfreg.com under *Data & Tools*. Not all 308 installation files were accessed and downloaded due to a number of the tools being proprietary. The 2780 unique hash values were then compared to the newest 2016 NSRL hash data set.

[6] http://www.nirsoft.net/utils/hash_my_files.html (last accessed 2016-02-10).

Results and discussion

Extended taxonomy of anti-digital forensics

The taxonomy was characterized by the need to identify all aspects of anti-digital forensics that may exist on a system, as it is crucial that first responders, investigators, and researchers understand what forms of anti-forensic activity may exist on a system or network. Our resultant taxonomy is depicted in Fig. 1. We discuss and present the results of our efforts in the sections that follow.

Data hiding

With regards to *data hiding*, the original taxonomy included *encryption*, *steganography* and *other forms of data hiding* as its sub categories. Supplemental in our extended taxonomy is the identification of important subcategories, as well as an articulation of the *other forms of data hiding* category (i.e. *data contraception*, *filesystem manipulation*, *hard disk manipulation*, *memory hiding*, and *network-based hiding*, respectively).

Encrypting data on a machine is the quickest method to prevent access to stored data. The extended taxonomy expounded largely upon the *encryption* category, due to a multitude of encryption techniques that can take place on a system, or the peripherals of a system, on different data types. It is important to differentiate between these forms of encryption to better prepare practitioners with what they may encounter on the field, or researchers for what they may encounter during their research.

Disk encryption was identified as a vital category; the exponential growth of storage media on systems has led to the development of tools that can encrypt the full volume of a hard drive. Additionally with the overall growth in user produced data we see a paralleled increase in the usage of databases to store the data; *database encryption* is thus another form of *data hiding*.

It was also deemed necessary to identify *folder encryption* (both single user and multi-user), as well as *file encryption* (both single and multi-user). Instances of *email encryption*, *application encryption*, *cloud service encryption*, *portable disk encryption*, and *mobile device encryption* are increasing, and thus included in our proposed extended taxonomy. This is evident due to the large amounts of tools identified for these encryption purposes.

Of interest were also tools that allow for or enhance *data-in-transit/network protocol encryption*, as encrypted network traffic can greatly hinder the network forensic process. These types of tools are gaining popularity, especially by users with privacy concerns. The extended taxonomy thus included these forms of encryption.

The area of *steganography* was not greatly adjusted from what was identified in the original taxonomy; all that was modified was the identification of the four forms of *steganography* that may occur digitally: *video*, *audio*, *images*, and *text*. What was included, however, was *filesystem steganography*, as this was an emerging steganography type

identified in our research (e.g., Magikfs[7] – a steganographic file system for Linux).

A challenge we faced in the creation of our extended taxonomy was the categorization of "encryption suites", i.e., software suites that provide numerous cryptographic features. Future work could involve a methodology developed to successfully categorize these tools.

The following subcategories of *data hiding* were identified in the research as an articulation of the *other forms of data hiding* category from the original taxonomy; *Data contraception* was introduced as a new category within *data hiding*. It can be defined as anti-forensic activities that produces little to no digital trace evidence, thus having great potential to defeat forensic investigations. This category includes *syscall proxying*, *remote library injections*, *direct kernel object manipulation (DKOM)*, and *portable apps*.

Added to the category of *data hiding* were the subcategories of *filesystem manipulation* and *hard disk manipulation*. These subcategories of *data hiding* were deemed part of the new taxonomy due cases existing in which items are hidden on systems through the deliberate manipulation of filesystems and hard disks.

Memory hiding was seen as another form of *data hiding* for inclusion into the extended taxonomy. Utilizing a system's memory to hide data is potentially devastating to forensic investigations due to the high level of training required to defeat these attempts.

Using a network to hide data must also be considered in digital forensic investigations; such anti-forensic activities would include Virtual Private Network (VPN) tunneling and the use of proxy servers and terminal emulators. Thus, a category identified as *network-based hiding* was established as well.

Artifact wiping

The *artifact wiping* category was extended to include the different types of "wiping" (i.e., the deliberate destruction of data that could be used as evidence) which was not fully affirmed in the original taxonomy.

Based on our analysis of the collected data set, there are vast amounts of software that can be used to wipe different forms of data: *files*, *disks*, *removable/portable disks*, *logs*, *metadata*, and *registries*. A large number of tools performed more than one form of wiping; they were classified as *generic data wiping*.

Artifact wiping, alongside *data hiding*, based on the large amount of tools that were found for each of those respective categories, shown in Table 2, is likely to remain a popular form of evidence destruction and manipulation, based on the considerable number of tools that can perform these functions.

Trail obfuscation

Trail obfuscation, i.e., the deliberate activity to disorient and divert a forensic investigation on a digital system or network included what Rogers (2006) had originally specified (i.e. *log cleaners*, *different forms of spoofing*, *misinformation*, *backbone hopping*, *zombie accounts*, and *trojan commands*). Added was *P2P networking*, which is a popular form of *trail obfuscation*; P2P software has been noted as easily being misused for criminal activity (e.g., copyright infringement, malware dissemination, etc.).

It is likely that *trail obfuscation* activities will increase, alongside the rise in cybercrime, given that criminals will naturally attempt to mitigate "fingerprints" that provide evidence of their crimes, and might employ the tools and techniques at their disposal to confuse and mislead investigators.

Attacks against forensic tools and processes

Direct attacks against the software used to investigate digital media, as well as against the processes employed during these investigations have the potential to be the most devastating anti-digital forensic activity to an investigation.

This section of the original taxonomy was extended to include key areas that have not been formerly addressed; *program packers*, *anti-reverse engineering*, and *attacks against the integrity of investigating parties*.

Program packers can be used against forensic tools and processes by hindering the forensic process as they allow for the compression, and, in some cases with certain program packers, the encryption of data.

Anti-reverse engineering also needs to be considered in a taxonomy of anti-forensics as criminals will naturally attempt to deflect attempts by investigators to extract information from suspected artifacts. While *program packers* are technically a form of anti-reverse engineering, the two were separately categorized; as more specified forms of

Table 2
Number of tools per category.

Data hiding (153)	**Artifact wiping (113)**
Encryption (127)	File wiping (27)
Disk (46)	Disk wiping (28)
Email (9)	Removable disk wiping (3)
File (19)	Generic data wiping (23)
Filesystem (9)	Registry wiping (29)
Data-in-transit/network protocol (14)	Disk degaussing/destruction techniques (2)
Password (7)	Metadata wiping (1)
Mobile device (8)	**Trail obfuscation (38)**
Portable drive (5)	P2P networking (33)
Application (2)	IP address spoofing (1)
Cloud service (3)	Data fabrication (2)
Programming language (2)	Data misdirection/ misinformation (1)
Library (3)	Proxy server (1)
Steganography (16)	**Attacks against forensic tools & processes (4)**
Image (4)	Program packers (4)
Text (9)	
Filesystem (3)	
Data contraception (1)	
Syscall proxying (1)	
Filesystem manipulation (2)	
Slack-space hiding (2)	
Memory hiding (1)	
Memory-sharing emulators (1)	
Network-based data hiding (6)	
Terminal emulators (3)	
VPN tunneling (3)	

[7] http://magikfs.sourceforge.net/ (last accessed 2016-02-10).

anti-reverse engineering are identified, the taxonomy will be updated to include them.

A newer concept added to the taxonomy was *attacks against the integrity of forensic investigators*, which was added to the taxonomy, given that it is within the realm of possibility that criminals may attempt a smear campaign, frame, or perform other malicious activities to attack the integrity of organizations tasked with the legal investigation of computer criminal cases.

Possible indications of anti-digital forensic activity

Investigators are now forced to develop a "situational awareness" when it comes to digital forensics, due to the sheer volume of criminal cases, increasing at an exponential rate. Being able to quickly and efficiently locate any *potential* of anti-digital forensic activity is now a requirement for successful investigations.

What needs to be understood is that there may be activity or software on a system that may not appear to be criminal or illicit, but a practitioner will still need to be open to the fact that there are tools that can be used for anti-forensic purposes, despite not being created with anti-forensics in mind.

Possible indications of anti-digital forensic activity can be rather straightforward; the very presence of anti-forensic software and files can obviously be giveaways. The installation files for cryptographic software, for example, should offer an investigator the "low-hanging fruit" indication that data may be encrypted on a system.

An example of a possible indication of anti-digital forensic activity would be an encrypted virtual machine on a host system. Such a scenario would have the benefit of hiding all application, temporary file storage, and other items in a virtual environment, and would be stored as a single file, which later can be deleted.

For anti-forensics to be better understood and developed, as well as that of *anti-anti-forensics*, i.e., the mitigation and solution strategies arrayed against anti-forensics, this category should be further articulated and expounded upon by the digital forensic community.

Comparing anti-forensic tool hashes to the NSRL

A Python script was written to acquire 2780 unique MD5 and SHA1 hash values of the anti-forensic tool installation-related files, and was compared against the newest 2016 Reference Data Set (RDS).[8] Out of the 2780 unique hashes, 423 distinct hashes were found to be in the most current 2016 version of the RDS at the time of writing this paper. The hash data set was also compared against the hashes of NIST provided operating systems installed on virtual machines, with no further matches beyond the 423 found in the RDS. This was an important finding, as the unmatched hashes would be examples of *Presence of anti-forensics tools/files/installation files* under the category of *Possible indications of anti-digital forensics*, as represented in our newly expanded taxonomy of anti-digital forensics. This finding warrants further research in creating known hash sets related to anti-forensic tools.

[8] This was conducted with the help of Douglas White at NIST.

Quantitative analysis of the data set of anti-forensic tools

Table 2 displays the number of tools found at various levels of the taxonomy, i.e., the specific forms of *data hiding*, *artifact wiping*, *trail obfuscation*, and *attacks against forensic tools and processes*. These tools and their respective anti-forensic capabilities are recorded within our categorical data set, which can be downloaded from http://www.unhcfreg.com under *Data & Tools*.

To note, the majority of tools categorized were of the primary *data hiding* and *artifact wiping* categories, and within those, *file, filesystem, disk, and email encryption* made up the majority of *data hiding* tools, and *file, disk, generic data, and registry wiping* made up the majority of *artifact wiping* tools found in our data set. This should be expected, as these tools have historically been prevalent in computing.

Additionally, with regards to *trail obfuscation*, a large number of *P2P networking* software was found during the research. This finding, as well as that of the previously mentioned large quantities of encryption and data wiping tools found, should be of interest to the forensic community.

Fig. 2 displays the identified instances of operating systems and platforms for which specific anti-digital forensic tools were developed. It should not come as a surprise that *Windows*, *Macintosh*, and *Linux* made up the vast majority of identified operating systems in the research, considering that they are the most prominent operating systems.

Multi-platform/unspecified was a category with a high number, reflecting that numerous tools had versions for more than one unspecified operating system, or no operating system was specified by the developing party. Some of these tools were simply described as "cross-platform", or no specific platforms were identifiable during the research.

The presence of anti-digital forensic tools for mobile operating systems, e.g., Windows Phone, iOS, and Android, in the view of the researchers, is a component of anti-forensic research that will require more attention, due to the growing ubiquity of mobile devices in our society, and thus the paralleled growth in their adoption for criminal activity.

We also compiled a list of the countries of origin of a number of the anti-digital forensic tools, as well as the developing party's organizational type (not all, due to a number of the sources for the tools not alluding to their respective countries of origins or a specific organization type).

With regards to Fig. 3, it is of interest that a large number of anti-digital forensic tools were identified as originating from Germany and Finland. It would be interesting to further examine why these two countries were reflected in the results, as opposed to say, Russia and China. Additionally, further research on how the anti-digital forensic landscape has changed over time would be interesting, by exploring if the amounts of tools from each of

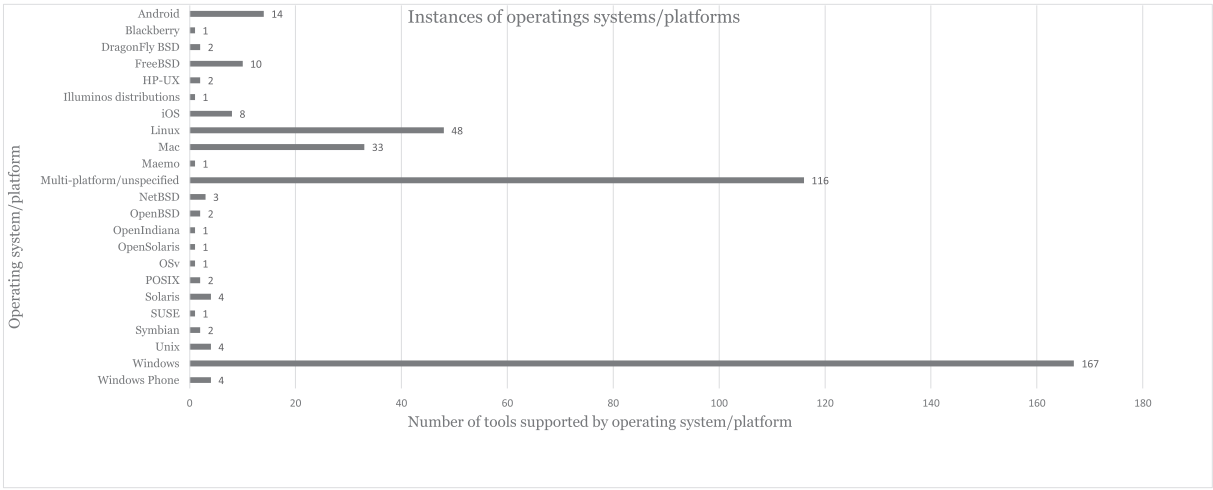


Fig. 2. Identified operating systems and platforms.

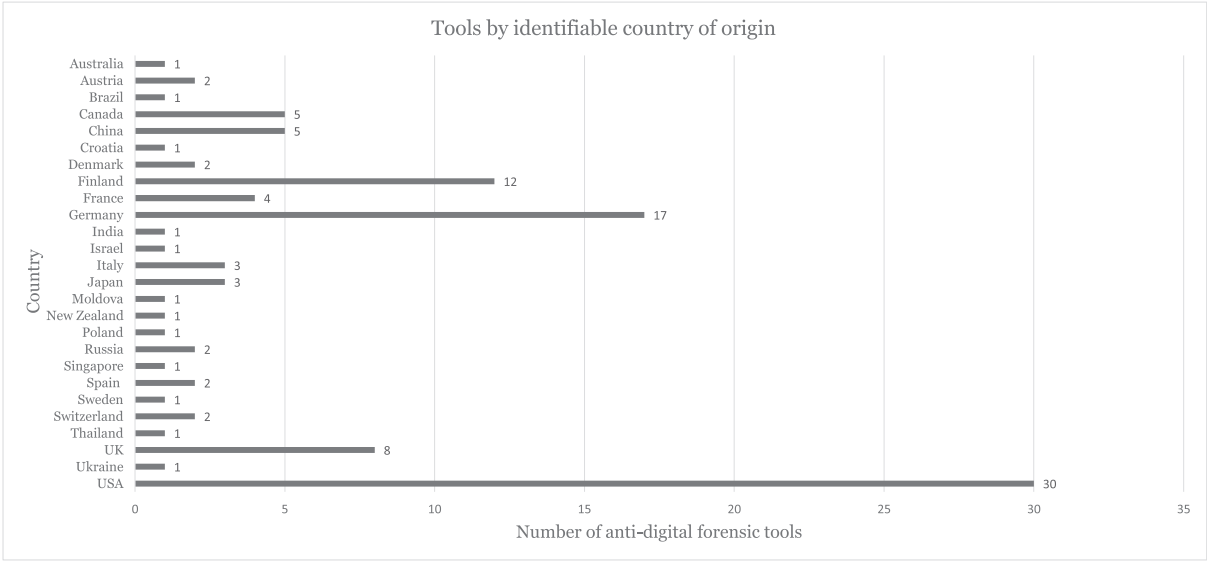


Fig. 3. Identified countries of origin of a number of the anti-forensic tools.

these countries of origins have experienced any dramatic increases or decreases.

155 anti-digital forensic tools from the data set had data on their organizational type of the developing entity. It was found that 46 of the tools were developed by non-commercial entities. The non-commercial organizations most certainly developed the anti-digital forensic tools that were designed with "privacy" and "open-source" philosophies in mind, e.g., P2P networking, steganography, and encryption tools. 98 of the 155 tools were created by commercial entities. This is to be expected, due to the large number of *data hiding* and *artifact wiping* software that is certainly being developed by commercial enterprises for corporate and personal use. A much smaller amount of tools were found to be created by military and academic organizations (one and ten, respectively).

Further research into the origins of anti-digital forensic tools and their developing parties, alongside other variables, could help facilitate the continued mitigation efforts of the anti-digital forensic problem.

Limitations

A considerable limitation of this work was that the sheer number of software tools that may be considered "anti-forensic" in nature is incredibly vast and continuously growing, and it is extremely difficult to determine the entire scope of the domain. This is, however, not as much a "limitation" as it is an opportunity for future research endeavors.

An additional limitation of the categorical data set and extended taxonomy was the failure to fully cover newer, yet hard to articulate fields of anti-digital forensics. For example, not only is the domain of "cloud service forensics"

still nascent and not fully defined, the field of "anti-cloud service forensics" needs research attention as well.

Conclusion and future work

The goal of this work was the following: to create a categorical data set that would be useful to the digital forensic community through the collection and organization of 308 anti-forensic tools, including specification of their possible anti-forensic usages on a system. Another goal was to create an extended classification of the original anti-forensics taxonomy, with the purpose of encapsulating the possibilities within the domain of anti-forensics.

Future work could include expanding the scope of the categorical data set to include more tools, of which there are many. Based on the results, collecting identifiable information on anti-digital forensic tools, and then formulating it into an accessible body of knowledge has the potential to benefit and assist digital forensic practitioners.

With the migration of the Internet into devices not usually connected to networks, i.e., the "internet-of-things", anti-digital forensic software and tools will follow this digital migration. We can only assume that anti-digital forensic activity will reach these newly-connected devices. Expanding the taxonomy to include not just the classifications of anti-digital forensic tools and methods, but also the forms of digital devices that such activity could occur on would be beneficial to researchers and practitioners as well.

It would be interesting to see a similar methodology, i.e., that of the collection, identification, and categorization of software tools extended to other fields of the information assurance domain. For example, resources established that could identify and centralize information on software tools used by malicious hackers to penetrate systems and networks could prove useful to the forensic and security communities alike.

Lastly, ways of automating the process of classifying anti-forensic tools may be of interest to scientists working in computational linguistics, as this could possibly be done by parsing metadata of tools online and leveraging machine learning. Continued research on this work, as well as in the domain in general, would prove beneficial to mitigating the growing problem of anti-digital forensics.

Acknowledgments

We would like to thank Sidharth S. Nandury and Mohammad M. Hasan for their participation in the tool collection and hashing efforts – specifically for their analysis of the tools. We would also like to thank Douglas White at NIST for helping us run our hashes against the newest 2016 RDS hash set.

References

Albano P, Castiglione A, Cattaneo G, De Santis A. A novel anti-forensics technique for the android os. In: Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on. IEEE; 2011. p. 380–5.
Anobah M, Saleem S, Popov O. Testing framework for mobile device forensics tools. J Digit Forensics, Secur Law JDFSL 2014;9:221.
Baggili I, BaAbdallah A, Al-Safi D, Marrington A. Research trends in digital forensic science: an empirical analysis of published research. Digit Forensics Cyber Crime 2012:144–57 [Springer].
Berinato S. The rise of anti forensics. 2007.
Blunden B. Anti-forensics: the rootkit connection. In: Black Hat USA 2009 Conference Proceedings; 2009. p. 10 [Citeseer].
Böhme R, Kirchner M. Counter-forensics: attacking image forensics. Digit Image Forensics 2013:327–66 [Springer].
Brand M. Forensic analysis avoidance techniques of malware. 2007.
Dahbur K, Mohammad B. Toward understanding the challenges and countermeasures in computer anti-forensics. Cloud Comput Adv Des Implement Technol 2012:176.
Fairbanks KD, Lee CP, Xia YH, Owen III HL. Timekeeper: a metadata archiving method for honeypot forensics. In: Information Assurance and Security Workshop, 2007. IAW'07. IEEE SMC. IEEE; 2007. p. 114–8.
Foster J, Liu V. Catch me if you can…. Blackhat Briefings; 2005.
Garfinkel S. Anti-forensics: techniques, detection and countermeasures. In: 2nd International Conference on i-Warfare and Security; 2007. p. 77.
Geiger M. Evaluating commercial counter-forensic tools. DFRWS; 2005.
Grugq. The art of defiling: defeating forensic analysis. Blackhat briefings; 2005.
Harris R. Arriving at an anti-forensics consensus: examining how to define and control the anti-forensics problem. Digit Investig 2006;3:44–9.
James JI. Indicators of anti-forensics. 2014.
Kessler GC. Anti-forensics and the digital investigator. In: Australian Digital Forensics Conference; 2007. p. 1.
Liu V, Brown F. Bleeding-edge anti-forensics. 2006.
Pajek P, Pimenidis E. Computer anti-forensics methods and their impact on computer forensic investigation. Glob Secur Saf Sustain 2009:145–55 [Springer].
Peron CS, Legary M. Digital anti-forensics: emerging trends in data transformation techniques. 2005.
Rekhis S, Boudriga N. A system for formal digital forensic investigation aware of anti-forensic attacks. Inf Forensics Secur IEEE Trans 2012;7:635–50.
Rogers M. Anti-forensics: the coming wave in digital forensics. Retrieved September, 7. 2006.
Shirani B. Anti-forensics. High Technology Crime Investigation Association; 2002.
Smith A. Describing and categorizing disk-avoiding anti-forensics tools. J Digit Forensic Pract 2007;1:309–13.
Sremack JC, Antonov AV. Taxonomy of anti-computer forensics threats. IMF 2007:103–12.
Stamm MC, Lin WS, Liu K. Forensics vs. anti-forensics: a decision and game theoretic framework. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE; 2012. p. 1749–52.
Thuen C. Understanding counter-forensics to ensure a successful investigation. 2007.
Wundram M, Freiling FC, Moch C. Anti-forensics: the next step in digital forensics tool testing. In: 2013 Seventh International Conference on IT Security Incident Management and IT forensics; 2013. p. 83–97 [IEEE].