Chapter 11

FORENSIC ANALYSIS OF VOLATILE INSTANT MESSAGING

Matthew Kiley, Shira Dankner and Marcus Rogers

Abstract Older instant messaging programs typically require some form of installation on the client machine, enabling forensic investigators to find a wealth of evidentiary artifacts. However, this paradigm is shifting as web-based instant messaging becomes more popular. Many traditional messaging clients (e.g., AOL Messenger, Yahoo! and MSN), can now be accessed using only a web browser. This presents new challenges for forensic examiners due to the volatile nature of the data and artifacts created by web-based instant messaging programs. These web-based programs do not write to registry keys or leave configuration files on the client machine. Investigators are, therefore, required to look for remnants of whole or partial conversations that may be dumped to page files and unallocated space on the hard disk. This paper examines the artifacts that can be recovered from web-based instant messaging programs and the challenges faced by forensic examiners during evidence recovery. An investigative framework for dealing with volatile instant messaging is also presented.

Keywords: Instant messaging, forensic analysis, volatile information, artifacts

1. Introduction

The popularity of instant messaging has exploded during the last decade. From a humble beginning as a UNIX command line application, instant messaging has become one of the most popular forms of communication. During the period of growth, traditional client-based messaging programs such as AOL Instant Messenger (AIM) have dominated. In fact, active AIM subscribers currently number more than 50 million [15]. However, newer web-based programs are becoming increasingly popular. E-Buddy, a web-based messaging program, has 35 million desktop subscribers and more than five million mobile users [1].

Due to its popularity and purported privacy, instant messaging is being exploited by criminals, especially online predators.

Web-based and mobile messaging services are valuable sources of evidence. However, dealing with volatile instant messaging requires entirely different investigative procedures. Forensic analysis no longer involves merely locating archived or deleted messages, and stored "buddy" lists.

This paper presents a brief overview of volatile instant messaging and discusses approaches for conducting an investigation involving a web-based messaging program. Artifacts and other forensically-significant information that can be obtained from four popular web-based instant messaging programs are examined in detail. Finally, an investigative framework for dealing with volatile instant messaging is outlined.

2. Volatile Messaging

Techweb [12] defines instant messaging as the process of "exchanging text messages in real-time between two or more people logged into a particular instant messaging service." Volatile instant messaging, on the other hand, is a relatively new concept, which has not been formally defined. We adopt an operational definition for the concept: "real-time messaging between two or more people using a web interface." This means that a user with access to a public terminal or web browser can engage in instant messaging without having to access a traditional client like AOL Instant Messenger or MSN. Implied in the definition is the concept of volatility. After the web browser is closed or the machine is shut down, no records of user activity or chat log archives are (conceivably) retained. This is the primary difference between volatile instant messaging and its traditional counterpart.

Traditional instant messaging relies on the existence of an installed client program (e.g., Yahoo Messenger or MSN). Most programs require the user to enter an online handle and password from a previously created account. However, this information can be falsified as little, if any, verification is performed [7]. The one benefit of user authentication (i.e., "logging in") is that the messaging server can archive the IP address of the user [15]. This makes it possible to pinpoint a user to a specific computer or geographical location.

The messaging server typically marks the user as online upon successful authentication and sign on. The program then displays a list of currently logged on "buddies" from the user's contact list. Although the first message is sent through the main servers, subsequent messages originate directly from the client machine, reducing traffic to the messaging servers [5]. This poses a potential problem in forensic investigations because conversations are not logged by messaging servers.

The upside of client-based messaging is that information can be recovered from a suspect's machine. Recent studies [2, 4, 10] report that the forensic analysis of instant messaging programs provides a variety of evidence, including chat logs, file transfers and registry artifacts.

Web-based only or volatile messaging programs require a different investigative approach from client-based messaging programs. This is because there are no installed programs and very little data may remain after a browser is closed. The next section examines four popular web-based only messaging programs and discusses what, if any, evidence may be retained and recovered.

3. Methodology

This paper reports on the results of tests conducted on four web-based instant messaging programs: (i) AIM Express, (ii) Google Talk, (iii) Meebo, and (iv) E-Buddy. The four web-based programs were chosen because of the popularity of their service and instant messaging client. The tests used a Dell Latitude 600 laptop with 1 GB RAM, Windows XP Professional Service Pack 2 and a 60 GB hard disk formatted with NTFS. Internet Explorer version 6.0.2900.2180 was used as the web browser for chat communications.

AIM Express and Google Talk are web-based clients that run their own protocol [13]. Meebo and E-Buddy, on the other hand, are browser-based clients that rely on other instant messaging services (e.g., Yahoo, MSN or AOL) [3].

The machine settings were verified prior to conducting the tests. The default virtual memory size was set at 768 MB to 1,536 MB, and the registry was checked to ensure that the page file is not erased during shut down [9]. Test data was created by conducting three different conversations for each messaging program. The conversations were limited to two participants and lasted three to four minutes. The frequency of the conversations closely imitated real-life scenarios; suspects generally engage in multiple, short conversations with their victims. The conversations were initiated by another machine, after which the laptop user replied to the message with unique phrases that would help identify the conversation.

The first step in the forensic examination was to acquire a bit-stream image of the laptop. Access Data's Forensic Toolkit (FTK) Imager version 2.5.1 and a Tableau T5 IDE write blocker with a 2.5 inch adapter were used for image acquisition. After acquiring and verifying the im-

Table 1. Unique phrases used as keywords.

AOL	Google Talk	Meebo	E-Buddy
bannnnanas	fuzzie logyck	meebomeebo	functionza
weirdtheme	spaces spled wrong	thisfoodisok	documnt this consrvation
this is a space	toomany	generastso	999-222-2222

age of the laptop hard drive under FTK Imager, the file was indexed using FTK version 1.7.1 build 07.06.22. Prior to reviewing the image, a keyword list containing distinct phrases used during the conversations was created (Table 1). Keyword searches based on the list were run on the indexed drive, resulting in a relatively fast sweep of the hard drive image. Unfortunately, this yielded fewer results than expected, making it necessary to perform a live (un-indexed) search with FTK.

Runtime DiskExplorer for NTFS version 3.03 was then used to examine the hard drive image at a lower level. Sector-by-sector searches were conducted to find the distinct phrases used during the conversations. This method was necessary due to the nature of volatile messaging. After the browser is closed and the page file contents are erased, data often resides in unallocated space until the operating system re-allocates the cluster. Performing a cursory search using an indexed image typically yields limited results in the case of volatile messaging.

Table 2. Artifacts from volatile messaging clients.

Program	Time Estimate	Conversation Details	Screen Names	Buddy List Details
AIM Express	X	X	X	X
Google Talk	X	X	X	X
Meebo	X			
E-Buddy	X		X	

4. Results

Table 2 lists the artifacts discovered in the four volatile messaging clients. Evidence of forensic value was retrieved from every volatile messaging client; however, complete chat logs were not recoverable.

Artifacts were found in various Internet file caches used by Internet Explorer. Each cache holds a different piece of data. The `History.IE5` directory contains an `Index.dat` file, which maintains a log of the user's Internet history without caching the content. This file is crucial to reconstructing a suspect's browsing history because the file contains the URL of the site visited, the last time the page was visited, and the number of times the page was viewed [6]. Also, several sub-directories within `History.IE5` show the date ranges for the logged entries.

The `Temporary Internet Files\Content.IE5` sub-directory stores cached web pages and images that the user has viewed, and makes them readily accessible should the site be visited again. This was implemented to reduce the time needed to load web pages; however, it also provides the forensic examiner with valuable information about user activity. In addition, the `Cookies` sub-directory contains files that web pages place on the user's computer. These "cookies" are used by web sites to track user behavior and maintain personalized settings.

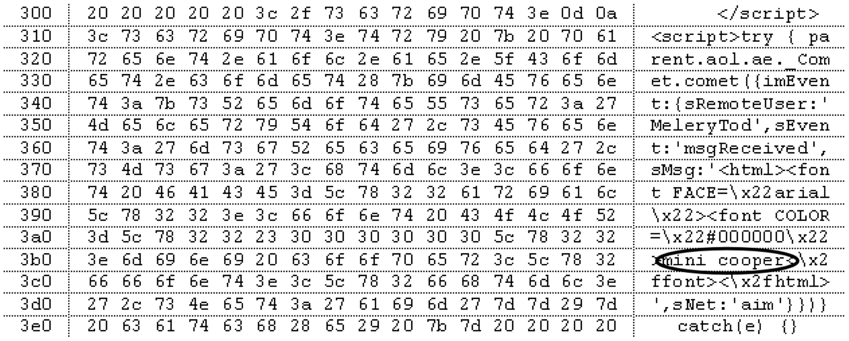
Many of the remaining artifacts were found in the drive free space (i.e., unallocated space on the drive). They consisted of screen names and, in the case of AIM Express, fragments of the buddy list. Snippets of AIM Express and Google Talk conversations were also found in the same location. Windows XP is known to use this space to store data that does not have to remain in memory or be saved on the hard drive. Note that this data is eventually overwritten.

Screen names were found in the `pagefile.sys` set of files. The operating system uses a page file to store information that should be in physical memory, but is not because it is used infrequently. The size of the page file is variable, but within a specified range; by default, the Windows XP range is 756 MB to 1,512 MB [14]. The forensic implications of modifying this range were not investigated in this study.

4.1 AIM Express

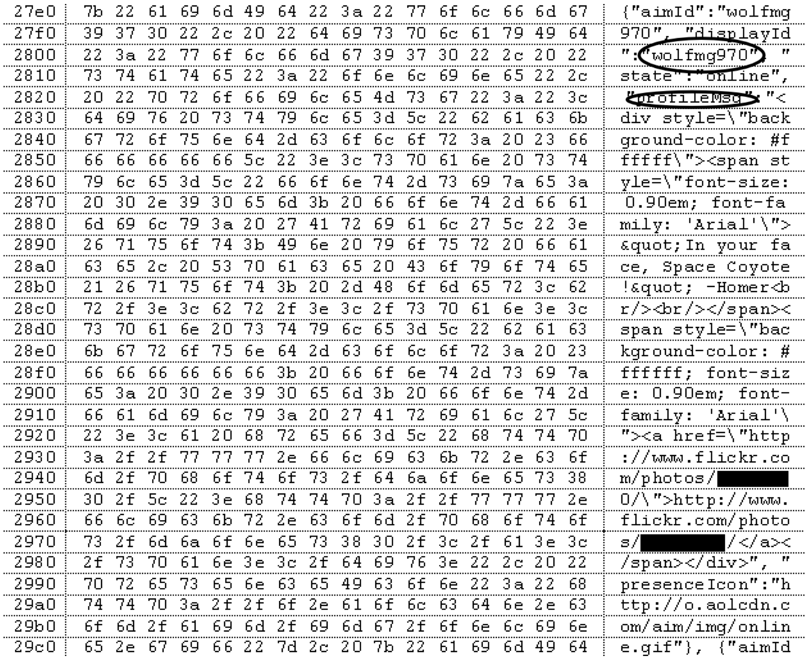
AIM Express left behind several artifacts, including snippets of conversations, details of the buddy list and approximate times when the conversations took place. The buddy list is extremely helpful in forensic investigations; this list can be used as a reference point to establish a social network. The approximate times of conversations can be estimated based on `Index.dat` entries made by AIM Express; these times can be used to construct timelines and sequences of key events.

Snippets of the other user's conversations and the buddy list were also found in the file slack and `pagefile.sys` file (Figure 1). This seems to agree with the observations of Dickson [2], except that this data was found on the hard disk rather than in RAM. In traditional instant messaging programs, such as AIM, chat logs are stored in files under locations specified by the user or in default locations such as the `Program Files` directory. Web-based conversations, unless specifically logged by



```
300  20 20 20 20 20 3c 2f 73 63 72 69 70 74 3e 0d 0a       </script>
310  3c 73 63 72 69 70 74 3e 74 72 79 20 7b 20 70 61   <script>try { pa
320  72 65 6e 74 2e 61 6f 6c 2e 61 65 2e 5f 43 6f 6d   rent.aol.ae._Com
330  65 74 2e 63 6f 6d 65 74 28 7b 69 6d 45 76 65 6e   et.comet({imEven
340  74 3a 7b 73 52 65 6d 6f 74 65 55 73 65 72 3a 27   t:{sRemoteUser:'
350  4d 65 6c 65 72 79 54 6f 64 27 2c 73 45 76 65 6e   MeleryTod',sEven
360  74 3a 27 6d 73 67 52 65 63 65 69 76 65 64 27 2c   t:'msgReceived',
370  73 4d 73 67 3a 27 3c 68 74 6d 6c 3e 3c 66 6f 6e   sMsg:'<html><fon
380  74 20 46 41 43 45 3d 5c 78 32 32 61 72 69 61 6c   t FACE=\x22arial
390  5c 78 32 32 3e 3c 66 6f 6e 74 20 43 4f 4c 4f 52   \x22><font COLOR
3a0  3d 5c 78 32 32 23 30 30 30 30 30 30 5c 78 32 32   =\x22#000000\x22
3b0  3e 6d 69 6e 69 20 63 6f 6f 70 65 72 3c 5c 78 32   >mini cooper<\x2
3c0  66 66 6f 6e 74 3e 3c 5c 78 32 66 68 74 6d 6c 3e   ffont><\x2fhtml>
3d0  27 2c 73 4e 65 74 3a 27 61 69 6d 27 7d 7d 29 7d   ',sNet:'aim'}})}
3e0  20 63 61 74 63 68 28 65 29 20 7b 7d 20 20 20 20    catch(e) {}
```

Figure 1. Conversation snippet from slack space.



```
27e0  7b 22 61 69 6d 49 64 22 3a 22 77 6f 6c 66 6d 67   {"aimId":"wolfmg
27f0  39 37 30 22 2c 20 22 64 69 73 70 6c 61 79 49 64   970", "displayId
2800  22 3a 22 77 6f 6c 66 6d 67 39 37 30 22 2c 20 22   ":"wolfmg970", "
2810  73 74 61 74 65 22 3a 22 6f 6e 6c 69 6e 65 22 2c   state":"online",
2820  20 22 70 72 6f 66 69 6c 65 4d 73 67 22 3a 22 3c    "profileMsg":"<
2830  64 69 76 20 73 74 79 6c 65 3d 5c 22 62 61 63 6b   div style=\"back
2840  67 72 6f 75 6e 64 2d 63 6f 6c 6f 72 3a 20 23 66   ground-color: #f
2850  66 66 66 66 66 5c 22 3e 3c 73 70 61 6e 20 73 74   fffff\"><span st
2860  79 6c 65 3d 5c 22 66 6f 6e 74 2d 73 69 7a 65 3a   yle=\"font-size:
2870  20 30 2e 39 30 65 6d 3b 20 66 6f 6e 74 2d 66 61    0.90em; font-fa
2880  6d 69 6c 79 3a 20 27 41 72 69 61 6c 27 5c 22 3e   mily: 'Arial'\">
2890  26 71 75 6f 74 3b 49 6e 20 79 6f 75 72 20 66 61   &quot;In your fa
28a0  63 65 2c 20 53 70 61 63 65 20 43 6f 79 6f 74 65   ce, Space Coyote
28b0  21 26 71 75 6f 74 3b 20 2d 48 6f 6d 65 72 3c 62   !&quot; -Homer<b
28c0  72 2f 3e 3c 62 72 2f 3e 3c 2f 73 70 61 6e 3e 3c   r/><br/></span><
28d0  73 70 61 6e 20 73 74 79 6c 65 3d 5c 22 62 61 63   span style=\"bac
28e0  6b 67 72 6f 75 6e 64 2d 63 6f 6c 6f 72 3a 20 23   kground-color: #
28f0  66 66 66 66 66 66 3b 20 66 6f 6e 74 2d 73 69 7a   ffffff; font-siz
2900  65 3a 20 30 2e 39 30 65 6d 3b 20 66 6f 6e 74 2d   e: 0.90em; font-
2910  66 61 6d 69 6c 79 3a 20 27 41 72 69 61 6c 27 5c   family: 'Arial'\
2920  22 3e 3c 61 20 68 72 65 66 3d 5c 22 68 74 74 70   "><a href=\"http
2930  3a 2f 2f 77 77 77 2e 66 6c 69 63 6b 72 2e 63 6f   ://www.flickr.co
2940  6d 2f 70 68 6f 74 6f 73 2f 64 6a 6f 6e 65 73 38   m/photos/[redacted]
2950  30 2f 5c 22 3e 68 74 74 70 3a 2f 2f 77 77 77 2e   0/\">http://www.
2960  66 6c 69 63 6b 72 2e 63 6f 6d 2f 70 68 6f 74 6f   flickr.com/photo
2970  73 2f 6d 6a 6f 6e 65 73 38 30 2f 3c 2f 61 3e 3c   s/[redacted]/</a><
2980  2f 73 70 61 6e 3e 3c 2f 64 69 76 3e 22 2c 20 22   /span></div>", "
2990  70 72 65 73 65 6e 63 65 49 63 6f 6e 22 3a 22 68   presenceIcon":"h
29a0  74 74 70 3a 2f 2f 6f 2e 61 6f 6c 63 64 6e 2e 63   ttp://o.aolcdn.c
29b0  6f 6d 2f 61 69 6d 2f 69 6d 67 2f 6f 6e 6c 69 6e   om/aim/img/onlin
29c0  65 2e 67 69 66 22 7d 2c 20 7b 22 61 69 6d 49 64   e.gif"}, {"aimId
```

Figure 2. Screen name and profile message in `fetchBuddyInfo.htm` file.

the user, are stored in temporary Internet directories that may or may not remain after the browser is closed. If these directories have been deleted or overwritten, more powerful forensic tools are required to view conversations in drive free space or file slack.

The `fetchbuddyInfo.htm` file, which is found under the `Temporary Internet Files\Content.IE5` directory within the profile's local settings, contained expanded buddy list information for the screen names obtained from the laptop (Figure 2). This information is valuable in cases where additional profile evidence is necessary. A profile often lists personal interests and hobbies, possibly even a home address. In addition, the expanded profile can provide investigative clues about the suspect's behavior and potential contacts, and help determine geographic areas of activity.

The `Index.dat` entries in `Temporary Internet Files\Content.IE5` show the screen name of the user as well as the time of the conversation. This allows an investigator to make an estimate of when the conversation took place. Finally, the user's screen name can be found in the following files: `$Logfiles`, `$MFT records`, `username@aimexpress.aol[1].txt` and `aimtoday.aim[1].txt`. Although these files may not provide crucial evidence, they can be used to corroborate other events.

4.2 Google Talk

Google Talk left several artifacts in the `Temporary Internet Files\Content.IE5` directory, e.g., the `accountinfo.htm` file, which displays the screen name used to sign on to Google. More importantly, the data gathered from slack space showed portions of all three conversations from both parties. These conversations were found by running keyword searches on the unique phrases used to distinguish the conversations. It is important to note that un-indexed searches were used to obtain these results; a normal indexed search yielded no results. Entries made in the `Index.dat` file within the `History.IE5` directory were also discovered. These entries can be used to correlate the time the user logged into gmail and the interface through which Google Talk was accessed.

4.3 Meebo and E-Buddy

Details about Meebo and E-Buddy conversations could not be found. The two programs function as true volatile messaging clients – virtually all the information about a conversation disappears after it ends. This is partly due to the heavy use of JavaScript on both websites. By maintaining a constant server-side connection via JavaScript, the site is able to maintain the appearance of a desktop application [8]. However, this has the effect of limiting the amount of information that can be gathered from the hard drive. Ultimately, the most useful artifacts found were the `Index.dat` entries, which showed when the E-Buddy and Meebo websites were accessed. In addition, the `ebuddy.htm` file in the `Temporary Internet Files` folder retains the screen name that the user used to sign on to the service.

5. Investigative Framework

Having discussed the artifacts that can be recovered from web-based instant messaging programs, we present a preliminary framework for investigators. This framework has three phases: recognition, formulation and search.

- **Recognition:** The first step in searching for evidence of volatile messaging is to identify if and when a web-based instant messaging conversation took place using the suspect machine. This is accomplished by searching for the existence of temporary Internet files or `Index.dat` entries that indicate the suspect signed on to a messaging service. For example, AIM conversations are indicated in temporary Internet files (e.g., `fetchBuddyInfo.htm`) while Google Talk conversations are identified by the presence of the `AccountInfo.htm` file. In situations where the Internet history or cache have been erased or are unavailable, manual indexed and non-indexed searches using the files mentioned above or search terms such as `.Ebuddy` may also yield results. Note that E-Buddy uses named servers (e.g.,"Kentucky") for logging in clients.

- **Formulation:** The formulation phase uses data gathered from the recognition phase to populate the list of possible screen names and other keywords used as input in the search phase. Snippets of previous instant messaging conversations may also be used to populate the list. In addition, any unique or misspelled words known by the investigator should be included in the list of search terms as they are likely to be found in chat conversations [11].

- **Search:** The search phase uses indexed and un-indexed searches to locate volatile messaging artifacts. Fast indexed searches that use the list created during the formulation phase should be performed first. If the results are inconclusive or incomplete, "live" or un-indexed searching is necessary. This is especially true for items found in slack or unallocated space because text residing in these locations may not be properly indexed by the forensic tool. The results from this phase can be used in subsequent searches.

The most challenging aspect of an examination is finding proof that a volatile messaging conversation ever took place. However, once evidence of this activity is found, search terms may be compiled and executed. Complete conversations may never be uncovered. Nevertheless, extensive live and un-indexed searches often yield successful results.

6. Conclusions

Web-based instant messaging presents challenges for forensic examiners due to the volatile nature of the data and artifacts created by the messaging programs. Forensic evidence is recoverable after these programs have been used, but investigators must know certain elements of the conversations in order to perform string searches. Even so, time-consuming sector-by-sector searches are required to uncover all the potential evidence.

Our research has revealed that several useful items of information can be recovered; these include the list of user contacts, snippets of conversations and the approximate time of the last conversation. In most cases, multiple instances of these items are found; they can be used to help corroborate other pieces of evidence found on the target system. The investigative framework proposed for the four web-based instant messaging programs considered in our study formalizes the task of evidence recovery. However, additional research is required to test the validity of this framework on other browsers and instant messaging clients.

Acknowledgements

This work was partially supported by the National Science Foundation under ITR Grant No. 0428554.

References

[1] Australian IT, E-Buddy gets growth message (www.ebuddy.com/press/auit_article.pdf), November 7, 2006.

[2] M. Dickson, An examination into AOL Instant Messenger 5.5 contact identification, *Digital Investigation*, vol. 3(4), pp. 227–237, 2006.

[3] A. Ghag, Top 10 web-based instant messengers (www.tech2.com/india/topstuff/websites-internet/top-10-webbased-instant-messengers/2892/0), 2006.

[4] W. Gillam, Instant messaging artifacts for cyber investigations, Unpublished manuscript, Department of Computer and Information Technology, Purdue University, West Lafayette, Indiana, 2006.

[5] A. Grossman, No don't IM me: Instant messaging, authentication, and the best evidence rule, *George Mason Law Review*, vol. 13(6), pp. 1309–1340, 2006.

[6] K. Jones and R. Belani, Web browser forensics, Part 1 (securityfocus.com/infocus/1827), 2005.

[7] D. Juhnke and D. Stenhouse, Instant messaging: What you can't see can hurt you (in court) (www.forensics.com/pdf/Instant_Messaging.pdf), 2005.

[8] Meebo, Meebo Forum (forum.meebo.com/viewtopic.php?t=12476).

[9] Microsoft Corporation, How to clear the Windows paging file at shutdown, Microsoft Help and Support, Redmond, Washington (support.microsoft.com/kb/314834), 2007.

[10] New York State Computer Forensic Workgroup, Messaging: A forensic view, presented at the *Ninth Annual New York State Cyber Security Conference* (www.cscic.state.ny.us/security/conferences/security/2006/Presentations/hurbanek.swf), 2006.

[11] J. Reust, Case study: AOL Instant Messenger trace evidence, *Digital Investigation*, vol. 3(4), pp. 238–243, 2006.

[12] Techweb, Instant messaging (www.techweb.com/encyclopedia/defineterm.jhtml?term=instantmessaging), 2007.

[13] H. Tschabitscher, Top 10 free email services (email.about.com/cs/freeemailreviews/tp/free_email.htm).

[14] D. Waddington and D. Hutchison, Resource partitioning in general purpose operating systems: Experimental results in Windows NT, *ACM SIGOPS Operating Systems Review*, vol. 33(4), pp. 52–74, 1999.

[15] Yahoo! IP address (info.yahoo.com/privacy/us/yahoo/ipaddress/details.html), 2008.