PLEASE
LEAVE BOOKS
BELOW FOR
RESHELVING
14th

4. File systems

We focus in this chapter on the main storage of a computer or device, the storage which holds the operating system. Once we have a logical or physical image of it, we should survey the image before starting detailed analysis. Even though many images look alike because of their common software, small differences can guide you in further investigation. Generally speaking, it is easier to analyze broad details of a drive than to analyze lower-level details. Many lower-level details are specific to the operating system and are often poorly documented, like descriptions of the metadata records about files. Fortunately, most investigations need not get far into the details of a drive, though malware investigations can be an exception.

Much of what you see when you look at a drive image is the work of the operating system and major software. That means a forensic investigator must be familiar with operating systems and software and know where to find the key data in each. Yes, you can scan a drive as a sequence of bytes and just look for keywords or patterns while ignoring file boundaries, as we discuss in chapters **Error! Reference source not found.** and **Error! Reference source not found.**, but us ing the operating-system records often helps you find what you want much more quickly.

4.1. The NPS forensic corpus

4. Examples in this book come from our NPS forensic corpus. This consists of 4131 drive images comprising 249,630,276 million files. 3437 of these drives with 62,258,249 million files (the “Real Data Corpus”) were obtained by Prof. Simson Garfinkel by purchase of used digital equipment in 27 countries. 247 of the drives in the NPS corpus with 162,182,108 million files came from classroom and laboratory computers at our school; since these are centrally managed by our information-technology department, many files were identical on more than one drive. 447 additional drives with 2,518,919 files came from research sponsors, many of which were small mobile devices. Of the around 250 million files, about 60% were deleted. 11,762,672 of these were “orphan” files, files that had lost their names and much of their metadata, but they were not included in these totals; orphan files can provide small items like email addresses and Web links.

Most drives were collected 2005 to 2015, a period in which fewer protections like encryption provided obstacles to forensic analysis. They primarily came from workstations and computers. Some came from mobile devices, and a few came from servers. Most were owned by people who could install software, and the software loaded on them reflects their interests. Around 20% of the drives were unreadable, either due to failures of their hardware or deliberate erasure by their owners.

We have been asked if we can make our corpus publicly available. Unfortunately, we cannot give unrestricted access because it contains much personal information, and as part of the U.S. Government, we are subject to stringent laws on privacy of personal data. Parts of the data have been released and can be released, however, to researchers who are willing to sign sharing agreements, provided the data can be sanitized for release. However, if you want drive data, used computer sites and stores can be a good source for your own purchase. Much of our corpus was purchased as used equipment from stores.

Large collections of drives like ours can be useful, so it is desirable for many organizations that do forensics to archive their old images. Some reasons are:

- Statistical analysis of the collection can show the typical patterns on drives, enabling easier recognition of anomalies.
- Software and sites evolve over time, and it is often important to see the patterns in how they change.
- Rare file types are easier to characterize if you can get many examples.
- Malware campaigns usually target only one or a few machines at a time, so you may need to collect data from many drives to see just one instance of a new malware type.

4.2. Identifying the operating system

The operating system (top-level software) used by a drive affects its analysis, so it is important to identify the operating system and its version at the start. Most of the NPS corpus use versions of the Windows operating system because of its popularity, some computers use Apple operating systems, and mobile devices use Android and IOS. Many image-analysis tools automatically identify the operating system for you. If not, usually some obvious clues occur:

- A directory "WIN95" or "WIN98" indicates Windows 95 or 98.
- A "WINDOWS" directory and NTFS-related directories indicate at least Windows NT.
- "Windows 2000" was used for Windows 2000.
- A "Windows" directory was used since then.
- A directory "Program Files" occurs in all Windows versions.
- A "WINNT" directory indicates Windows NT or XP.
- Windows has directories "Documents and Settings" and "i386" through Windows XP.
- Windows has a directory "Users" in Windows 10.
- Windows systems starting with XP have a file HIBERFIL.SYS and a directory "System Volume Information".

Forensic tools for drive analysis try to spare you this analysis and present a uniform view of the file system, independent of the operating system.

4.3. The front of the main drive

The layout of the main drive of a computer or device differs with the operating system and hardware, but almost always the key data for running it is stored at its beginning, in the lowest addresses on the drive. For Windows systems it includes:

- Some mostly blank header information. Traditionally this was 63 bytes, but mobile devices are using more of it.
- Information about the volumes (partitions) of a drive.
- The *master boot record* (listed as "$MBR" in Windows systems) which contains most of the initialization code for a computer or device when it is turned on. The first steps of startup (*booting*) are usually kept in *firmware* (reconfigurable hardware), and eventually transfer control to secondary storage.
- The *master file table* (listed as $MFT in modern Windows systems) which lists all the files and directories, where they are stored, and other important metadata about them. The master boot record gives the start of this table. Entries are 1024 bytes long on Windows systems (but a file can be described by multiple entries) and contain file names, directory names, sizes, timestamps, deletion statuses, access rights, and locations on secondary storage. If a file is fragmented and stored at several disjoint locations (as happens with files repeatedly expanded like log files), the entries specify the locations of all fragments.
- The first 16 entries in the Windows Master File Table are reserved for special files such as the location of the table backup, the location of the logging data, allocation statuses of

major partitions of the file system, the disk areas with faulty data, and general access-control information. The names of these entries are also preceded by the "$" character so you do not confuse them with regular files.

As mentioned in section **Error! Reference source not found.**, drives may be partitioned into " volumes" or other groupings holding fundamentally different kinds of data. Volumes are necessary when drives run virtual machines so that one machine cannot interfere with another, but they are also useful when parts of storage need extra protection from runaway programs and malware, as for the backup files for the operating system. Each volume has a separate Master File Table which forensic tools can identify. Figure 1 shows the volumes on one drive in our corpus as reported by the SleuthKit tool. Often you see empty volumes in a forensic image like 01, 03, 09, 13, and 15 below that were preallocated by a vendor but never used.

```
/corp/nus/drives/IN/IN10-0014/IN10-0014.E01
DOS Partition Table
Offset Sector: 0
Units are in 512-byte sectors
     Slot    Start        End          Length       Description
00:  Meta    0000000000   0000000000   0000000001   Primary Table (#0)
01:  -----   0000000000   0000000062   0000000063   Unallocated
02:  00:00   0000000063   0019535039   0019534977   NTFS (0x07)
03:  -----   0019535040   0019551167   0000016128   Unallocated
04:  Meta    0019551105   0078156224   0058605120   Win95 Extended
(0x0f)
05:  Meta    0019551105   0019551105   0000000001   Extended Table (#1)
06:  01:00   0019551168   0039102209   0019551042   NTFS (0x07)
07:  Meta    0039102210   0058653314   0019551105   DOS Extended (0x05)
08:  Meta    0039102210   0039102210   0000000001   Extended Table (#2)
09:  -----   0039102210   0039102272   0000000063   Unallocated
10:  02:00   0039102273   0058653314   0019551042   NTFS (0x07)
11:  Meta    0058653315   0078156224   0019502910   DOS Extended (0x05)
12:  Meta    0058653315   0058653315   0000000001   Extended Table (#3)
13:  -----   0058653315   0058653377   0000000063   Unallocated
14:  03:00   0058653378   0078156224   0019502847   NTFS (0x07)
15:  -----   0078156225   0078165359   0000009135   Unallocated
```

Figure 1: Example volume information for a drive reported with the MMLS command in the SleuthKit tool.

To retrieve a file in the forensic image of a drive, the forensic tool and the operating system must know:

- The start of its volume (the volume's "offset").
- The address of the file in the volume starting at 0. For many forensic tools such as SleuthKit, these are the Inode numbers found when a drive is imaged.
- If the file is embedded within a container file like a zip compression, the address within the container. Containers may contain other containers.

4.4. Inspecting file metadata

The file directory on a system and its metadata are often the best place to start an investigation. Drive-analysis tools often supplement the metadata from the operating system with hash values computed on the file, *inode* (secondary-storage index) number, and classification of the file type as discussed in chapter **Error! Reference source not found.**. Some of this may be missing for l ogical images like those of mobile devices. Nonetheless, this will tell you much about what is on a drive, allowing you to make decisions about what files to open.

NIST, the U.S. National Institute of Standards and Technology, has a standard format DFXML for exchange of digital-forensics metadata (NIST, 2021). It is provided by many forensic tools. It is based on XML, a data-interchange format based itself on the Web language HTML, and is quite simple. Figure 2 shows example metadata of a file in DFXML. The file is called "rbtemp.cab" and is stored in the WINDOWS directory and subdirectory SYSBCKUP. It is in volume 1, is a true file (indicated by the "r"), occupies 922,943 bytes, is not deleted, has inode number 2642 on the magnetic disk that stored it, has protection bits 511 (read and write access for the owner and read access for everyone else), was created, modified, and accessed on August 14, 2007, is in Microsoft Cabinet archive format based on the front of the file, is not fragmented, and has two hash values that can be looked up.

```
<fileobject>
    <filename>WINDOWS/SYSBCKUP/rbtemp.cab</filename>
    <partition>1</partition>        volume of drive on which the file resides
    <id>2252</id>
    <name_type>r</name_type>         whether this is a file, directory, or link
    <filesize>922943</filesize>          size in bytes
    <unalloc>1</unalloc>       whether file is deleted
    <used>1</used>
    <inode>2642</inode>      index number of file in secondary storage
    <meta_type>1</meta_type>
    <mode>511</mode>         access rights for the file
    <nlink>0</nlink>
    <uid>0</uid>             one hash on the file
    <gid>0</gid>             another hash on the file, used by Microsoft operating systems
    <mtime>2007-08-14T22:15:22Z</mtime>          time last modified
    <atime>2007-08-14T07:00:00Z</atime>            time last accessed
    <crtime>2007-08-14T22:15:13Z</crtime>          time created
    <libmagic>Microsoft Cabinet archive data, 922943 bytes, 4 files</libmagic>   file classification
    <byte_runs>           secondary-storage byte addresses of all fragments of the file
     <byte_run file_offset='0' fs_offset='15761920' img_offset='15794176' len='32768'/>
    </byte_runs>
    <hashdigest type='md5'>871c3297316af0e235b39196eef757ab</hashdigest>
    <hashdigest    type='sha1'>b15d4972c9af0b798f3d0882c29e682b544e1717</hashdigest>
</fileobject>
```

Figure 2: Example of DFXML metadata format.

Analysis tools can retrieve deleted files from physical images, and also some deleted files from "recycle bins" in logical images. Deleting a file on most operating systems does not erase the file, but merely marks its addresses for possible reuse. That means that deleted data may be useful in a criminal investigation because users may not realize it is still around. *Erasing* tools can deliberately write over the deleted data to destroy it, but they are not often used, and their use can be a tip-off to illegal or criminal activity. Data on drives in the 1970s and 1980s needed to be erased several times due to residual magnetism, but this is less a problem today (NIST, 1988).

Often older deleted files are only reused in part, and you may still be able to gather evidence from the rest of the file. Even "orphan" files that have lost their metadata can be useful with the techniques of file carving and reassembly discussed in chapter **Error! Reference source not found.**.

4.5. Surveying file types on a drive

Often a good early step in forensic analysis of a drive is to analyze the distribution of file types. If you are investigating accounting fraud, drives with many spreadsheets are important; if you are investigating child pornography, drives with many pictures are important; if you are investigating a malware attack, executable files in the operating system are important. Table 1 gives a breakdown of percentages of file types in our full NPS corpus of the approximately 250 million files in 4131 images based on the *extensions* on the file name. File extensions are the characters in a file name after the last period symbol, like "txt" for the name "security.syslog.txt", or "None" if there is no period.

Table 1: Percentages of files in the full NPS corpus by their file-extension type .

File extension type	Percentage	File extension type	Percentage
None	5.32%	Operating system	6.72%
Executable	10.08%	Script or source code	10.74%
Configuration file	2.48%	Log file	0.34%
Graphics	14.78%	JPEG or camera image	3.17%
Video	0.33%	Audio	1.38%
Microsoft Word	0.32%	Other document type	2.03%
Spreadsheet	0.30%	Presentation	0.06%
Web page	8.92%	XML or data interchange	3.01%
Database	0.59%	Queries	0.25%
Email or message	0.21%	Other Microsoft Office	0.10%
Link	0.48%	Help	0.60%
Temporary file	0.94%	Copies	0.19%
Compressed or encoded	1.04%	Geography	0.18%
Dictionary	0.06%	Index	4.17%
Integer extension	0.79%	Form	0.01%
Update	1.34%	Security	0.18%
Hardware-related	0.07%	Network	0.08%
Game-related	2.14%	Engineering	0.79%
Science-related	0.26%	Virtual machine or image	0.08%
Miscellaneous	0.05%		

Quick Question 4-1: Why isn't more video reported in the table when people use video all the time?

Another way to classify files is by the directories in which they reside. For instance, a directory "pics" contains pictures, "Web downloads" contains Web pages, and "Windows" contains executables for the Microsoft Windows operating system. Classifying files this way focuses on their purpose rather than their format, so the classes usefully supplement the file-extension data. Table 2 shows the breakdown of the types of immediate directories for our corpus, the enclosing directories in which a file occurs.

Table 2: Percentages of files in the full NPS corpus by their immediate directory.

Immediate directory type	Percentage in our corpus
Top level	2.32%
Operating system	12.18%
Hardware	3.14%
Logs and backup	5.08%
Temporary files	7.88%
Help	7.43%
Pictures	8.48%
Audio	0.88%
Video	0.21%
Web-related	4.14%
Data	2.64%
Programs	4.06%
Documents	7.84%
Sharing	1.25%
Security	0.36%
Mail	0.10%
Games	0.55%
Installation	5.13%
Applications	17.61%
Miscellaneous	0.56%

As a sample application of this data, Figure 6 shows some specific directory classifications for our corpus.

Software:
3 textconv 76
3 britannica 76
3 corel 76
3 corel5w 76
3 corel40 76
3 corel draw 12 76
3 coreldraw graphics suite 13 76
3 coreldraw graphics v8.0 76
3 coreldraw graphics v9.0 76
3 corel 12 76
3 winfax 76
3 icq 76
3 office xp 76
3 pubwiz 76
3 components 76
3 ui 76
3 nwclient 76
3 nwclient4.9sp1a 76
3 client 76
3 clientcon 76
3 zrmodels 76
3 unconvtab 76
3 ancillary 76
3 calweek 76
3 the microsoft network 76
3 enu 76
3 bitware 76
3 outlookexpress 76
3 remote assistance 76

Pictures:
3 cliparts 64
3 color books 64
3 colorfilm 64
3 cursores 64
3 dahla picnic 1387-1-3 64
3 filtergraphs 64
3 flag 64
3 foto,pintura 64
3 fotografias de lenin 64
3 framebuttons 64
3 frames(masks)_for_photoritesp 64
3 friendicon 64
3 gif photos 64
3 graphics-full_existing 64
3 gray quadtones 64
3 gray tritones 64
3 h.agha truck & dahla in mobail date
1383.12.20 64
3 handahar city from air 64
3 helmend & kar date 2005.7.12 124 64
3 home pic 64
3 icons_content 64
3 icons_handlers 64
3 icons_mediacategory 64
3 icons_menuitems 64
3 icons_osd 64
3 icons_settings 64
3 icons_setupwizard 64
3 image library 64
3 interiors 64

Figure 3: Examples of directory names assigned to software and pictures.

Calculating the distribution of file types and directories of a drive is very useful because it can tell you what is distinctive about the drive. Appendix A has a Python program to create this from the DFXML metadata file for a drive, and it also has a program to create DFXML data for a drive by navigating the directory hierarchy. Appendix A also includes text files based on the NPS corpus listing mappings of 15,075 extensions to 45 classes, 7,400 immediate directories to 21 classes, and 7,067 top-level directories to 12 classes. Section **Error! Reference source not found.** d iscusses additional methods for identifying the type of a file that are useful on deleted files and file fragments.

4.6. Surveying the file directory of example drive TH with Autopsy

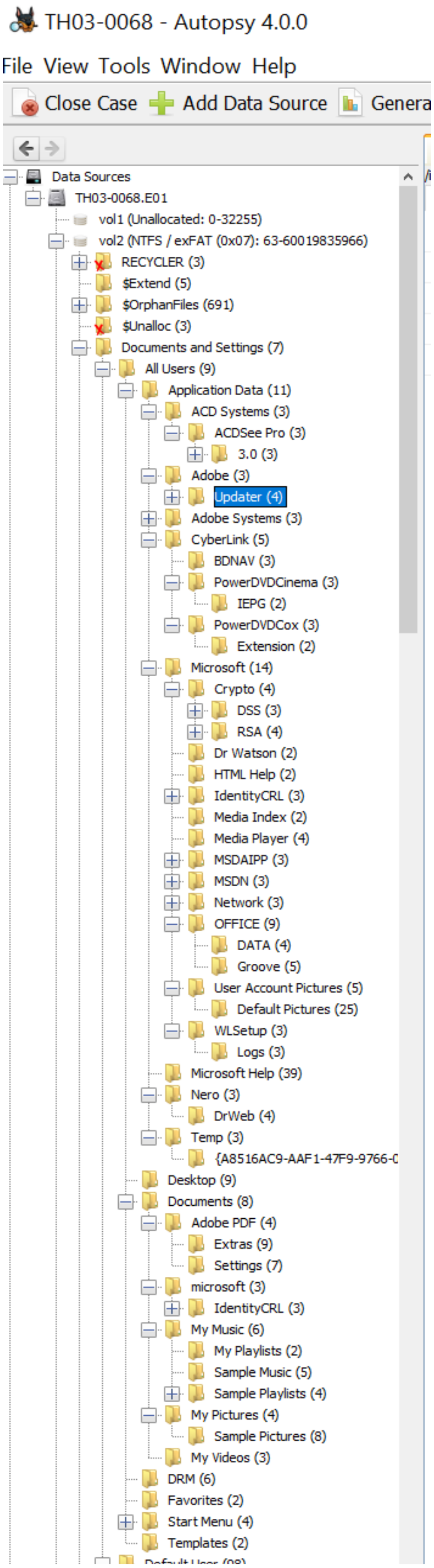
We use the forensic tool Autopsy (www.autopsy.com) for many examples of drive analysis in this book since it is a free open-source tool that is relatively easy to use. It provides a graphical user interface to forensic data as an extension of "The Sleuth Kit" (TSK) command-line tool (www.sleuthkit.com) which has been around a while. We start with the first steps of an investigation once the drive image has been loaded into Autopsy. Often the first steps are important since they must reduce the amount of data considered to an amount for efficient analysis.

Our first example here is a drive from Thailand, which we call the “TH” drive, purchased like most of our drives with used computer equipment. 26,685 files were found by Autopsy on the drive, which is low for a machine running the Windows operating system, so apparently many user files were deleted before we acquired it.

We skip the details of acquisition that produced an “.E01” file. We started Autopsy and loaded the file, asking it to give us a directory listing and identify several kinds of important files. Figure 4 shows the top of the file system overview shown by Autopsy. Indentation indicates subdirectories and subfiles; “+” means that a line can be expanded to show them, and “-“ means the line has already been fully expanded. Numbers in parentheses at the ends of line indicate the number of such subdirectories and subfiles.

Some unallocated bytes occur at the beginning of the drive labeled as “Volume 1”, then a very large partition “Volume 2” for the rest of the drive extending to storage address 60,019,835,966 (so this volume holds at least 60 gigabytes). This drive uses the Windows operating system because it refers to the Windows file system NTFS (and also its ancestor FAT) in the Volume line. The display also shows several directories associated with Windows such as MICROSOFT, Media Player, and OFFICE (for Microsoft Office software). This had a Windows XP operating system, but that is not explicitly mentioned.

Often in investigation of a Windows machine you want to examine the files created by a user. Many in Windows XP are stored in the Documents and Settings directory which we have expanded on the screen. (Much of this has now moved to a “Users” directory under more recent versions of Windows.) At the top level, Documents and Settings has subdirectories for each user, plus a few general directories like “All Users” and “Default User” for data that applies more broadly. Here we see references to Microsoft programs loaded on this machine that could give information about its use.

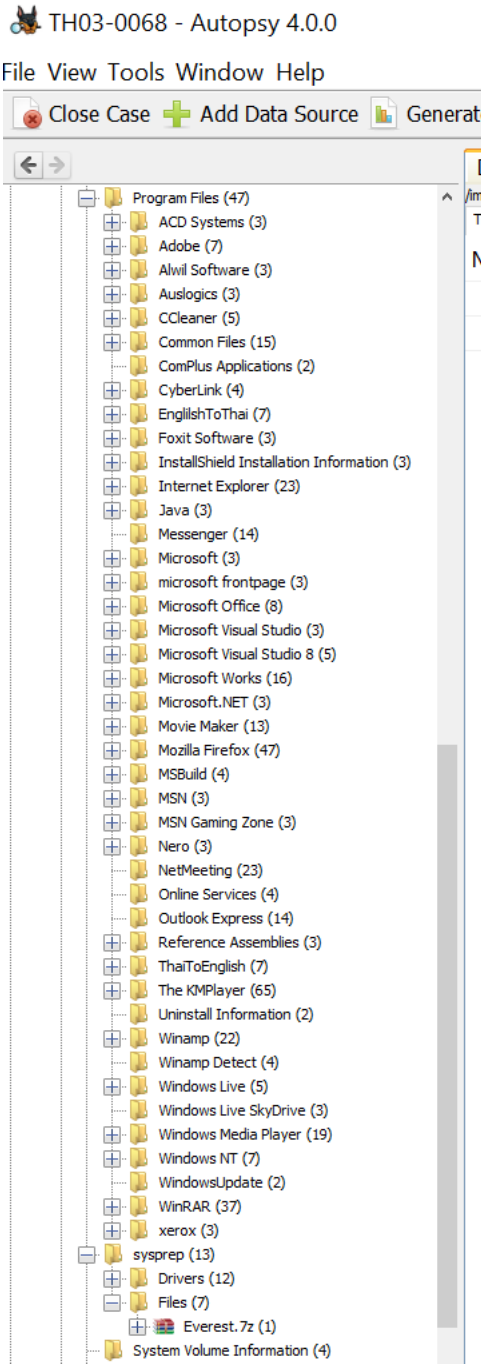


TH03-0068 - Autopsy 4.0.0

```
Data Sources
  TH03-0068.E01
    vol1 (Unallocated: 0-32255)
    vol2 (NTFS / exFAT (0x07): 63-60019835966)
      RECYCLER (3)
      $Extend (5)
      $OrphanFiles (691)
      $Unalloc (3)
      Documents and Settings (7)
        All Users (9)
          Application Data (11)
            ACD Systems (3)
              ACDSee Pro (3)
                3.0 (3)
            Adobe (3)
              Updater (4)
            Adobe Systems (3)
            CyberLink (5)
              BDNAV (3)
              PowerDVDCinema (3)
                IEPG (2)
              PowerDVDCox (3)
                Extension (2)
            Microsoft (14)
              Crypto (4)
                DSS (3)
                RSA (4)
              Dr Watson (2)
              HTML Help (2)
              IdentityCRL (3)
              Media Index (2)
              Media Player (4)
              MSDAIPP (3)
              MSDN (3)
              Network (3)
              OFFICE (9)
                DATA (4)
                Groove (5)
              User Account Pictures (5)
                Default Pictures (25)
              WLSetup (3)
                Logs (3)
            Microsoft Help (39)
            Nero (3)
              DrWeb (4)
            Temp (3)
              {A8516AC9-AAF1-47F9-9766-C
          Desktop (9)
          Documents (8)
            Adobe PDF (4)
              Extras (9)
              Settings (7)
            microsoft (3)
              IdentityCRL (3)
            My Music (6)
              My Playlists (2)
              Sample Music (5)
              Sample Playlists (4)
            My Pictures (4)
              Sample Pictures (8)
            My Videos (3)
          DRM (6)
          Favorites (2)
          Start Menu (4)
          Templates (2)
        Default User (98)
```

Figure 4: Top of file hierarchy for TH03-0068, a drive purchased in Thailand.

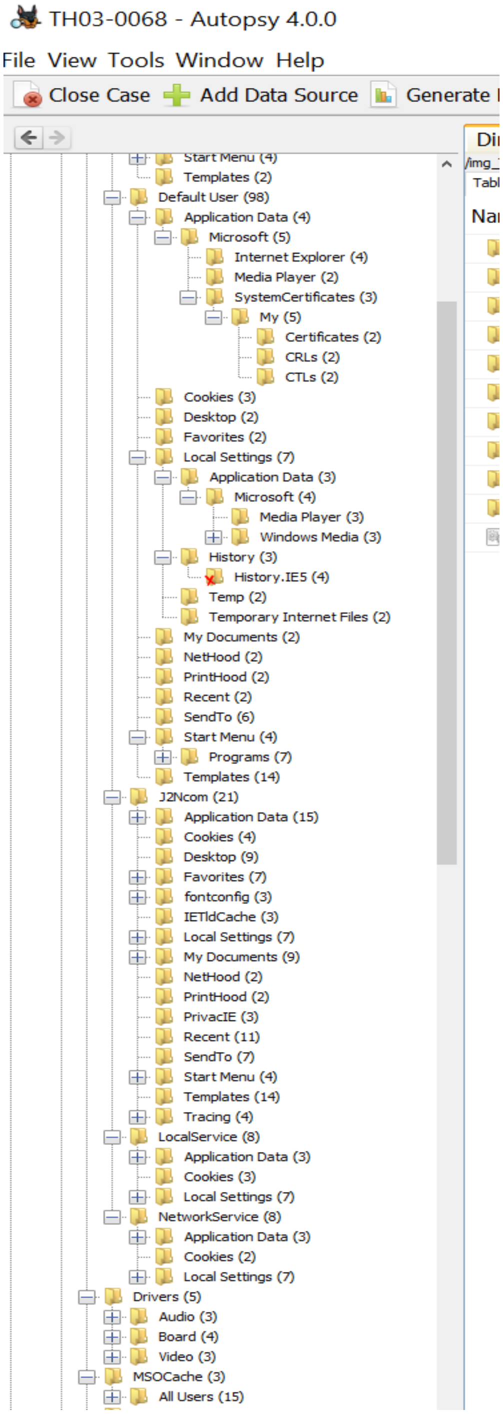
Figure 5 shows an expansion of the "Program Files" directory. This contains software executables, and often gives good indications about the use of a drive. Most software is standard except for EnglishToThai and ThaiToEnglish which say something about the drive origin. Since this is mostly business software with no games, it is reasonable to think this drive came from a computer used in a business.



```
Program Files (47)
    ACD Systems (3)
    Adobe (7)
    Alwil Software (3)
    Auslogics (3)
    CCleaner (5)
    Common Files (15)
    ComPlus Applications (2)
    CyberLink (4)
    EnglishToThai (7)
    Foxit Software (3)
    InstallShield Installation Information (3)
    Internet Explorer (23)
    Java (3)
    Messenger (14)
    Microsoft (3)
    microsoft frontpage (3)
    Microsoft Office (8)
    Microsoft Visual Studio (3)
    Microsoft Visual Studio 8 (5)
    Microsoft Works (16)
    Microsoft.NET (3)
    Movie Maker (13)
    Mozilla Firefox (47)
    MSBuild (4)
    MSN (3)
    MSN Gaming Zone (3)
    Nero (3)
    NetMeeting (23)
    Online Services (4)
    Outlook Express (14)
    Reference Assemblies (3)
    ThaiToEnglish (7)
    The KMPlayer (65)
    Uninstall Information (2)
    Winamp (22)
    Winamp Detect (4)
    Windows Live (5)
    Windows Live SkyDrive (3)
    Windows Media Player (19)
    Windows NT (7)
    WindowsUpdate (2)
    WinRAR (37)
    xerox (3)
sysprep (13)
    Drivers (12)
    Files (7)
        Everest.7z (1)
System Volume Information (4)
```

Figure 5: Program Files directory for TH03-0068.

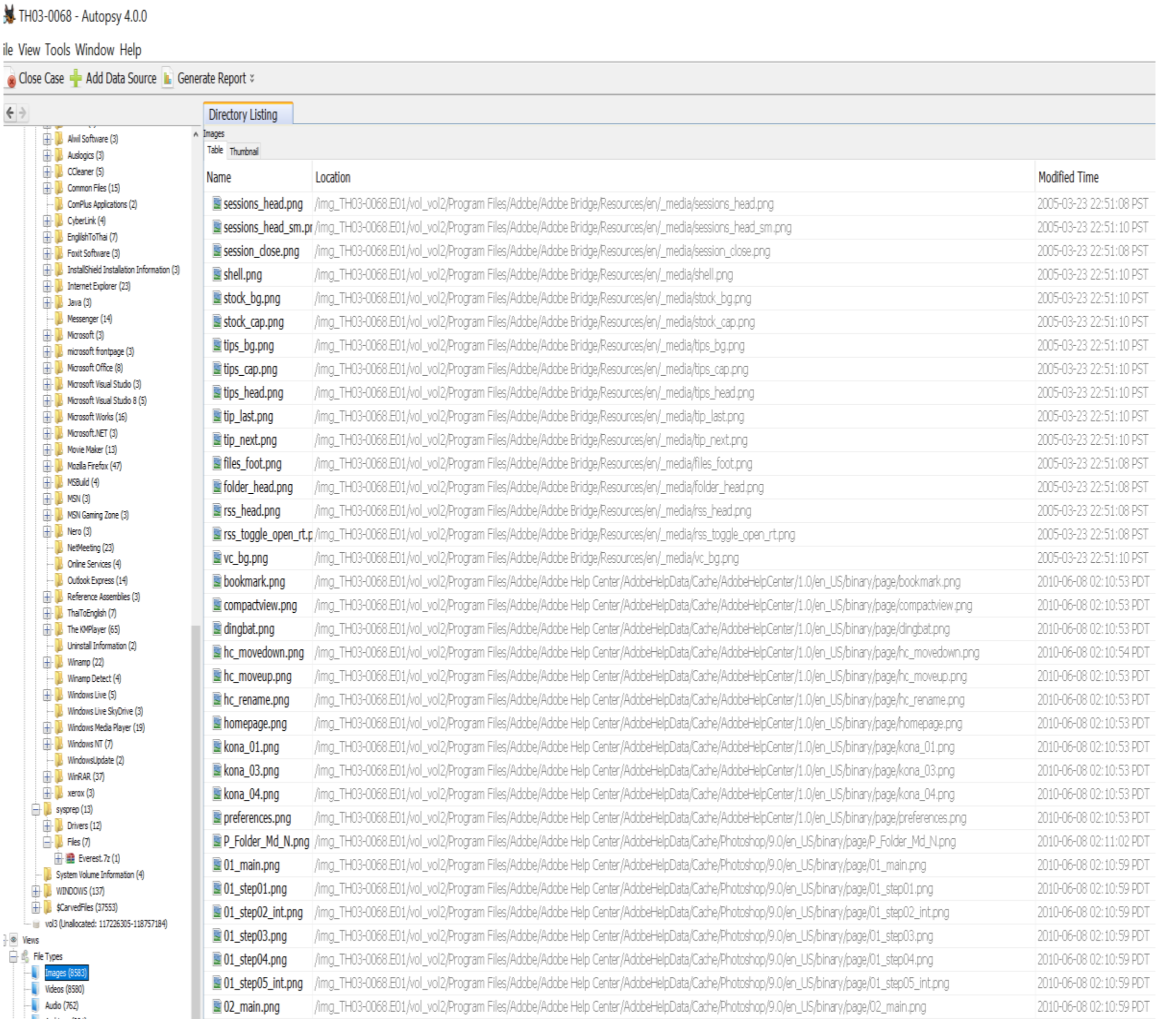
Figure 6 shows some of the files on this drive under the Default User directory under Documents and Settings. Apparently logins were not required on this drive and much activity was attributed to the "default user". User files like these can help in an investigation because they indicate activities. Here Cookies, Favorite, History, and My Documents could be particularly useful.



```
    Start Menu (4)
    Templates (2)
  Default User (98)
    Application Data (4)
      Microsoft (5)
        Internet Explorer (4)
        Media Player (2)
        SystemCertificates (3)
          My (5)
            Certificates (2)
            CRLs (2)
            CTLs (2)
    Cookies (3)
    Desktop (2)
    Favorites (2)
    Local Settings (7)
      Application Data (3)
        Microsoft (4)
          Media Player (3)
          Windows Media (3)
      History (3)
        History.IE5 (4)
      Temp (2)
      Temporary Internet Files (2)
    My Documents (2)
    NetHood (2)
    PrintHood (2)
    Recent (2)
    SendTo (6)
    Start Menu (4)
      Programs (7)
    Templates (14)
  J2Ncom (21)
    Application Data (15)
    Cookies (4)
    Desktop (9)
    Favorites (7)
    fontconfig (3)
    IETldCache (3)
    Local Settings (7)
    My Documents (9)
    NetHood (2)
    PrintHood (2)
    PrivacIE (3)
    Recent (11)
    SendTo (7)
    Start Menu (4)
    Templates (14)
    Tracing (4)
  LocalService (8)
    Application Data (3)
    Cookies (3)
    Local Settings (7)
  NetworkService (8)
    Application Data (3)
    Cookies (2)
    Local Settings (7)
Drivers (5)
  Audio (3)
  Board (4)
  Video (3)
MSOCache (3)
  All Users (15)
```

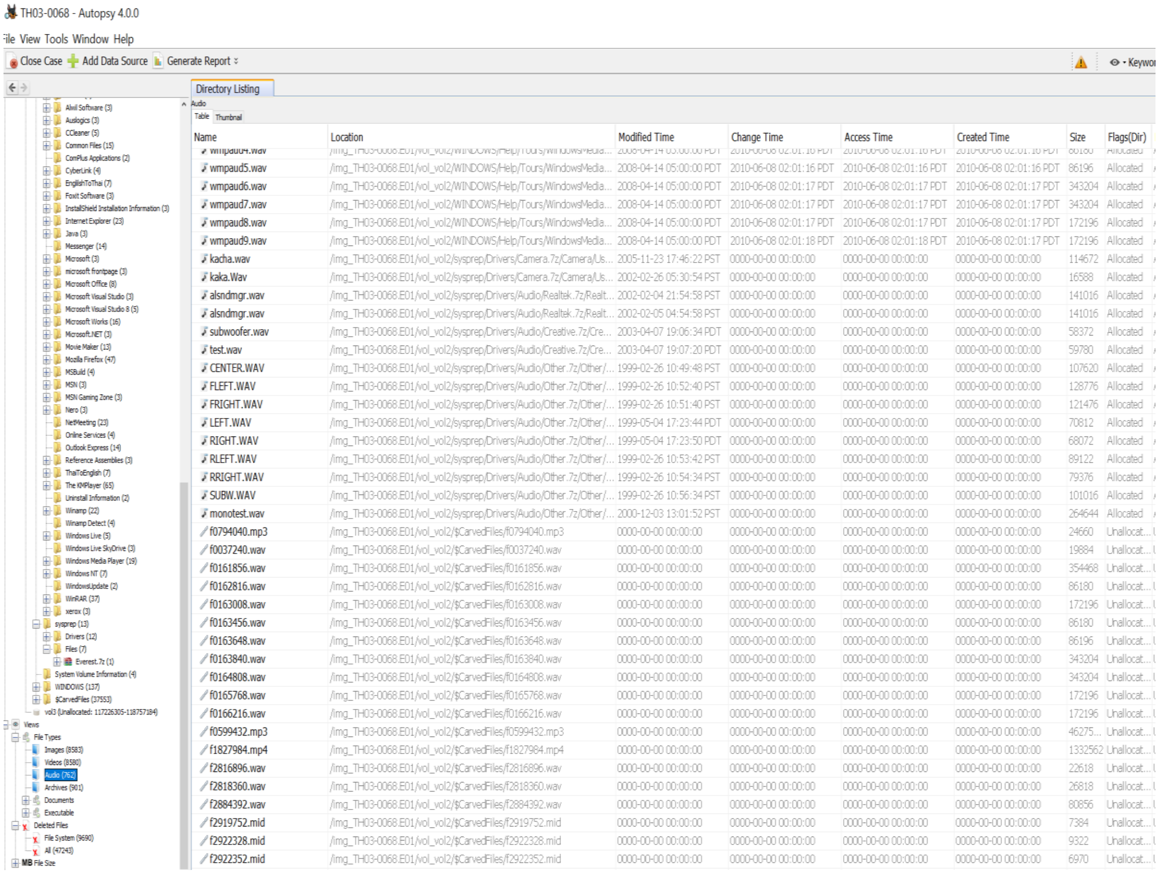
Figure 6: Files for the "default user" on this drive.

Autopsy also creates artificial (“virtual”) directories of file types likely to be useful evidence. Figure 7 lists some paths of pictures (called “images” here, not to be confused with the forensic meaning). But those shown are uninteresting since they all come from the Program Files directory and are likely graphics for programs. They also are uninteresting because their names are short, around the same length, and the timestamps are mostly identical, usually indicators of software origin. Figure 8 similarly shows some audio files from the drive. These are almost entirely in “wav” format, a Microsoft audio format usually used with graphics, so they are not interesting.



Name	Location	Modified Time
sessions_head.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/sessions_head.png	2005-03-23 22:51:08 PST
sessions_head_sm.pr	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/sessions_head_sm.png	2005-03-23 22:51:10 PST
session_close.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/session_close.png	2005-03-23 22:51:08 PST
shell.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/shell.png	2005-03-23 22:51:10 PST
stock_bg.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/stock_bg.png	2005-03-23 22:51:10 PST
stock_cap.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/stock_cap.png	2005-03-23 22:51:10 PST
tips_bg.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/tips_bg.png	2005-03-23 22:51:10 PST
tips_cap.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/tips_cap.png	2005-03-23 22:51:10 PST
tips_head.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/tips_head.png	2005-03-23 22:51:10 PST
tip_last.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/tip_last.png	2005-03-23 22:51:10 PST
tip_next.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/tip_next.png	2005-03-23 22:51:10 PST
files_foot.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/files_foot.png	2005-03-23 22:51:08 PST
folder_head.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/folder_head.png	2005-03-23 22:51:08 PST
rss_head.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/rss_head.png	2005-03-23 22:51:08 PST
rss_toggle_open_rt.p	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/rss_toggle_open_rt.png	2005-03-23 22:51:08 PST
vc_bg.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Bridge/Resources/en/_media/vc_bg.png	2005-03-23 22:51:10 PST
bookmark.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/bookmark.png	2010-06-08 02:10:53 PDT
compactview.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/compactview.png	2010-06-08 02:10:53 PDT
dingbat.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/dingbat.png	2010-06-08 02:10:53 PDT
hc_movedown.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/hc_movedown.png	2010-06-08 02:10:54 PDT
hc_moveup.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/hc_moveup.png	2010-06-08 02:10:53 PDT
hc_rename.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/hc_rename.png	2010-06-08 02:10:53 PDT
homepage.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/homepage.png	2010-06-08 02:10:53 PDT
kona_01.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/kona_01.png	2010-06-08 02:10:53 PDT
kona_03.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/kona_03.png	2010-06-08 02:10:53 PDT
kona_04.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/kona_04.png	2010-06-08 02:10:53 PDT
preferences.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/AdobeHelpCenter/1.0/en_US/binary/page/preferences.png	2010-06-08 02:10:53 PDT
P_Folder_Md_N.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/P_Folder_Md_N.png	2010-06-08 02:11:02 PDT
01_main.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_main.png	2010-06-08 02:10:59 PDT
01_step01.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_step01.png	2010-06-08 02:10:59 PDT
01_step02_int.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_step02_int.png	2010-06-08 02:10:59 PDT
01_step03.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_step03.png	2010-06-08 02:10:59 PDT
01_step04.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_step04.png	2010-06-08 02:10:59 PDT
01_step05_int.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/01_step05_int.png	2010-06-08 02:10:59 PDT
02_main.png	/img_TH03-0068.E01/vol_vol2/Program Files/Adobe/Adobe Help Center/AdobeHelpData/Cache/Photoshop/9.0/en_US/binary/page/02_main.png	2010-06-08 02:10:59 PDT

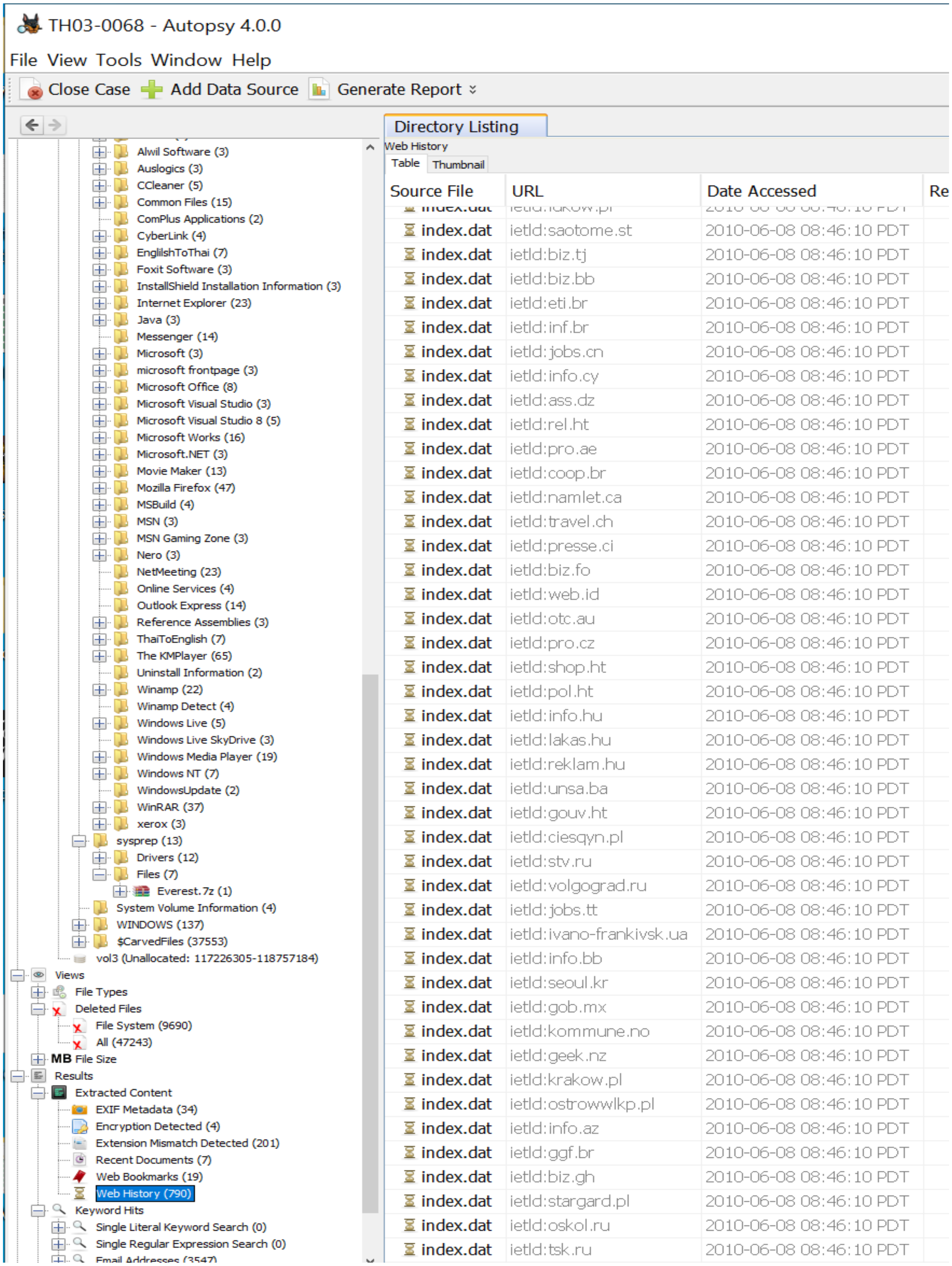
Figure 7: Some example picture files on this drive.



Name	Location	Modified Time	Change Time	Access Time	Created Time	Size	Flags(Dir)
wmpaud4.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:16 PDT	2010-06-08 02:01:16 PDT	2010-06-08 02:01:16 PDT	86180	Allocated
wmpaud5.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:16 PDT	2010-06-08 02:01:16 PDT	2010-06-08 02:01:16 PDT	86196	Allocated
wmpaud6.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	343204	Allocated
wmpaud7.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	343204	Allocated
wmpaud8.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	2010-06-08 02:01:17 PDT	172196	Allocated
wmpaud9.wav	/img_TH03-0068.E01/vol_vol2/WINDOWS/Help/Tours/WindowsMedia...	2008-04-14 05:00:00 PDT	2010-06-08 02:01:18 PDT	2010-06-08 02:01:18 PDT	2010-06-08 02:01:17 PDT	172196	Allocated
kacha.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Camera.7z/Camera/Us...	2005-11-23 17:46:22 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	114672	Allocated
kaka.Wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Camera.7z/Camera/Us...	2002-02-26 05:30:54 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	16588	Allocated
alsndmgr.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Realtek.7z/Realt...	2002-02-04 21:54:58 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	141016	Allocated
alsndmgr.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Realtek.7z/Realt...	2002-02-05 04:54:58 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	141016	Allocated
subwoofer.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Creative.7z/Cre...	2003-04-07 19:06:34 PDT	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	58372	Allocated
test.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Creative.7z/Cre...	2003-04-07 19:07:20 PDT	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	59780	Allocated
CENTER.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:49:48 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	107620	Allocated
FLEFT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:52:40 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	128776	Allocated
FRIGHT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:51:40 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	121476	Allocated
LEFT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-05-04 17:23:44 PDT	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	70812	Allocated
RIGHT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-05-04 17:23:50 PDT	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	68072	Allocated
RLEFT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:53:42 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	89122	Allocated
RRIGHT.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:54:34 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	79376	Allocated
SUBW.WAV	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	1999-02-26 10:56:34 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	101016	Allocated
monotest.wav	/img_TH03-0068.E01/vol_vol2/sysprep/Drivers/Audio/Other.7z/Other/...	2000-12-03 13:01:52 PST	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	264644	Allocated
f0794040.mp3	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0794040.mp3	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	24660	Unallocat...
f0037240.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0037240.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	19884	Unallocat...
f0161856.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0161856.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	354468	Unallocat...
f0162816.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0162816.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	86180	Unallocat...
f0163008.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0163008.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	172196	Unallocat...
f0163456.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0163456.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	86180	Unallocat...
f0163648.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0163648.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	86196	Unallocat...
f0163840.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0163840.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	343204	Unallocat...
f0164808.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0164808.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	343204	Unallocat...
f0165768.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0165768.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	172196	Unallocat...
f0166216.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0166216.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	172196	Unallocat...
f0599432.mp3	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f0599432.mp3	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	46275...	Unallocat...
f1827984.mp4	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f1827984.mp4	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	1332562	Unallocat...
f2816896.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2816896.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	22618	Unallocat...
f2818360.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2818360.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	26818	Unallocat...
f2884392.wav	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2884392.wav	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	80856	Unallocat...
f2919752.mid	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2919752.mid	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	7384	Unallocat...
f2922328.mid	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2922328.mid	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	9322	Unallocat...
f2922352.mid	/img_TH03-0068.E01/vol_vol2/$CarvedFiles/f2922352.mid	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00	6970	Unallocat...

Figure 8: Some example audio files.

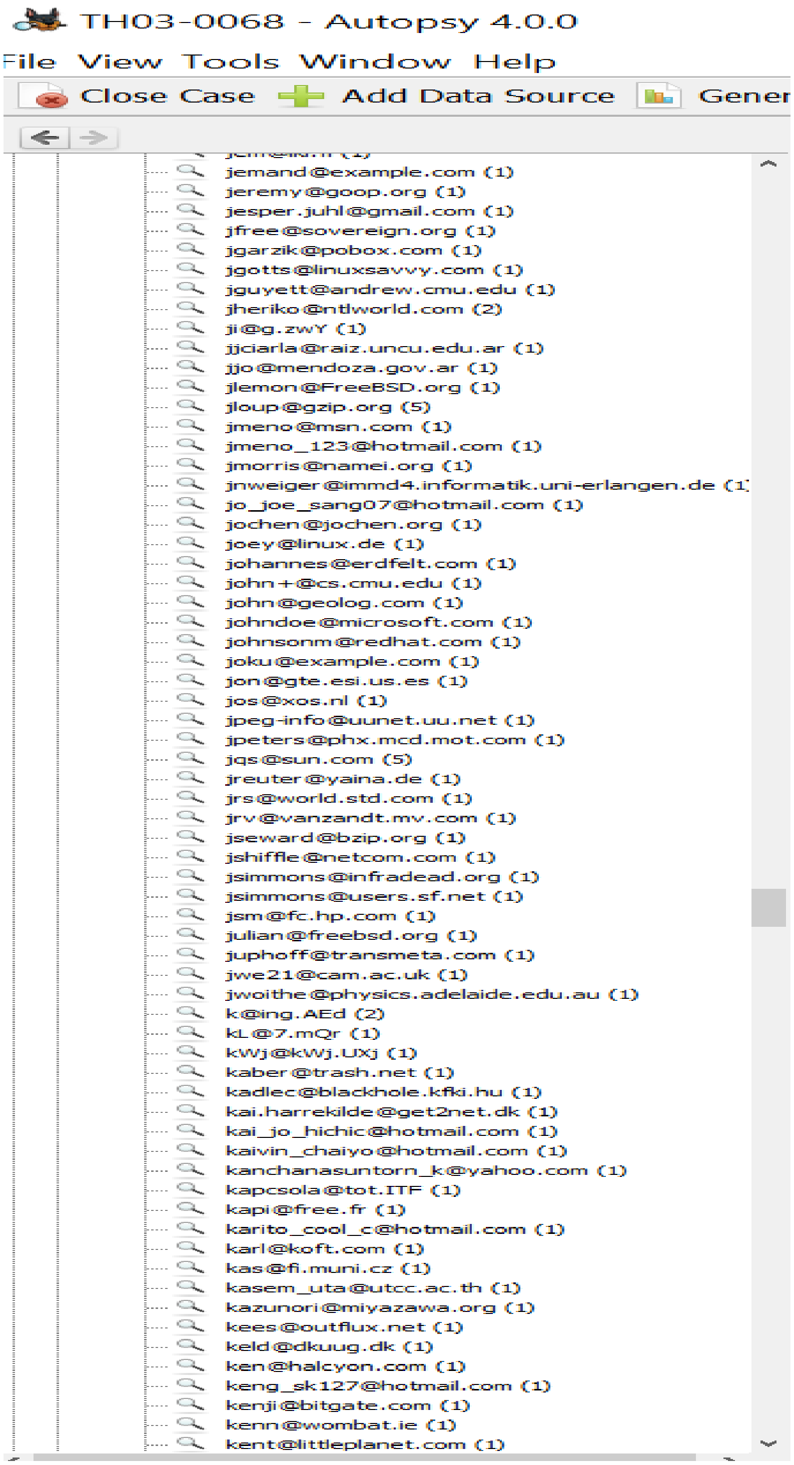
Web history can help in investigating a drive. But the Web history of this drive is not too interesting (Figure 9) since the entries seem to be pages visited automatically.



Source File	URL	Date Accessed	Re
index.dat	ietld:lukow.pl	2010-06-08 08:46:10 PDT	
index.dat	ietld:saotome.st	2010-06-08 08:46:10 PDT	
index.dat	ietld:biz.tj	2010-06-08 08:46:10 PDT	
index.dat	ietld:biz.bb	2010-06-08 08:46:10 PDT	
index.dat	ietld:eti.br	2010-06-08 08:46:10 PDT	
index.dat	ietld:inf.br	2010-06-08 08:46:10 PDT	
index.dat	ietld:jobs.cn	2010-06-08 08:46:10 PDT	
index.dat	ietld:info.cy	2010-06-08 08:46:10 PDT	
index.dat	ietld:ass.dz	2010-06-08 08:46:10 PDT	
index.dat	ietld:rel.ht	2010-06-08 08:46:10 PDT	
index.dat	ietld:pro.ae	2010-06-08 08:46:10 PDT	
index.dat	ietld:coop.br	2010-06-08 08:46:10 PDT	
index.dat	ietld:namlet.ca	2010-06-08 08:46:10 PDT	
index.dat	ietld:travel.ch	2010-06-08 08:46:10 PDT	
index.dat	ietld:presse.ci	2010-06-08 08:46:10 PDT	
index.dat	ietld:biz.fo	2010-06-08 08:46:10 PDT	
index.dat	ietld:web.id	2010-06-08 08:46:10 PDT	
index.dat	ietld:otc.au	2010-06-08 08:46:10 PDT	
index.dat	ietld:pro.cz	2010-06-08 08:46:10 PDT	
index.dat	ietld:shop.ht	2010-06-08 08:46:10 PDT	
index.dat	ietld:pol.ht	2010-06-08 08:46:10 PDT	
index.dat	ietld:info.hu	2010-06-08 08:46:10 PDT	
index.dat	ietld:lakas.hu	2010-06-08 08:46:10 PDT	
index.dat	ietld:reklam.hu	2010-06-08 08:46:10 PDT	
index.dat	ietld:unsa.ba	2010-06-08 08:46:10 PDT	
index.dat	ietld:gouv.ht	2010-06-08 08:46:10 PDT	
index.dat	ietld:ciesqyn.pl	2010-06-08 08:46:10 PDT	
index.dat	ietld:stv.ru	2010-06-08 08:46:10 PDT	
index.dat	ietld:volgograd.ru	2010-06-08 08:46:10 PDT	
index.dat	ietld:jobs.tt	2010-06-08 08:46:10 PDT	
index.dat	ietld:ivano-frankivsk.ua	2010-06-08 08:46:10 PDT	
index.dat	ietld:info.bb	2010-06-08 08:46:10 PDT	
index.dat	ietld:seoul.kr	2010-06-08 08:46:10 PDT	
index.dat	ietld:gob.mx	2010-06-08 08:46:10 PDT	
index.dat	ietld:kommune.no	2010-06-08 08:46:10 PDT	
index.dat	ietld:geek.nz	2010-06-08 08:46:10 PDT	
index.dat	ietld:krakow.pl	2010-06-08 08:46:10 PDT	
index.dat	ietld:ostrowwlkp.pl	2010-06-08 08:46:10 PDT	
index.dat	ietld:info.az	2010-06-08 08:46:10 PDT	
index.dat	ietld:ggf.br	2010-06-08 08:46:10 PDT	
index.dat	ietld:biz.gh	2010-06-08 08:46:10 PDT	
index.dat	ietld:stargard.pl	2010-06-08 08:46:10 PDT	
index.dat	ietld:oskol.ru	2010-06-08 08:46:10 PDT	
index.dat	ietld:tsk.ru	2010-06-08 08:46:10 PDT	

Figure 9: Web history for this drive.

Most tools can also collect important small data items (*artifacts*) like email addresses (Figure 10). We will discuss this more in chapter **Error! Reference source not found.**. Nearly all of these w ere found within software files, judging by the business names and the "(1)" annotations which say the address occurred only once on the drive. Thus they are likely mostly advertising links.



```
jemand@example.com (1)
jeremy@goop.org (1)
jesper.juhl@gmail.com (1)
jfree@sovereign.org (1)
jgarzik@pobox.com (1)
jgotts@linuxsavvy.com (1)
jguyett@andrew.cmu.edu (1)
jheriko@ntlworld.com (2)
ji@g.zwY (1)
jjciarla@raiz.uncu.edu.ar (1)
jjo@mendoza.gov.ar (1)
jlemon@FreeBSD.org (1)
jloup@gzip.org (5)
jmeno@msn.com (1)
jmeno_123@hotmail.com (1)
jmorris@namei.org (1)
jnweiger@immd4.informatik.uni-erlangen.de (1
jo_joe_sang07@hotmail.com (1)
jochen@jochen.org (1)
joey@linux.de (1)
johannes@erdfelt.com (1)
john+@cs.cmu.edu (1)
john@geolog.com (1)
johndoe@microsoft.com (1)
johnsonm@redhat.com (1)
joku@example.com (1)
jon@gte.esi.us.es (1)
jos@xos.nl (1)
jpeg-info@uunet.uu.net (1)
jpeters@phx.mcd.mot.com (1)
jqs@sun.com (5)
jreuter@yaina.de (1)
jrs@world.std.com (1)
jrv@vanzandt.mv.com (1)
jseward@bzip.org (1)
jshiffle@netcom.com (1)
jsimmons@infradead.org (1)
jsimmons@users.sf.net (1)
jsm@fc.hp.com (1)
julian@freebsd.org (1)
juphoff@transmeta.com (1)
jwe21@cam.ac.uk (1)
jwoithe@physics.adelaide.edu.au (1)
k@ing.AEd (2)
kL@7.mQr (1)
kWj@kWj.UXj (1)
kaber@trash.net (1)
kadlec@blackhole.kfki.hu (1)
kai.harrekilde@get2net.dk (1)
kai_jo_hichic@hotmail.com (1)
kaivin_chaiyo@hotmail.com (1)
kanchanasuntorn_k@yahoo.com (1)
kapcsola@tot.ITF (1)
kapi@free.fr (1)
karito_cool_c@hotmail.com (1)
karl@koft.com (1)
kas@fi.muni.cz (1)
kasem_uta@utcc.ac.th (1)
kazunori@miyazawa.org (1)
kees@outflux.net (1)
keld@dkuug.dk (1)
ken@halcyon.com (1)
keng_sk127@hotmail.com (1)
kenji@bitgate.com (1)
kenn@wombat.ie (1)
kent@littleplanet.com (1)
```

Figure 10: Sample email addresses on the drive.

4.7. Statistical distributions on the TH drive

As mentioned in section 4.5, statistical measures of file distributions can also be very helpful in understanding the use of a drive. Usually digital-forensic examiners have far too much data to carefully examine it all, and need some quick guides to which drives and which parts of those drives would be most useful to analyze. Also, a graph can indicate drive features that are anomalous, and which should be explored in criminal or malware investigations where deliberate concealment has occurred. We show graphs from software we wrote; Autopsy and other tools, however, have their own plotting software that you may prefer.

Often the graphs for a drive are based on a considerable range of numbers, as with file sizes that range from 0 to billions of bytes. For these graphs, it is conventional to use the logarithms to the base e=2.71828182 (the natural logarithms, abbreviated "ln") rather than the logarithms to the base 2 shown in chapter 2. Table 3 shows some representative values. Logarithms to the base 2 are just the natural logarithms multiplied by 1.442695.

Table 3: Natural logarithms of some base values.

ln(X)	X	ln(X)	X	ln(X)	X	ln(X)	X
0.0	1.0000	**1.0**	2.7182	**2.0**	7.3890	**3.0**	20.086
4.0	54.598	**5.0**	148.41	**6.0**	403.43	**7.0**	1,096.6
8.0	2,981.0	**9.0**	8,203.1	**10.0**	22,026	**11.0**	59,874
12.0	162,575	**13.0**	442,413	**14.0**	1,202,604	**15.0**	3,269,017
16.0	8,886,111	**17.0**	24,154,953	**18.0**	65,659,969	**19.0**	178,482,301
20.0	485,165,195	**21.0**	1,218,815,734	**22.0**	3,584,912,846	**23.0**	9,744,803,446

4.7.1. Overall timestamp distributions on the drive

An important part of file metadata is the timestamps for key times associated with a file. Digital systems can use four kinds of timestamps:

- Creation time, crtime: When the file was first written on the imaged drive, not when the vendor created it.
- Modification time, time: When the file was last modified. This can be at the vendor for a software file that has been unchanged since it was downloaded. That means modification times can be earlier than creation times of files.
- Access time, abime: When the file was last accessed (read or written) on the drive.
- Metadata modification time, ctime: When the metadata for the file was last changed on the drive. Linux and Unix systems traditionally offer this instead of crtime.

Of the four, creation times are the most useful forensically. Once an operating system is installed on a drive and initialized, often the creation times are the best indicators of user activity since they refer to creation times on the specific hardware on which the files reside. How users create files can indicate the pattern of their activity. Modification times, on the other hand, are only really helpful for files a user modified on the drive. Access times can be inaccurate for older drives or much-modified files because it is hard for an operating system to keep up to date with every access, and some do not try. Metadata modification times are often the same as creation times, but can differ for many forensically uninteresting reasons such as name and access-control changes. Nonetheless, all four times can provide useful information; chapter 10 discusses them in more detail.

Figure 11 shows the overall distribution of file times for the TH drive. It shows that the drive was primarily used 2002-2010 though some files were put on it earlier. Note the spikes in access times in 2010 and 2012 when the files were probably being copied before selling the drive. Blue is drawn first, then green, and then red, so that red can overwrite green, and both red and green can overwrite blue. There isn't much green in the Figure because many creation times were the same as the modification times, and were mostly software files and Web downloads.

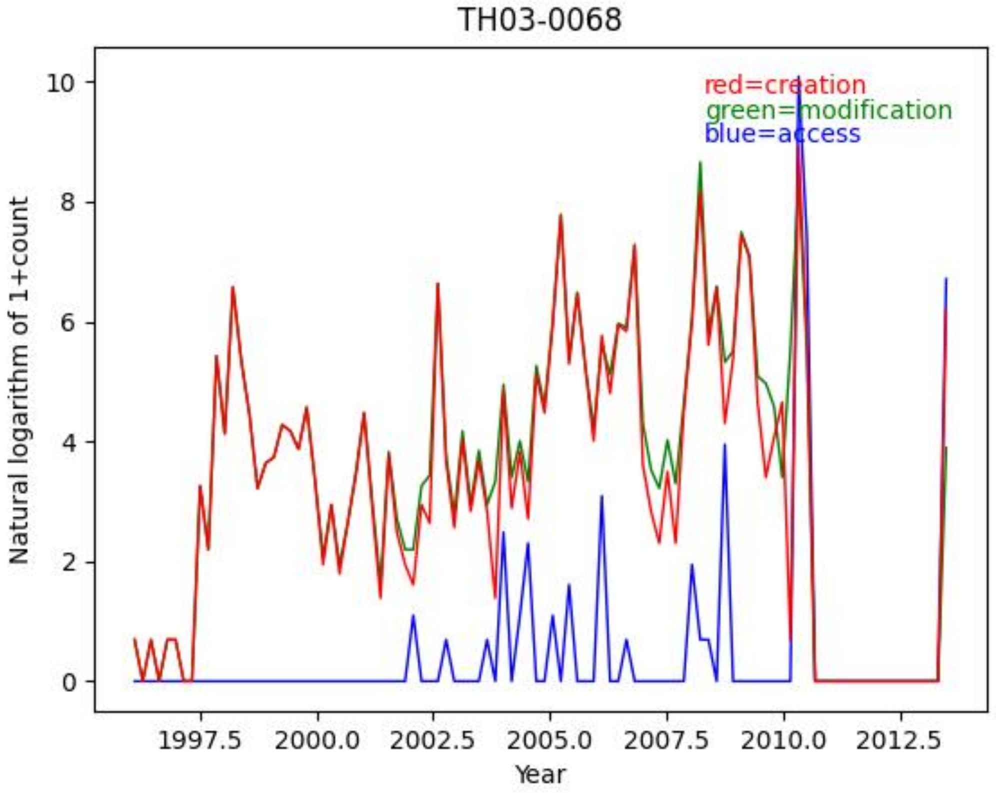


Figure 11: Overall histogram of times of the TH03-0068 drive.

It is often useful to look at time distributions at several levels of granularity to learn more about the type of use. Figure 12 shows the histogram of times within the week for this drive. Daily patterns can be seen, with spikes in use on Saturday, Monday, and Tuesday evenings. ("Thursday midnight" means the first hour on Thursday, so the first day includes all of Thursday.) The Saturday and Tuesday spikes are accompanied by increased user activity, but the Monday spike is not, suggesting it represents automatic downloads whereas the others represent deadlines for employees, assuming what we said in section 4.6 that this was a business computer. Daily and weekly data can benefit from application of the Fourier transform to identify periodicities that suggest regularly scheduled tasks in a user's life or an organization's regular activities. It is also useful to find periodicities to make it easier to see suspicious activity at unusual times like the middle of the night.

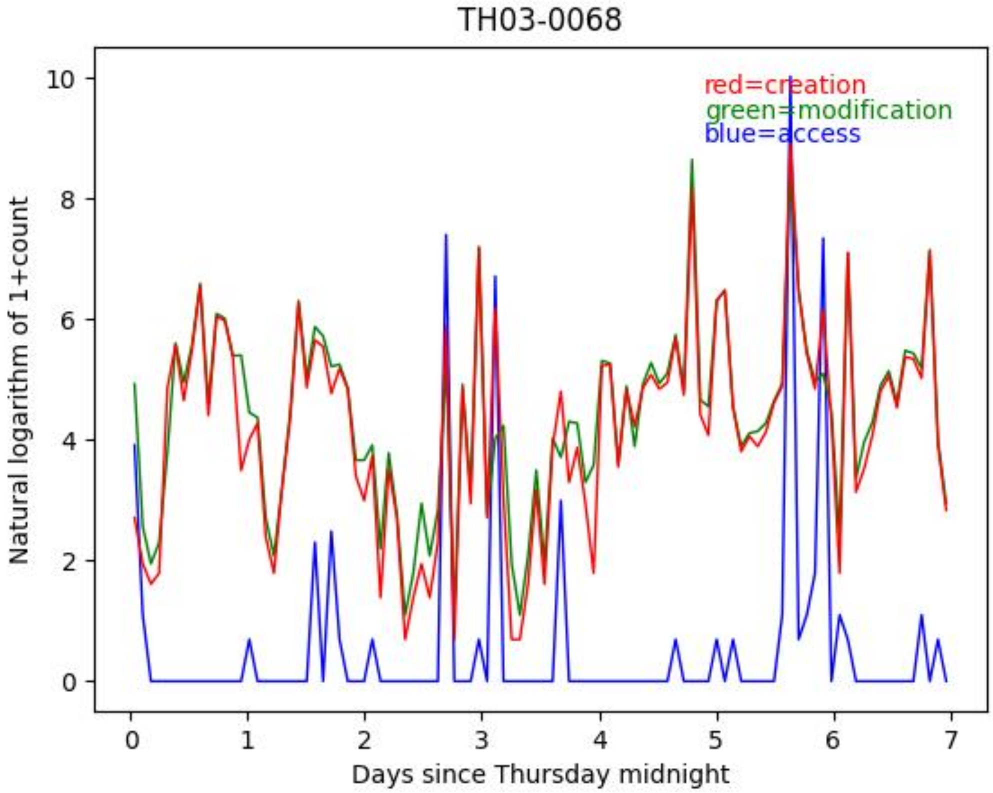


Figure 12: Weekly pattern of use of the TH03-0068 drive.

Figure 13 shows the daily pattern over all files of the drive. Creation and modification times track one another well, suggesting users create many files that are never modified. Two big spikes in access time occur in early in the morning, probably automatic backup, and 1600-1800 is probably users finalizing and saving their work for the day, or sending it out. Some activity occurs through the night, suggesting automatic monitoring on ongoing processes like building sensors or orders being received from around the world.

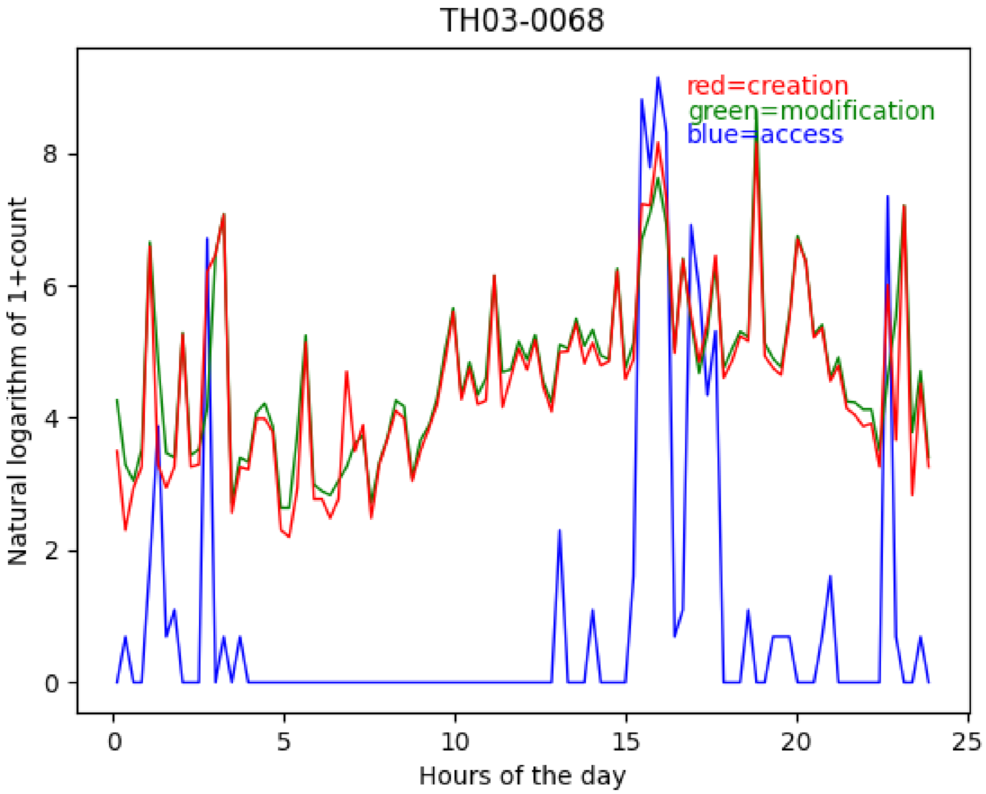


Figure 13: Daily pattern of use of the TH03-0068 drive.

Finally, Figure 14 shows the hourly pattern of the drive. This is pretty random, though it shows a few more accesses occur early in each hour. A peak at the top of the hour on the left side represents some timestamps that have been rounded to the nearest hour by lazy software.

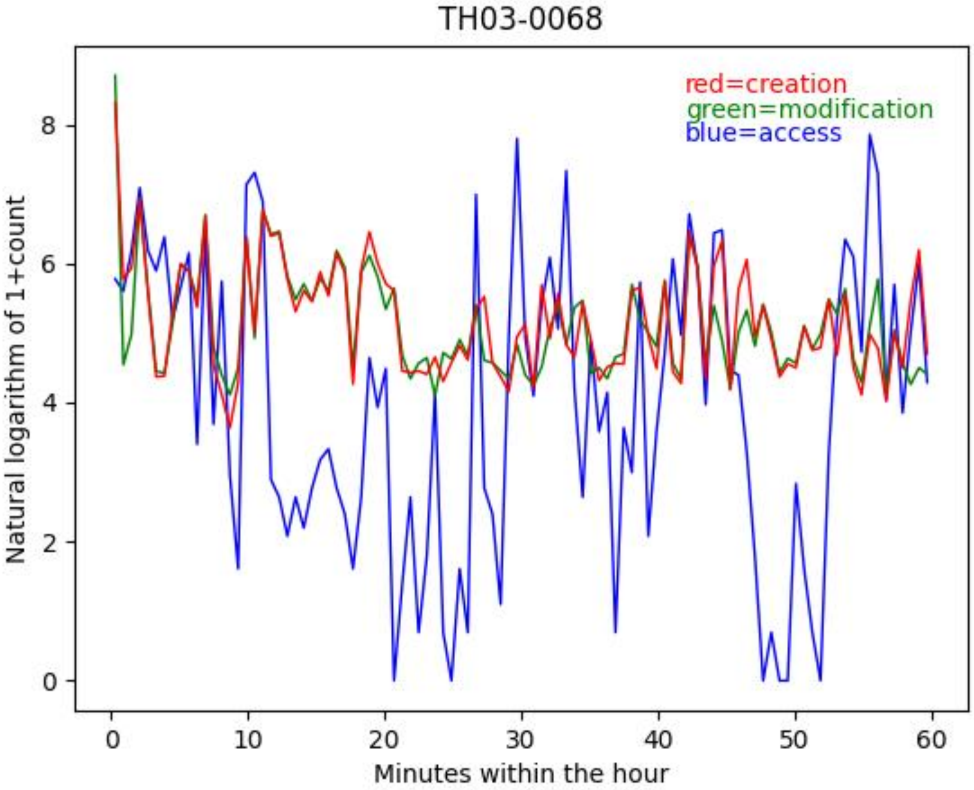


Figure 14: Hourly patterns of use of the TH03-0068 drive.

4.7.2. Other interesting distributions for the drive

We can also plot other distributions of numeric quantities on the drive, like file sizes (Figure 15); Appendix A has a Python program that creates these. Chapter 2 showed the distribution for the whole corpus, but the peak for drive TH is sharper and departs more from a normal distribution. This shows that standard software practice tends to create a narrower range of file sizes than use might suggest. Note also a more obvious tailing off to the right, indicating that drive TH has more large files than a normal distribution would predict. Note also that more bumps on both sides occur when the amount of data is small and allows for more variations in the logarithms of the counts.

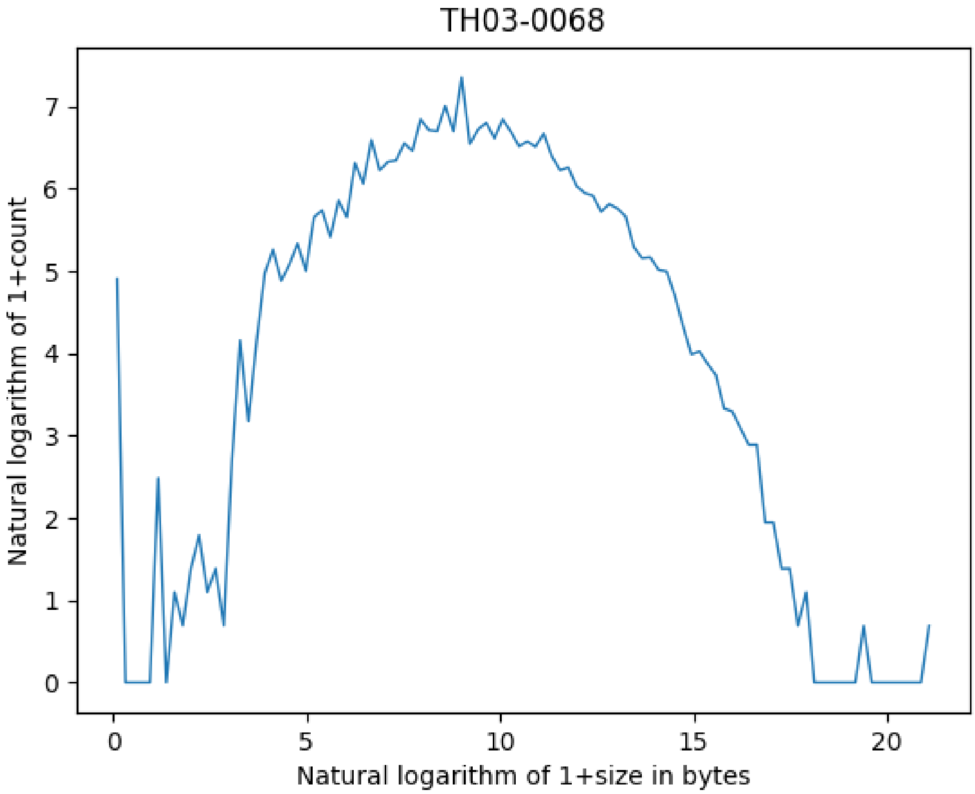


Figure 15: Distribution of the logarithm of file sizes on drive TH03-0068.

Quick Question 4-2: Why are there so many zero-size files instead of just having the file deleted?

Figure 16 shows the distribution of file depths in the file hierarchy (the number of subdirectories on the full path to the file). This is typical of computers as most software is no more than 8 levels deep. Deeper files on drives usually indicate some kind of specialized data collection.

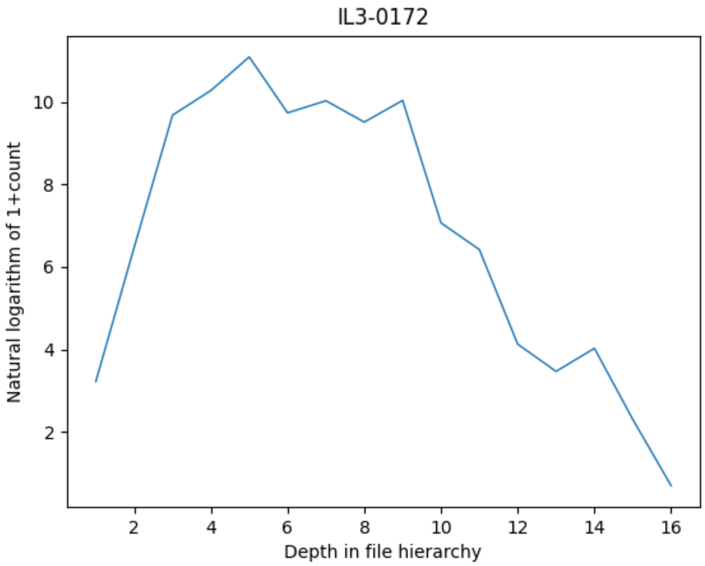


Figure 16: Distribution of depths of files on TH03-0068.

Figure 17 shows the distribution of “inode” numbers for the files this drive. Inode numbers are integers usually assigned in the order in which files were created on the secondary storage, so dips indicate deletion of large blocks of files originally created at the same time and likely interrelated. For this drive, the deleted files were probably user files since we saw so few when viewing the file hierarchy. Note that scattered gaps in the inode numbers do not necessarily means deletions occurred at different times, as the inode numbers corelate only with creation times; often large deletions on a drive occur together when a user is repurposing or selling the drive.

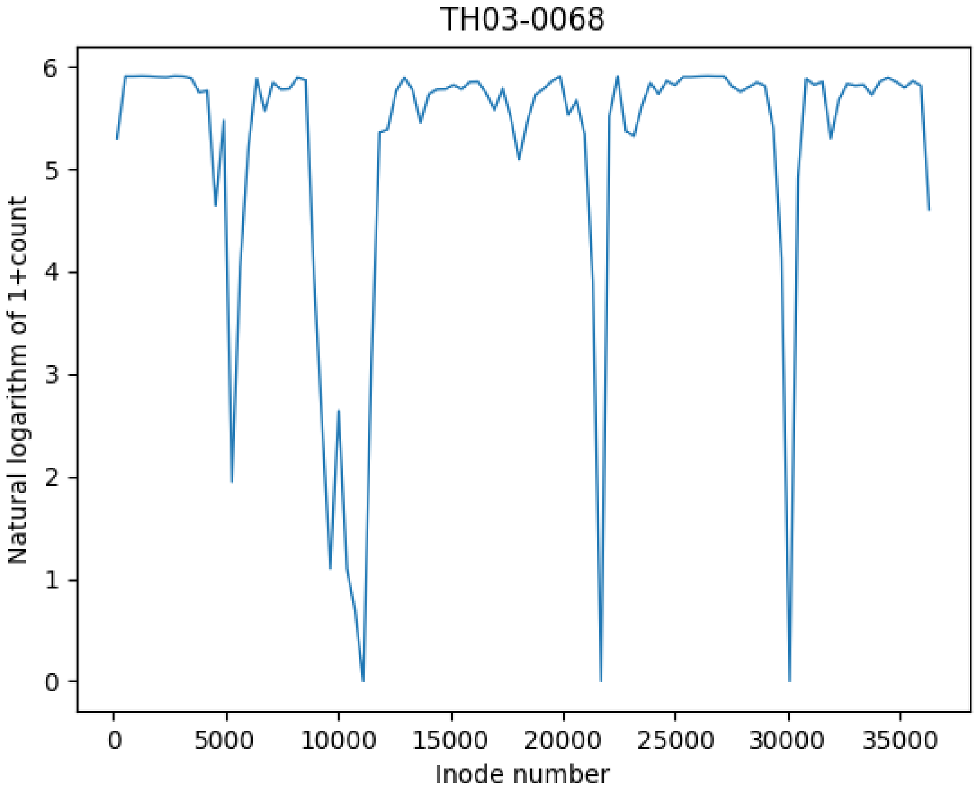


Figure 17: Distribution of inode numbers on TH03-0068.

Figure 18 shows the distribution of the number of files in our corpus having a matching hash value to a file on this drive, a rough indication of the popularity of the files on this drive. A dramatic transition between unpopular files and popular ones occurs at around a logarithm of 3.5, corresponding to 33 drives. Apparently nearly all the files remaining on the drive are files occurring on at least 33 drives in the corpus, so most user files have been deleted.

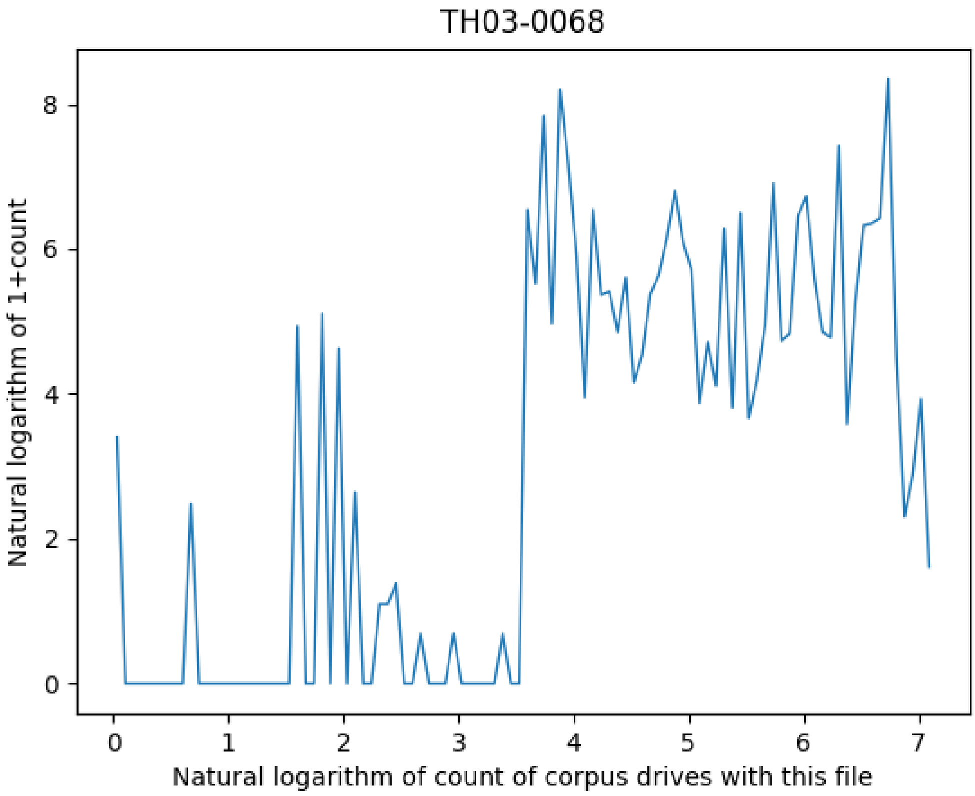


Figure 18: Distribution of number of drives having a file with the same hash value as one on TH03-0068.

Finally, we can examine the distribution of file types for the drive. Here are the ratios of fractions of file extension types of TH03-0068 to averages for our corpus. Large numbers indicate file types that TH03-0068 has in an unusually large number; small numbers indicate file types that it has in an unusually small number. Noteworthy types are the operating system (low), graphics (high), temporary files (very low), Web files (low), general documents (high), spreadsheets (low), links (high), help (high), executables (high), geographic (low), dictionaries (high), queries (high), forms (high), and games (low). These are also consistent with a business machine on which many sensitive files have been deleted.

- none: 1.02561
- operating system: 0.40541
- graphics: 1.78243
- camera images: 0.98119
- temporary files: 0.00797
- Web: 0.34323
- general document: 1.71126
- Microsoft Word: 0.58553
- presentations: 0.93686
- database: 1.29572

- other Microsoft Office: 1.00431
- spreadsheets and business: 0.37474
- email and messaging: 0.53535
- links: 1.5302
- compressed or encoded: 0.54831
- help and training: 5.06527
- audio: 0.67888
- video: 0.51101
- program source: 0.17935
- executables: 1.98524
- disk image: 1.3116
- data interchange: 1.60106
- log: 0.52905
- geographic: 0.04164
- copies and backup: 0.27613
- dictionaries: 16.11392
- query: 8.48417
- codenumber: 0.03321
- index: 0.06445
- form: 13.49073
- configuration: 2.21521
- installs and updates: 3.07624
- security: 0.70785
- known malicious: 0.0
- network: 1.21791
- multipurpose: 1.16016
- encrypted: 0.45802
- hardware: 2.40906
- unassigned: 0.0
- games: 0.11207
- engineering: 0.0876
- science: 0.25944
- signals: 0.0
- virtual machine: 0.56211
- miscellaneous: 0.0

4.7.3. Classifying users

The informal analysis we just described of the use and user of a drive can be made more systematic. A first step is to classify the drive as a business user, home or personal user, or server machine.

Daily timestamp counts (specifically, the sum of creation and modification counts) for five example drives in our corpus are shown in Figure 19, and their hourly totals are shown in Figure 20. It can be seen that drive 29 (in purple) is hardly used on weekends and at night, so it is likely a business computer. Drive 403, on the other hand, is used more on weekends and nights (note the sharp peak at 2000), so it is likely owned by a home user who works during the day. Drive 994 was unused on Sunday, but is used during evening hours on other days of the week, so it is likely a business like a restaurant that is only open in evenings. Drives 695 and 855 have a more even pattern than the others, suggesting they are server machines, with 855 in particular having international use to explain its activity around the clock.

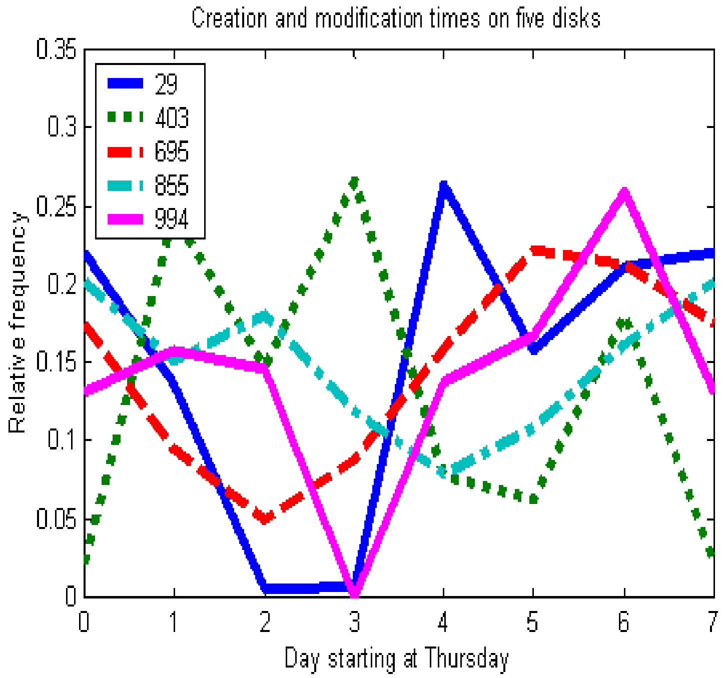


Figure 19: Daily timestamp totals (sum of creation and modification counts) for five sample drives.

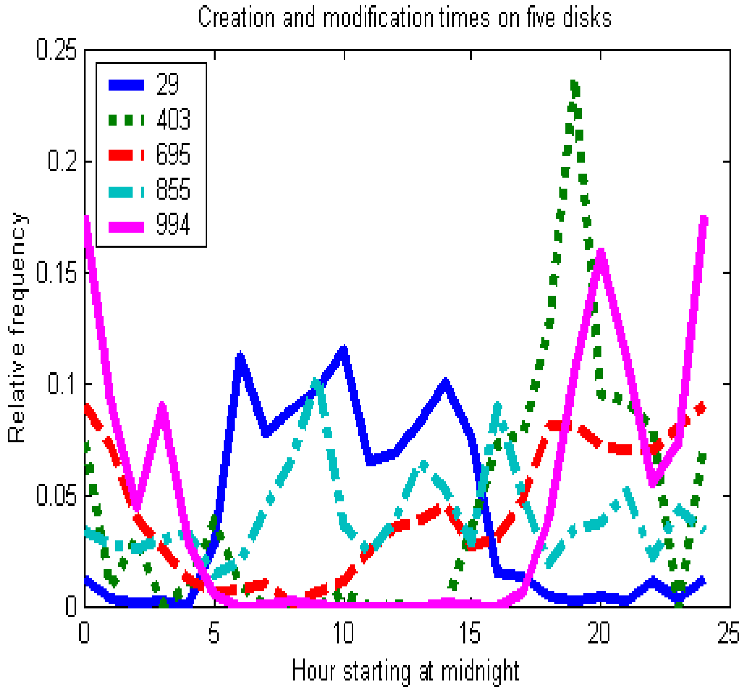


Figure 20: Hourly timestamp totals (sum of creation and modification counts) for five sample drives.

We see that good clues for general classification are the timestamp count ratios between day and evening, day and night, and weekday and weekend. We can measure the distance between this three-element vector and ideal ratios to find the closest match (a form of *case-based reasoning* from artificial intelligence). Our experiments with a sample of drives suggested the following ideal ratios:

Weekday ratio	**Day ratio**	**Evening ratio**	**Case description**	**Inferred disk count in a sample**
0.8	0.8	0.0	Business user	146
0.8	0.0	0.8	Evening business user	115
0.8	0.4	0.4	Day-evening business user	66
0.8	0.0	0.0	Night business user	5
0.3	0.3	0.6	Home user	57
0.8	0.3	0.6	Weekday evening user (either business or home)	65
0.2	0.45	0.45	Day-evening home user	1
0.2	0.1	0.1	Night home user	21
0.5	0.5	0.5	Local server	7
0.5	0.33	0.33	International server with 24-hour use	81
-	-	-	Remainder	15

Classification of users can use other kinds of data besides timestamps, and combine the evidence using methods of the machine-learning area of artificial intelligence. Figure 21 shows a possible taxonomy of users that could be learned from their distribution of file types. Digital professionals can be easily distinguished from amateurs from the kinds of software they run, and subtypes of each can be recognized by the particular software they use. Machine-learning methods can help establish statistical characteristics of the user types, possibly from the words they use as discussed in chapter 9.

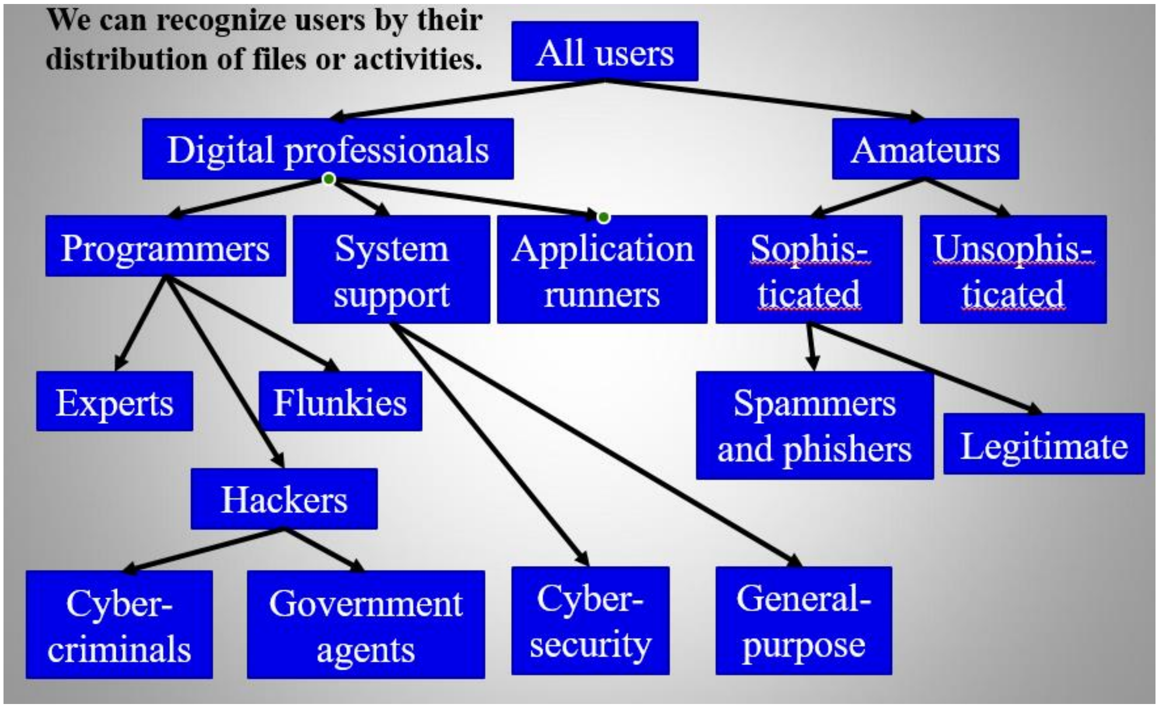

We can recognize users by their distribution of files or activities.

Figure 21: A taxonomy of users of digital technology.

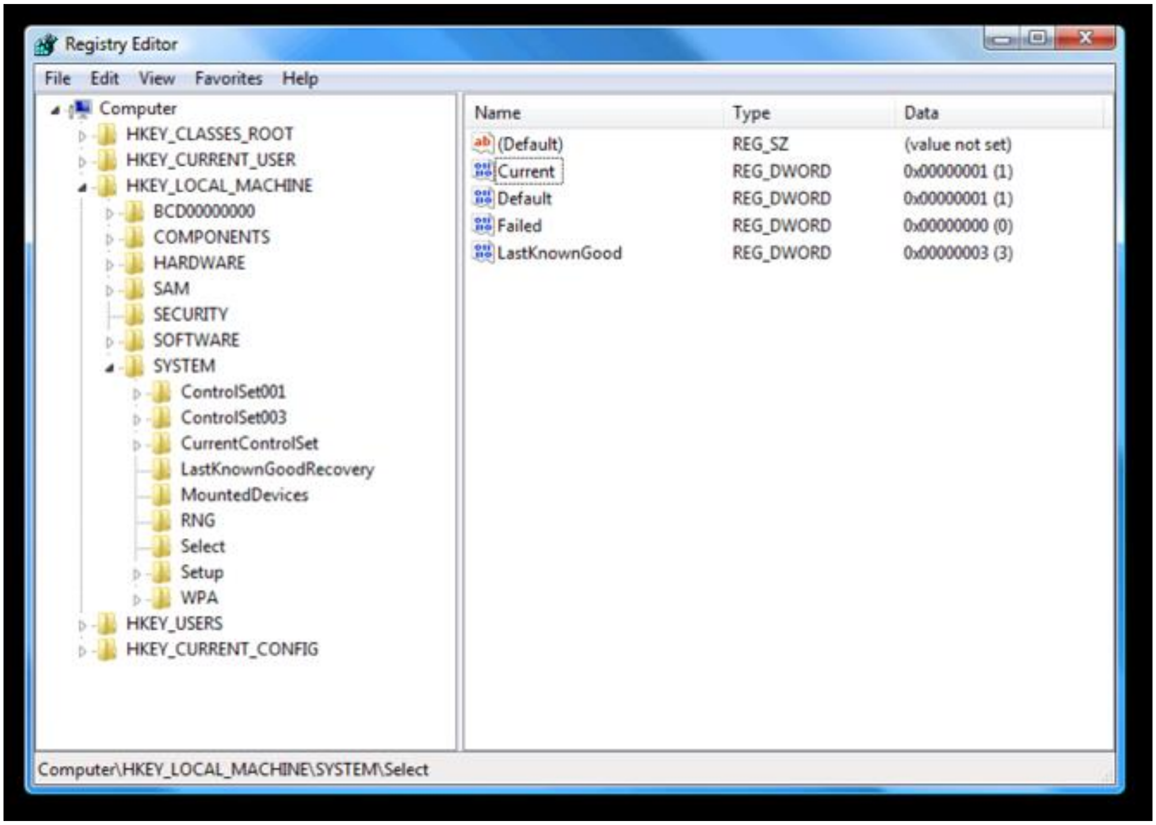
4.8. The Windows Registry

Additional useful information about a file system is stored with the operating system. The Microsoft Windows operating system has made this easier by its “Registry” which you can access through the “regedit” tool. The Registry is a kind of database of metadata about a particular system. It includes installation information, parameter settings (“configurations”), logs, and system-monitoring data. Some of this can help in investigation. Bear in mind that Microsoft makes it difficult to use third-party tools to access the registry for security reasons, so you generally must use a Microsoft tool.

The Registry is especially useful for malware investigations, where you can see what has been installed recently and what software parameters have been changed. But it is also useful to check what software is present and what activity has occurred recently, which helps many investigations.

Figure 22 shows an example top-level screen for the Registry as viewed by the Regedit tool used on a 2010-manufacture Windows system. Items shown are:

- HKEY_LOCAL_MACHINE or HKLM: Settings specific to this computer
- HKEY_CURRENT_CONFIG or HKCC: Information generated about the current session
- HKEY_CLASSES_ROOT or HKCR: Information about registered applications
- HKEY_CURRENT_USER or HKCU: Settings specific to the current logged-in user
- HKEY_USERS or HKU: Settings for all users
- HKEY_PERFORMANCE_DATA : Runtime performance data.
- HKEY_DYN_DATA: Data about hardware devices



- Computer
  - HKEY_CLASSES_ROOT
  - HKEY_CURRENT_USER
  - HKEY_LOCAL_MACHINE
    - BCD00000000
    - COMPONENTS
    - HARDWARE
    - SAM
    - SECURITY
    - SOFTWARE
    - SYSTEM
      - ControlSet001
      - ControlSet003
      - CurrentControlSet
      - LastKnownGoodRecovery
      - MountedDevices
      - RNG
      - Select
      - Setup
      - WPA
  - HKEY_USERS
  - HKEY_CURRENT_CONFIG

Name	Type	Data
(Default)	REG_SZ	(value not set)
Current	REG_DWORD	0x00000001 (1)
Default	REG_DWORD	0x00000001 (1)
Failed	REG_DWORD	0x00000000 (0)
LastKnownGood	REG_DWORD	0x00000003 (3)

Computer\HKEY_LOCAL_MACHINE\SYSTEM\Select

Figure 22: Example top-level view of the Microsoft Registry Editor.

Registries are big; 400,000 registry cells was typical in Windows 7, and it has increased since then. But often forensic analysis needs focus on only a few things:

- Most-recently-used (MRU) lists
- USB devices attached
- Software that starts automatically
- Wireless nodes referenced

Most-recently-used lists can quickly indicate user priorities and activities in the form of the files and directories the user was using recently. Often separate lists are kept for different software. Data for Microsoft-related software is stored in the Windows Registry, and is displayed to the user when viewing directories; data for other software is usually stored in their installation directories. Most-recently-used lists can especially help in studying exploits to see what the attacker was changing or stealing, since attackers do not waste time on a system.

4.9. Exercises

4-1. (*) You are investigating possible financial fraud in an organization. Which metric would be the most useful in finding relevant drives?

A. Count of document files on the drive
B. Fraction of document files on the drive
C. Count of spreadsheet files on the drive
D. Fraction of spreadsheet files on the drive
E. Standard deviation of modification time minus creation time

4-2. (*) We suspect that a spy exfiltrated files from a Windows system with a USB storage device. Where is it best to look?

A. Windows event log
B. Windows Internet Information Service (IIS) logs
C. Windows Registry under CurrentControlSet/Enum/USBSTOR
D. Windows Registry MRU (most recently used) data
E. Windows Registry under REG_LINK

4-3. (*) Answer each of the following questions in 50 words or less.

(a) Suppose you discover that a workstation drive has a high proportion of picture files, but no picture-processing software other than an image viewer. What can you conclude about the purpose of the drive?

(b) Suppose you discover that a workstation drive has many similar files with different dates. What can you conclude about the purpose of the drive?

(c) Suppose you discover that a workstation drive has many files with unusual extensions that you don't see often, and several kinds of unfamiliar software. It also has many files containing long lists of numbers that do not vary much through the rows of the files. What can conclude about the purpose of the drive?

(d) Suppose you discover that a drive from a personal phone has hundreds of records of phone calls and emails every day. What can you conclude about the purpose of the drive?

4-4. Suppose you are interested in determining what time of day a smartphone has generally been used. Which timestamps on files are the most useful (of creation, modification, access, and metadata changes) and why? (100 words or less)

4-5. Get a hex editor such as the Neo Hex Editor (https://www.hhdsoftware.com/free-hex-editor) or Hex Fiend for Macintosh. These display the contents of a file in hexadecimal, transcribing any printable characters (usually on the right side of its window). Download from the Web the file mysteryfiles22a.zip from the course materials site. This contains five files whose names were changed from their originals.

(a) Examine the files with the hex editor. Identify the major sections of the file based on the byte types and their pattern, and give the address ranges of each section within the file, and try to find in particular the header and data sections. Note those sections that are ASCII characters, constant values, apparently random values, and blanks; these observations provide clues to the purpose of the file. Also note any recognizable ASCII strings that provide clues. You may not look up any signatures, but must describe what you see in the file.

(b) Using your analysis of part (a), propose the most reasonable interpretation of what kind of file it is and what purpose it serves, giving your reasons. It may help to look up on the Web particular features in the file, but refer primarily to evidence in the file and not the Web. Then make reasonable guesses about what the major parts of the file are. You may not use any automated file-identification tools for this problem.

4-6. Follow the instructions for problem 1 but use the file mysteryfiles21b.zip instead.

4-7. Follow the instructions for problem 1 but use the file mysteryfiles20.zip instead. Then answer the following additional questions.

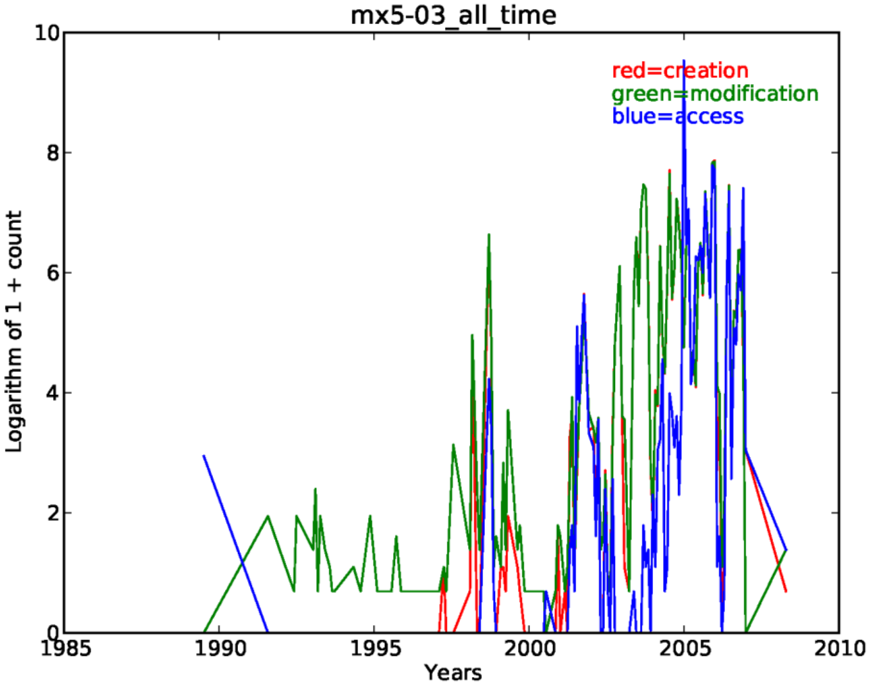
(c) File mystery23.mys has some parts with larger-than-usual numbers of "00" bytes like addresses b400-b800, c360-c3400 and cce0-ce80. What do they likely represent?

(d) File mystery25.mys has many "FF" bytes up until address 00027CF0. What do you think they represent? Could this be evidence of deception?

4-8. (a) Research the file extensions "ock", "bpi", "networkanalysis", and "xdp14" online by finding at least three sources describing each if possible. Summarize what kind of file they think each of them is.

(b) The file http://faculty.nps.edu/ncrowe/coursematerials/cs4677_18_hw2_p4.txt lists all occurrences of these file extensions in our main corpus. The first column is the MD5 hash code, the second is the drive name, the third is the full path to the file, the fourth is the inode number, and the fifth is 1 if the file is not deleted. Describe what additional insights about the file types that you get from this information, and note use that the online resources missed.

4-9. The picture below shows to overall activity for a Mexican drive in our corpus. The counts are per month. The vertical axis is the natural logarithm of one plus the count, and note that the natural logarithm of 1 is 0 and the natural logarithm of 15 is 2.7.



(a) (10 points) When did users likely start using the drive? Give your reasons.

(b) (10 points) What was likely happening 2003-2004 when modification times were more frequent than access times? Give your reasons.

(c) (10 points) How do you explain the first access times at the end of 1989? Give your reasons.

4-10. (*) For this and the next two questions, you will find on the course site a zip file called cs4677_hw2_files_23a.zip. It contains three directories containing drive-image data. For this and the next two questions, each of the three directories contains a listing of the files on a drive, a histogram of file types, a listing of email addresses found on the drive, and graphs of six important kinds of data about the drive. The columns of the "bigtable" files are MD5 hash value, size, whether undeleted (1) or deleted (0), creation timestamp, modification timestamp, access timestamp, full path, and "popularity" (number of drives in our corpus having a file with the same hash value). The listing of files in problems 1 and 2 includes only the undeleted files. Your answers to each part should not exceed 100 words, and much less is needed for some questions.

For 4-10 specifically, answer these questions as best you can for drive BS001-0034 (from the Bahamas) and give your reasoning. It may help to look up some things online, but try to focus on the data in the files.

(a) When were the starting time and the ending time of the period of usage of the drive by its users to the nearest year?

(b) What distinctive patterns were in the usage per day, week, or hour? Note that week plots start with all of Thursday, and year labels on the overall plot are centered on July 1. Some of the plots may show errors in estimating the time-zone correction, so note if that may have happened. Note that blue is written first, then green, and then red, so some of the blue and green may be covered up by red.

(c) How normal (in the mathematical sense) was the distribution of file sizes? What likely explains the differences from a normal distribution?

(d) What did the distribution of inode numbers tell you about usage of the drive?

(e) What major kinds of applications (software) were loaded on the drive, and which were unusual? It may help to look the unfamiliar names up in Google.

(f) Did you see anything distinctive in the distribution of file types? Bear in mind that less-common types have more variation.

(g) A file lists the email addresses found on the drive and their counts; also there may be a listing of phone numbers. What do they tell you about the usage of the drive?

(h) How many human users did the drive have? What kind of user or users were they?

(i) What kind of usage, did the drive have, and why? Try to distinguish business from personal usage. If business usage is apparent, try to distinguish what kind of business as precisely as you can.

4-11. (*) Answer questions (a)-(i) in problem 4-10 for the IL004-0010 drive (from Israel). It also had some phone numbers.

4-12. (*) Answer questions 1(a)-(i) for the SG001-7013 drive (from Singapore). Note that almost all its files are deleted.

4-13. (12 points) The file size given in the DFXML metadata for a file may differ from the size given in the “byte_runs” metadata for that file, the data indicating the actual addresses occupied by the file in secondary storage. Why?