Chapter 24

APPLYING FILTER CLUSTERS TO REDUCE SEARCH STATE SPACE

Jill Slay and Kris Jorgensen

Abstract Computer forensic tools must be both accurate and reliable so as not to miss vital evidence. While many investigations are conducted in sophisticated digital forensic laboratories, there is an increasing need to develop tools and techniques that could permit preliminary investigations to be carried out in the field. Pre-filtering electronic data in the field, before a computer is brought back to a laboratory for full investigation, can save valuable time. Filtering can also speed up in-house investigations by reducing search space size.

This paper discusses the application of automated tools based on filters. In addition to helping reduce the search space, filters can support specific tasks such as locating and identifying encryption software and hidden, encrypted or compressed files. Filters may be used to automate tedious examinations of temporary Internet files, Windows directories or illicit images. Also, filters can facilitate customized searches based on patterns encountered in investigations of common cases.

Keywords: Forensic tools, field investigations, filter design, filter clusters

1. Introduction

One of the main challenges in computer forensic investigations is the increasing capacity of storage media [1, 7]. Even small electronic devices can hold thousands of documents, images and other files. Because of the capacity of these devices, it is necessary to devise techniques that can partition the search space into smaller, more easily managed areas. Partitioning the search space makes it more feasible to locate specific data, whether or not attempts have been made to conceal it.

Extracting digital evidence from storage devices requires the use of forensically sound tools and techniques [7, 8]. In contrast to normal applications where computers are used to significantly decrease the time

needed to reach an objective, the priority in a computer forensic investigation is accuracy rather than speed [3]. It is often impractical to do a complete examination of all the devices encountered in an investigation. This problem is further aggravated by rapid increases in storage capacity [4]. Moreover, achieving the desired level of accuracy is particularly difficult given the exigencies in field investigations.

Several techniques exist for concealing data in electronic devices, ranging from simple techniques such as hiding a file in a large collection of other files to advanced techniques involving alternate data streams, encryption and steganography [2]. Given the availability of free software for data concealment, it is a distinct possibility that incriminating information may have a high level of protection applied to it.

A forensic examiner must not only have the tools for conducting investigations, but must also have a solid grounding in and knowledge of operating systems, file systems, file formats and information storage techniques [9]. With constantly changing technology, examiners must be kept up to date by training or self-learning. Even then, it is not possible for one individual to be an expert in all areas. This highlights the need for a team approach, with each member of the team having expertise in specific areas.

Forensic tools are available for performing tasks such as searching for file types, detecting encryption, recovering deleted files, locating concealed data and tracing email. Some advanced tools allow scripts to be written to tailor their operations to specific investigations.

While many investigations are conducted in sophisticated digital forensic laboratories, there is an increasing need to develop tools and techniques that could permit preliminary investigations to be carried out in the field, especially in remote locations and rural areas. Pre-filtering electronic data in the field, before a computer is brought back to a laboratory for full investigation, can save valuable time. Filtering can also speed up in-house investigations by reducing search space size.

This paper discusses the application of automated tools based on filters. In addition to helping reduce the search space, filters can support specific tasks such as locating and identifying encryption software, and hidden, encrypted or compressed files. Filters may be used to automate tedious examinations of temporary Internet files, Windows directories or illicit images. Also, filters can facilitate customized searches based on patterns encountered in investigations of common cases.

2. Exhaustive Search vs. Intelligent Search

Due to the massive capacity of modern storage media, the actual amount of incriminating data in many cases is a small percentage of the total data contained on seized media. Two approaches exist for locating evidence in a mass storage device, the "Tourist Approach" and the "Divide and Conquer Approach."

The Tourist Approach involves examining every bit of data on the device. While this approach is very time consuming and is often infeasible [12], notwithstanding human error, it will yield the data sought if such data, in fact, exists on the device. Of course, it may not be possible recover data that is protected using cryptography or steganography [5].

The Divide and Conquer Approach applies intelligence to partition the data into pre-established domains that are analyzed individually. The partitioning process does not, and probably should not, entail just one pass as further partitions could be sought that logically group the data into the smallest sets possible. Once the data is sorted into logical groupings, more specific approaches can be applied to each data set. Examples include automatically detecting images based on the presence of skin-colored pixels or culling files that contain specific keywords and phrases. By using a collection of divide-and-conquer partitioning strategies, the original data can be sifted automatically, enabling an examiner to significantly reduce search time by focusing on specific data sets.

3. Determining What to Exclude

One method to reduce the search space is to determine the data that is superfluous to the investigation, and eliminate this incidental data at the outset.

The level of skill of a suspect may decide the extent to which data is eliminated. The skill levels, and therefore the types of filters, range from a barely computer-literate user to an expert who can hide data in operating system files without impeding normal system use. In the case of a suspect with average computer skills, an investigator could ignore operating system files and scan user-created files. This can be accomplished by using filters designed to "bubble up" user data that is relevant to the investigation.

Often, the only data that actually pertains to a case is the data that the user has explicitly placed on the storage device. Most of the other incidental data can be safely eliminated without fear of ignoring crucial evidence. A prime example is incidental data placed on the device by the operating system to permit its operation. For example, in the case of

Microsoft Windows, the system dependent DLLs and executables need not be considered and can be eliminated during an initial search.

Some individuals may be competent enough to conceal data so that it cannot be found except by exceptional means, e.g., using second harmonic magneto resistive microscopy [10]. Therefore, all storage media from suspects with technical expertise should be immediately seized and sent to a laboratory for a complete investigation

During a thorough laboratory-based forensic investigation, the process of elimination of inconsequential files is done in a controlled manner. Toolkits utilizing hash sets allow for low-level matching of files to known signatures [11]. These hash sets are available from commercial vendors, law enforcement agencies and government organizations, e.g., the U.S. National Institute of Standards and Technology [11].

4. Designing Filters

To make the process of locating relevant information more efficient, tools must be developed that can narrow the search state space of a device being scrutinized. These tools are the digital forensic equivalent of filters in the real world.

A filter partitions data based on specific criteria. By combining multiple filters into filter clusters, it is possible to narrow down the target device information and attach a relevant suspicion level to data, which in effect ranks the data from most likely to be relevant to least likely to be relevant. Filters should be used in order of increasing specificity: the first filter should partition data as useful and not useful, and remaining filters should become more specific as the desired outcome is approached.

Filters need not be designed to produce wide partitions of the search space. Instead, filters can target individual files or programs, eliminating the need to have different search programs or strategies. In these instances, however, the filter loses its ability to be included as part of a logical formula. This is not necessarily a bad option as it precludes the use of more than one tool and allows the same design paradigm to be applied to search one specific case or to divide the space for other filters or human examination.

No two suspects are likely to hide data in exactly the same way. Therefore, given the myriad ways available for concealing data, a single filter type cannot handle every case. Instead, multiple filters should applied in series, each designed to give priority to a different aspect. This implies an iterative development process in which filters that are found to be effective are added to the toolkit. Also, if one filter cluster fails to provide meaningful results, others can be applied to the data. If the

filter clusters used in a field investigation do not provide useful information, the target device should be sent to a laboratory for a complete examination. Data recovered during the examination will have the side effect of producing useful filter clusters for future cases.

5. Filter Types

Filters are divided into four main categories based on logical principles [6]: inclusion, exclusion, grouped and isolated filters. By reducing the filter types to logical operations and using set theoretical properties, it is possible to strictly define filters in terms of their purpose.

- **Inclusion Filters:** These filters are defined in terms of information that should be included in the result, e.g., all JPEG files on a device.
- **Exclusion Filters:** These filters are defined in terms of information that should be excluded from the result, e.g., exclude all images smaller than 50mm $\times$ 50mm.
- **Grouped Filters:** These filters specifically target similar types of data in the general vicinity of each other. Depending on how the filter is specified, it could select files in the same directory, or files in neighboring directories, or 90% of the same type of files in the same directory or in neighboring directories.
- **Isolated Filters:** These filters specifically target dissimilar types of data in the general vicinity of each other. An example is a small number of files of one type that are dispersed among a large number of files of different types.

Intersection is a useful operation for combining the results of different filters. The intersection operation on filters yields four possible outcomes: (i) Accept everything from both filters, (ii) Accept all but the intersection, (iii) Accept the intersection, and (iv) Accept the difference.

A collection of filters can be used in different combinations to implement complex selection criteria similar to using logical formulas. Individual filters can also be organized into clusters, resulting in successive layers of inclusion and exclusion. The selection criteria of the initial filter determine the data that is operated on by the remaining filters. Data that is not within the filter's criteria is placed in the set of non-processed data. This set is of interest because it may contain concealed, malformed, unknown and/or deleted data.

The purpose of a forensic examination is to locate incriminating data. Therefore, it stands to reason that a suspect may have gone to some

trouble to deliberately hide incriminating data. If this is the case, the data being sought could fall into the excluded set and a search for concealed data should concentrate on this data set.

Collections of filters can be specified based on case experience. These filters may be embedded into automated tools designed for field investigations. By validating these tools against the requirements of forensic analyses, it is possible to ensure the authenticity of the recovered data.

6. Conclusions

Filters help reduce the size of the search state space in forensic investigations of storage media that hold large amounts of data. Pre-filtering electronic data in the field or during an in-house investigation can save valuable time. In addition, filters support specific tasks such as locating and identifying encryption software and hidden, encrypted or compressed files. Filters can automate tedious examinations of temporary Internet files, Windows directories or illicit images. Filters also facilitate customized searches based on patterns encountered in investigations of common cases. Finally, filters can help collect statistical information on common data hiding techniques and effective search strategies.

References

[1] M. Anderson, Hard disk drives: Bigger is not better (www.forensics-intl.com/art14.html), 2001.

[2] AntiOnline.com, Basic data hiding tutorial (www.antionline.com/printthread.php?threadid=251463&pagenumber=1), 2004.

[3] DIBS, The DIBS Methodology (www.dibsusa.com/methodology/methodology.html).

[4] M. Hannan and P. Turner, Australian forensic computing investigation teams: Research on competence, *Proceedings of the Seventh Pacific-Asia Conference on Information Systems*, 2003.

[5] ITsecurity.com, Encryption (www.itsecurity.com/security.htm?s=386), 2004.

[6] R. Johnsonbaugh, *Discrete Mathematics*, Prentice Hall, Englewood Cliffs, New Jersey, 2001.

[7] W. Kruse and J. Heiser, What exactly is computer forensics? (www.developer.com/java/other/article.php/3308361.01), 2004.

[8] R. McKemmish, What is forensic computing? *Australian Institute of Criminology: Trends & Issues in Crime and Criminal Justice*, pp. 1-6 (www.cit.uws.edu.au/compsci/computerforensics/Online%20Materials/ti118.pdf), 1999.

[9] D. Michaud, Adventures in computer forensics (www.sans.org/rr/papers/27/638.pdf), 2001.

[10] NIST, New Commerce Department magnetic microscope helps retrieve information from damaged or altered tapes (www.nist.gov/public/_affairs/releases/g00-108.htm), 2001.

[11] NIST, National Software Reference Library and Computer Forensics Tool Testing Project (www.nsrl.nist.gov/Project), 2003.

[12] M. Noblett, M. Pollitt and L. Presley, Recovering and examining computer forensic evidence, *Forensic Science Communications,* vol. 2(4), 2000.