

BULK_EXTRACTOR AND MEMORY ANALYSIS

Using Open-Source Tools for Rapid Memory Parsing

ABSTRACT

In digital forensics investigations, there is an increasing need to conduct detailed examinations of memory from servers and other computers. This paper proposes using an open-source tool to analyze these memory acquisitions: Bulk_Extractor.

Greg Tassone, Digital Forensic Investigator Bulk_Extractor and Memory Analysis

Overview

There is an increasing need to conduct detailed examinations of system memory during digital forensics investigations. Due to the complicated structure of such memory, certain tools are better suited for this task. One particularly useful tool for parsing memory acquisitions is the open source "**Bulk_Extractor**."

Bulk_Extractor is a multipurpose artifact scanning program written in C++. It is capable of scanning a memory acquisition, packet capture, disk acquisition files (E01, raw, etc.), single files, or entire directories. It then extracts and stores useful information into *Feature Files* that can be easily viewed with the software of your choice or processed with automated tools. Bulk_Extractor also creates *Histograms* of the features it locates, ranking artifacts by how often they occur. *NOTE: Bulk_Extractor does not acquire memory extractions by itself, but is effective at scanning and finding useful information in acquisitions gathered by other tools.*

Bulk_Extractor was primarily developed by Simson L. Garfinkel from the United States Naval Postgraduate School, along with several other colleagues. This tool is open-source software that is free to use for any purpose, licensed under a BSD-style license¹. It is compiled into different versions compatible with Windows, macOS, and Linux-based systems. It can be used from the command-line, as well as from a Java-based graphical user interface (GUI) to simplify using the many features offered by the tool.

This document will focus on using the GUI application in a Windows environment. For detailed information about using Bulk_Extractor in other environments, please refer to the main documentation located on the project Wiki within the public GitHub repository².

Installation Instructions

Dependencies

The core components of **Bulk_Extractor** can be run from the command-line. However, to run the GUI application, otherwise known as the "Bulk Extractor Viewer" or "BEViewer," you must have a Java runtime environment of version 1.6 or higher installed. If you already have a compatible version of Java installed, you can skip to the section **Bulk_Extractor Download and Installation.**

To determine if you have Java installed, you can check your Apps/Programs section of the Windows *Control Panel*, or you can run the following command from the Windows console ("cmd" window):

java -version

If Java is installed, you should receive a response detailing the active version on your system. openjdk version "1.8.0_322" OpenJDK Runtime Environment (Temurin)(build 1.8.0_322-b06) OpenJDK 64-Bit Server VM (Temurin)(build 25.322-b06, mixed mode)

Otherwise, you will receive an error message similar to:

'java' is not recognized as an internal or external command ... "

¹ License text and description for Bulk_Extractor: <u>https://github.com/simsong/bulk_extractor/wiki/Licensing</u> ² The public GitHub project repository for Bulk_Extractor is: <u>https://github.com/simsong/bulk_extractor/</u>

Java Download and Installation

1. Download Java: First some background: The Java language transitioned to opensource licensing beginning in 2007. As of 2019, there are many certified-complaint implementations of the Java runtime environments that can be used freely for commercial and non-commercial uses. The Apache Adoptium Java project provides such "Open" packages which are built for Windows. The Java version 1.8 runtime (version "8" in the new Java naming scheme) should be used since it is the oldest supported version produced by the Adoptium project. In recent testing, it was the only version compatible with bulk_extractor BEViewer. Download the Temurin 8 (LTS) package:



Figure 1: Adoptium download selector

😹 Eclipse Temurin JDK with Hotspot 8u322-b06 (x64) Setup 🛛 🗌 🗙						
Custom Setup						
Select the way you	want features to be installed.					\sim
Click the icons in th	ne tree below to change the w	ay fea	itures will t	oe installed.		
	vith Hotspot Add to PATH Associate .jar Set JAVA HOME variable		Eclipse Te Hotspot	emurin Developn	nent Kit with	1
	JavaSoft (Oracle) registry ke	ys	This featu hard drive selected. on your h	rre requires 186 e. It has 4 of 4 s The subfeatures ard drive.	MB on your ubfeatures s require 7K	в
<	2					
Location:	C:\Program Files\Eclipse Ado	ptium	jdk-8.0.32	2.6-hotspot\	Browse	
Re <u>s</u> et	Disk <u>U</u> sage		<u>B</u> ack	Next	Cancel	

Figure 2: Temurin JDK installation window

<u>https://adoptium.net/</u>

2. Install Java: Install the downloaded version of the Java OpenJDK even if you already have a different version of Java runtime installed (they can coexist on the same system). You will need Administrator permissions for the complete installation. During the install, customize the options to ensure *all features* are enabled.

Bulk_Extractor Download and Installation

- Download Bulk_Extractor: At the time of this writing, the version 2.0 series of Bulk_Extractor has been released. However, the most current version of BEViewer was last released within version 1.5.5 of Bulk_Extractor; an updated version of BEViewer for the 2.0 series is still in development. Download the version labeled *bulk_extractor-1.5.5-windowsinstaller.exe* from here:
 - https://downloads.digitalcorpora.org/downloads/bulk extractor/
- Install Bulk_Extractor: Launch the installer downloaded in Step #1. You will need Administrator permissions to install it properly. Keep the default options and complete the installation.
- After the installer completes, you should see the new program installed in your Start Menu labeled:

"Bulk Extractor 1.5.5"

Bulk Extractor 1.5.5 Seture Installation Ontions
Burk Extractor 1.5.5 Setup: Installation Options –
Check the components you want to install and uncheck the components you don't want to install. Click Install to start the installation.
Select components to install:
Space required: 46.6MB
Cancel Nullsoft Install System v2,46 Install

Figure 3: Bulk Extractor installation window

Using Bulk_Extractor

Launching the program

- Launch Bulk Extractor Viewer (BEViewer): You should see the new program installed in your Start Menu under "Bulk Extractor 1.5.5" – Launch the program BEViewer with Bulk Extractor 1.5.5 (64-bit). This should open the main BEViewer window.
- 2. **Open the Run window:** Inside the main BEViewer window, click the gear icon on the toolbar. This opens the "Run" window where you can configure scans and initiate a scan.



Figure 4: Bulk Extractor folder in the Windows Start Menu

Bulk Extractor Viewer			1000	\times
File Edit View Bookmarks Tools	Help			
× 🔒 🎭 🛊 🖈	Highlight:	☑ Match case		
Reports	Feature Filter Match case	Image File None Feature File None Feature None Feature None Image		
	Referenced Feature File None Referenced Feature None			
		🖲 Text 🔿 Hex 🕷 🖉 🖿		

Figure 5: BEViewer main window ("Run" option highlighted in the toolbar)

- 3. Select a memory acquisition: Inside the Run window, choose an acquisition file to be scanned. Since this example discusses memory forensics, choose a binary memory acquisition file that has been collected by another tool of your choice. You can either type in the full path to the acquisition file (*labeled as "Image file"*) or click the icon on the right side of the field to use the file-chooser.
- 4. **Configure an output directory and scan settings:** Inside of the Run window, select a directory where extracted artifacts (*Features*) will be stored. Then configure your scan settings to your liking. For this example, you can leave all scan settings at their defaults, but enable all **Scanners** on the right-side panel except *hashdb*, *sceadan*, and *wordlist* since they require additional setup. If you want to carve for JPEG image files more extensively, enable **Use Settable Options** and enter *jpeg_carve_mode=2*:

Bulk Extractor Viewer		Run bulk_extractor	×	×
File Edit View Bookmarks	Tools		_	
× 🚨 🖗 🔶 🔶	5	Required Parameters	Scanners	
	_			_
Reports	Featu	Image nie D: (rancapture)20220329.mem		
	East			
	Featu	General Options		
		Use Banner File	sceadan	
		Use Alert List File		
		Use Stop List File	✓ xor	
		Use Find Regex Text File	✓ accts	
		Use Find Regex Text	i aes	
		Use Random Sampling	✓ base64	
			✓ elf	
		Tuning Parameters	✓ email	
		Use Context window Size	🖌 exif	
		Use Page Size 16777216	🗹 find	
		Use Margin Size 4194304	🗹 gps	
		Use Block Size 512	🗹 gzip	
		Use Number of Threads	🗸 hiberfile	
		Use Maximum Recursion Depth 7	✓ httplogs	
	Refer	Use Wait Time 60	🗹 json	
	Refer	Parallelizing	🗸 kml	
		Use start processing at offset	✓ msxml	
		Use process range offset o1-o2	🗸 net	
		Use add offset to reported feature offsets	🗹 pdf	
			🗸 rar	
		Debugging Options	🗹 sqlite	
		Start on Page Number 0	✓ vcard	
		Use Debug Mode Number	✓ windirs	
		Erase Output Directory	vinlnk	
		Scanner Controls	✓ winpe	
		Use Plugin Directories	winprefetch	
		Use Settable Options jpeg_carve_mode=2	🗹 zip	
		Manage Queue Import Submit Run Cancel		

Figure 6: BEViewer "Run" window with scan options configured

5. Initiate the Scan: Once all settings are configured, click the Submit Run button. This will begin the scanning process, which may take a long time to complete depending on the size of your acquisition file. Once the scan is finished you should see a results window summarizing the operation. You can then close the Scan/Run window. NOTE: Although detailed statistics about the scan are stored within the file *report.xml* in the output directory, you might want to copy the simple summary from this window before closing it.

Julk_extractor Scan	×
Image File 20220329.mem Feature Directory BE_output	
Command	
bulk_extractor -o D:\ramcapture\BE_output -S jpeg_carve_mode=2 -e base16 -e facebook -e outlook -e xor D:\ramcap	oture
Progress 1.78% 14:49:38 Offset 654MB (1.78%) Done in 0:31:58 at 15:21:36	
bulk_extractor version: 1.5.5	
Input file: D:\ramcapture\20220329.mem	
Output directory: D:\ramcapture\BE_output	
Disk Size: 36767268864	
Threads: 16	
Cancel	

Figure 7: Scan/Run in progress - window

wilk_extractor Scan	×
Image File20220329.memFeature DirectoryBE_output	
Command	
bulk_extractor -o D:\ramcapture\BE_output -S jpeg_carve_mode=2 -e base16 -e facebook -e ou	itlook -e xor D:\ramcaptur
Progress Done bulk_extractor scan completed. Report BE_output is ready.	
*****	^
MD5 of Disk Image: 229ff7f3994faee2fa285de51c2ac591	
Phase 2 Shutting down scappers	
Thate 2. Shadding down Soumers	
Phase 3. Creating Histograms	
Phase 3. Creating Histograms Elapsed time: 3206.29 sec.	
Phase 3. Creating Histograms Elapsed time: 3206.29 sec. Total MB processed: 36767	
Phase 3. Creating Histograms Elapsed time: 3206.29 sec. Total MB processed: 36767 Overall performance: 11.4672 MBytes/sec (0.716701 MBytes/sec/thread)	
Phase 3. Creating Histograms Elapsed time: 3206.29 sec. Total MB processed: 36767 Overall performance: 11.4672 MBytes/sec (0.716701 MBytes/sec/thread) Total email features found: 12496	
Phase 3. Creating Histograms Elapsed time: 3206.29 sec. Total MB processed: 36767 Overall performance: 11.4672 MBytes/sec (0.716701 MBytes/sec/thread) Total email features found: 12496 Attempt to open D:\ramcapture\20220329.mem	

Figure 8: Scan/Run completed - Results window

Examine Bulk_Extractor Output with BEViewer

After successfully running **Bulk_Extractor** there will be "Feature Files" collected within the output directory. You can then continue to use **BEViewer** to examine these files in an organized way. NOTE: For this to be effective, the analyzed acquisition file(s) must still be available in the original location, as **BEViewer** will display raw data from the acquisition during your examination.

1. **Open the Scan Report:** On the top left of your **BEViewer** window, there should be a folder icon with the name of your configured output directory. This is your Scan Report folder from the previous steps. Select the report folder to load the report contents into the viewer, as the following figure shows:

Rulk Extractor Viewer			<u> </u>	×
File Edit View Bookmarks Tools	s Help			
× 🔒 🎭 🛊 🔺	Highlight:	✓ Match case		
Reports Rep	Feature File Match case Feature File None Referenced Feature File None Referenced Feature File None Referenced Feature None	Testre File None Feature None Feature None Testre None Testre None Testre None Test O Hex		

Figure 9: BEViewer main window with "BE_ouput" Scan Report selected/loaded

- 2. **Viewing Feature Data:** Selecting a particular report feature in the left pane will load relevant information into the panes toward the right.
 - For example, selecting the "domain.txt" report feature will load the contents of that feature file into the upper-middle pane. Then, to examine an individual result, you can select an entry within this upper-middle pane. This will open the raw acquisition file in the large pane on the right, highlighting the relevant data for that entry.
 - Note: the number listed to the left of each entry is the location/offset of that entry in the raw acquisition. This offset is expressed as the number of bytes from the start of the file where the entry is located.

Bulk Extractor Viewer			— [3	×
X 🔒 🍇 🚖 🔺	r Help Highlight:	☑ Match case			
Reports BE_output BE_OUTPU	Feature File Match case Feature File domain.txt 35066554 1.0.0 35066554 1.0.0 1115561 www.microsofl.com 1995305 dsymcb.com 1995406 ssymcd.com 1995406 ssymcb.com 199547 dsymcb.com 1995484 ssymcd.com 1996513 dsymcb.com 1996544 ssymcate.com 1996554 ssymantec.com 1996567 schemas.dmff.org 3219703 schemas.dmff.org 3219711 schemas.dmff.org 3219703 schemas.dmff.org 3219711 schemas.dmff.org 3219712 schemas.dmff.org 3219713 schemas.dmff.org 3219714 schemas.dmff.org 3219715<	Image File forensiclaptop.mem Preature File domain.bt Preature Tile www.microsoft.com Inage 1114412 1114112			~

Figure 10: BEViewer window displaying results from a single report feature

- This process is enhanced when selecting a "histogram" report feature, such as *domain_histogram.txt*. Once it is loaded, the histogram results are shown in the upper-middle pane. Selecting an entry from this upper-middle pane loads the specific results of each histogram entry into the lower-middle pane, as shown in the following figure. These results will continue to be correlated to the raw data from the acquisition file in the right-side pane.
- Note: In this example, the "n=[number]" on the left side of the upper-middle pane is the number of hits located for that entry during the scan. The location/offset will be shown to the left of each entry in the lower-middle pane.

Note: Match case Partie: Partie: Match case Partie: Parti: Parti: Parti: <th>Bulk Extractor Viewer File Edit View Bookmarks Tool</th> <th>s Help</th> <th></th> <th>- 0 X</th>	Bulk Extractor Viewer File Edit View Bookmarks Tool	s Help		- 0 X
Generating Service Scheme File Generating Service Scheme File File details in the server scheme Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Generating Scheme File Genea	× 🔒 🍕 🚖 ★	Highlight:	Match case	
🔿 Text 🖓 Hex 🖌 👘	Reports BE_output - @ aes_keys.bxt - @ alerts.bxt - @ con_bit - @	Feature Filter Match case Histogram File domain_histogram.txt n=10619 www.wicrosoft.com n=2936 cil.nicrosoft.com n=1497 schemas microsoft.com n=1497 schemas microsoft.com n=1130 www.wickigeert.com n=1063 cil.4.digicert.com n=901 40.0 n=685 www.bige.com n=637 libensing.microsoft.com n=635 go.ncbastl.comC n=456 op. digiert.comC n=459 oop. digiert.comC n=3519 www.wisco.com n=3519 www.wisco.go 5595341 www.wiscog 55953534 www.wiscog 73126359 www.wiscog 732233738 www.wiscog 73233501 www.wiscog 73233502 www.wiscog 73233503 www.wiscog 73233503 www.wiscog	Image File forensidaptop.mem Feature File forensidaptop.mem Feature File moman.txt Forensic Path 73124339 Feature File www.w3.org Image 731244160 To'.Hf0Kz, PT.gIKu.*w 73124224 h3.6.S.R\/951gy!el8.E, IF77., 5Y.	· · · · · · · · · · · · · · · · · · ·

Figure 11: BEViewer window displaying results from a "histogram" report feature

3. Filtering / Searching Feature Data: Due to the large number of entries in certain report features, BEViewer has built-in searching and filtering capabilities. To activate a filter, begin typing in the "Feature Filter" box in the upper-middle pane.

Bulk Extractor Viewer	s Help			- 0 X
× 🔒 🍕 🚖 🔺	Highlight:	Ma	tch case	
Reports BE_output aes_keys.bt aes_keys.bt con.bt con.bt con.bt domain_histogram.bt domain_histogram.bt email_ohistogram.bt email_ohistogram.bt email_histogram.bt email_histogram.bt facebook.bt hex.bt pi.btt facebook.bt pi.btt facebook.bt bex.bt gi.btt facebook.bt ul_stagram.bt ul_st ul_st	Highlight Feature File Match case digit Interpret file Histogram. Ele domain_histogram. Ext n=1160 cfl:3 digicert.com n=1161 cfl:4 digicert.com n=062 cfl:4 digicert.com n=063 cfl:4 digicert.com n=146 cosp. digicert.comOC n=120 cosp. digicert.comOL n=120 cosp. digicert.comOL n=120 cosp. digicert.comOL n=120 cosp. digicert.comOL n=121 cosp. digicert.comOL n=131 cosp. digicert.comOL n=14 cosp. digicert.comOL n=11 cosp. digicert.comOL 12997175 cosp. digicert.comOL 12997175 cosp. digicert.comOL 4201712447 cosp. digicert.comOL <t< th=""><th>Image File Feature File Feature File Forensic Path 16887680 816887680 816887680 816887680 816887872 816887872 816887872 816887872 816888128 816888128 816888320 816888320 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560</th><th>tch case forensidaptp.mem domain.bt. 316839911 ocsp.dgicert.com0L</th><th></th></t<>	Image File Feature File Feature File Forensic Path 16887680 816887680 816887680 816887680 816887872 816887872 816887872 816887872 816888128 816888128 816888320 816888320 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816888540 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560 816889560	tch case forensidaptp.mem domain.bt. 316839911 ocsp.dgicert.com0L	
	4850886158 ocsp. digicett com0L 4854811662 ocsp. digicett com0L 4980946999 ocsn. digicett com0L	816889920 816889984	cert.com/Di.+0@http://cacerts.digicert.com/DigiCertAssure dIDCodeSigningCA-1.crt0U0.0*HZ.Lb.Q.X Image: Comparison of the state of the	~

Figure 12: BEViewer window displaying filtered results from a "histogram" report feature, filtering on the term "digi"

4. Bookmarking Data: You can optionally Bookmark entries from report features within BEViewer. This creates a list of bookmarked entries that can be exported or easily referenced again at a future date. To bookmark a particular entry, select the desired entry and click the option in the toolbar labeled "Bookmark the selected feature" (icon of a star with a green arrow). To reference this bookmark in the future, you can click the option in the toolbar labeled "Manage bookmarks" (icon of a star without the green arrow).

Parsing Bulk_Extractor output

Most of the report features created by **Bulk_Extractor** are text-readable files stored in the output directory. These files can be easily parsed by external scripts and applications, if desired. NOTE: These files use a tabseparated-value (tab-delimited) format.

Some of the report features are constructed in XML or as SQLite database files. You may need to adjust your parsing software if you intend to use those files.

Learning Resources

The steps outlined here should be a good start to using **Bulk_Extractor** and **BEViewer** in general, and especially for parsing memory acquisitions. However, since the software also includes many features and capabilities not discussed here, following are resources for further learning:

- User Manual, by Jessica R. Bradley and Simson L. Garfinkel Written for version 1.4 of Bulk_Extractor, this manual still provides applicable and detailed information about the software. It is installed and accessible from your Windows Start Menu after the successful installation of the software, as a PDF file. It is also available from the Project Wiki, listed below.
- Bulk_Extractor 1.5 Overview, by Simson L. Garfinkel Presentation slides in PDF format describing many details of the software: <u>http://downloads.digitalcorpora.org/downloads/bulk_extractor/2014-07-17_BE15.pdf</u>
- Bulk_Extractor Project Wiki, Documentation Page Contains links to the User Manual, a Programmer's Manual, and a Worked Examples document. Although some of this information is outdated, much of it contains detailed, relevant information about the software: https://github.com/simsong/bulk_extractor/wiki/Documentation