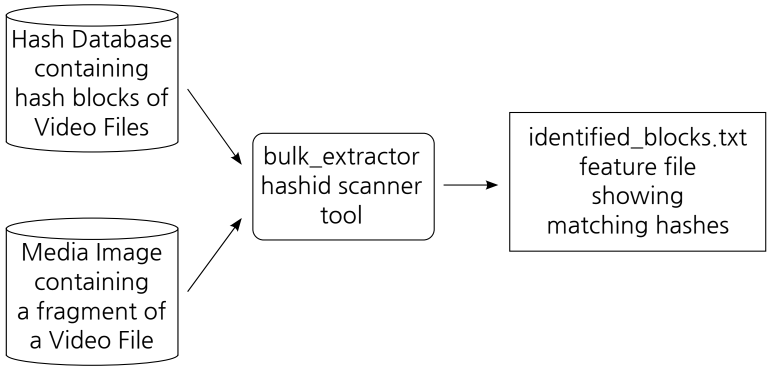
Use a Hash Database to Find Part of a Video File in a Media Image

Workflow:


Hash Database containing hash blocks of Video Files
Media Image containing a fragment of a Video File
bulk_extractor hashid scanner tool
identified_blocks.txt feature file showing matching hashes

Scan a Media Image for parts of a Video file.

Setup: Download the files required for this example:

- Hash Database containing hash blocks of Video Files: `www.digitalcorpora.org/downloads/hashdb/examples/db_of_video_hashes`
- Media Image containing a fragment of a Video File: `www.digitalcorpora.org/downloads/hashdb/examples/media_image_with_fragment`
- *bulk_extractor* built with the *hashdb hashid* scanner: `www.digitalcorpora.org/downloads/hashdb/bulk_extractor-1.4.1-windowsinstaller.exe`

Steps:

1. Install the two `.exe` files.
2. Run *bulk_extractor* built with the *hashdb hashid* scanner:
   ```
   $ bulk_extractor -o outdir -S db_of_video_hashes \
     media_image_with_fragment
   ```
3. View the feature file using an editor or *bulk_extractor Viewer*. For example type `vi outdir/identifie`
   An example match looks like this:
   ```
   102400 a2929e2d838b88973b4c4f3d7a96c6d9 1
   ```

Seeing hash `a2929e2d...` at Forensic path `102400` shows that something matched our database, but what? We find the file that contains this hash using a source lookup using this workflow:

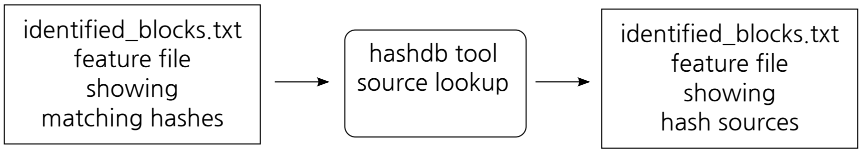

identified_blocks.txt feature file showing matching hashes
hashdb tool source lookup
identified_blocks.txt feature file showing hash sources

Figure 1: Look up the file that has the hash.

Steps:

1. Run the *hashdb* tool to obtain the lookup:
   ```
   $ hashdb get_hash_source outdir/identified_blocks.txt \
     identified_sources.txt
   ```

Now we view file `identified_sources.txt` to see features containing source information, like this one:

```
102400 a2929e2d838b88973b4c4f3d7a96c6d9   repository \
incriminating_video.mpg   8120000
```

indicating that the block was from file `incriminating_video.mpg`. specifically, the block at forensic path `102400` came from the block that is `8120000` bytes into video file `incriminating_video.mp`