A High Level Analysis of Deletion Habits on USB Thumb Drives

Mojtaba Al Fardan
Introduction and Motivation

● One important aspect of personal privacy is how to completely erase data from our personal storage devices.

● Normal deletion methods available from the operating systems are ambiguous and do not describe what they actually do.
Introduction and Motivation

- The purpose of my research is:
  - To find out the deletion habits of people
  - To find out what types of data that can be recovered

- This research is done on USB thumb drives
Drive Categories

- We purchased flash drives from different sellers in Amazon/Ebay.
- We also ran a buy back program where we got old USB drives from student/faculty/staff in exchange for a new one
Drives Analyzed

- 129 Amazon/Ebay USB flash drives
- 114 from the Buyback program
  - Had a survey component
Deletion Methods

- We examined what happens when using different deletion methods on FAT formatted drives
  - Normal deletion (trash can and delete key)
  - Quick format (or mkdosfs in Linux)
  - Full format (zero overwrite)
  - Other deletion methods
Normal Deletion

- When the user uses the delete key, what that does is replace the first 8 bits of the file name to E5 in hex.
Normal Deletion

Before

After
Quick Format

- When the user uses quick format all that does is delete the FAT tables
  - FAT table contains addresses of data for files and directories -- **not data**!
Full Format

- When the user uses full format option what that does is rewrite everything with zeros
Other Deletion Methods

• Other deletion methods:
  ○ Random characters overwrite
  ○ Single or alternating character overwrite
Data Classifications

- **Illicit**: Anything illegal or obtained by an illegal means or pornographic materials.
- **Corporate Confidential**: Anything related to companies but is not public, like employee data.
- **Corporate Commercial**: Anything related to companies that is public, like general information about the company or announcements or ads.
Data Classifications

- **Personal Identifiable Information**: Anything that identifies or corresponds to a person.
  - We used the National Institute of Standards and Technology definition of PII which includes full names, addresses, pictures of faces, identification numbers, social security numbers, credit card numbers, etc.
Process Overview

1. Identify metadata
2. Classify mounted data
3. Copy drive image
4. Forensic analysis
5. Record results
Identifying Metadata

- Take a picture of the drive
- Record drives details
  - size
  - brand
  - model
  - category
  - where and when it came from
If non-deleted data can be found, we classify the data into the categories mentioned earlier.
Copy Drive Image

- We create two copies:
  - Safe copy for backup
  - Working copy for forensic analysis
Forensic Analysis

- Manual analysis of deletion methods used (described earlier)
- Automatic recovery of files using photorec
  - Photorec recovers files not just photos
- Classification of files found from photorec
Record Results

- Enter the results found into the database
- Enter survey results (if applicable) into the database
The Survey
Special Cases

● Sometimes we get drives that have a high amount of sensitive information.
● One example is case B092 where there is a lot of personal identifiable information and illicit data.
Special Case

- Pictures
- Employment history
- Medical record
- Addresses
- SSN/tax form
- Drivers license
- Passport scan
- References
- Military record
- Resume
- Pornographic materials
Results

The Categories of The Data Found On USB Flash Drives

- Illicit
- Corporate Confidential
- Corporate Commercial
- Personal Identifiable Information

The chart shows the distribution of data categories with Amazon/Ebay and BuyBack categories.
Results

DISTRIBUTION OF DELETION METHODS OF AMAZON/EBAY USB FLASH DRIVES

- Never Deleted: 29%
- Delete: 36%
- Quick Format: 20%
- Full Format: 10%
- Other Deletion Methods: 5%

DISTRIBUTION OF DELETION METHODS OF BUYBACK USB FLASH DRIVES

- Never Deleted: 41%
- Delete: 19%
- Quick Format: 28%
- Full Format: 10%
- Other Deletion Methods: 2%
Results

- Number of drives with sensitive information
  - Buy Back drives: 69 of 114 (about 60%)
  - Ebay/amazon drives: 47 of 129 (about 37%)
Results
Results
Results

Males: 39 out of 73 drives had sensitive data on them.

Females: 29 out of 40 drives had sensitive data on them.
Results

Distribution Of Deletion Habits Across 18-35 Age Group

- Never Deleted: 42%
- Quick Format: 33%
- Delete: 14%
- Other Deletion Methods: 11%

Distribution Of Deletion Habits Across 36+ Age Group

- Never Deleted: 14%
- Delete: 26%
- Quick Format: 28%
- Full Format: 6%
### Results

What they said (Did you delete?)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>57</td>
<td>12</td>
</tr>
<tr>
<td>No</td>
<td>22</td>
<td>23</td>
</tr>
</tbody>
</table>
## Results

### What they said (Did you delete?)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>No</td>
<td>39</td>
<td>28</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>What they did (Actually Deleted) (Full Format Or Better)</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>No</td>
<td>68</td>
<td>32</td>
</tr>
</tbody>
</table>
Results

Deletion Methods Of Survey Participant Who Always Care About Deletion

<table>
<thead>
<tr>
<th>Number Of Drives</th>
<th>Data Is Recoverable</th>
<th>Data Is Not Recoverable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deletion Method Categorized By Possible Recoverability
Results

Deletion Methods And Perceptions Of Recoverability

Number Of Participants

Somewhat Hard | Hard To Recover

Data Is Recoverable | Data Is Not Recoverable

Participants Perception
Recap

- External storage devices have been around a long time
- Deletion methods have been around a long time too
- Yet users still don't know what these options mean!
  - Perhaps we need to engineer better deletion methods instead of relying on deletion education
Thank You For Listening

Questions?